Variation between the copula si "to be" and the l-clitics in Romani spoken in Mexico

Cristian Padure, Stefano de Pascale, Evangelia Adamou

Variation between the copula *si* 'to be' and the *l*-clitics

in Romani spoken in Mexico

Cristian Padure, Inalco
Stefano De Pascale, KU Leuven
Evangelia Adamou, CNRS (French National Centre for Scientific Research)

Address for correspondence:
evangelia.adamou@cnrs.fr

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

Abstract

This research aims to investigate the innovative use of the subject clitic pronouns *lo*, *la*, *le* in attributive predications, e.g., *O xabe lašo lo* 'The dinner is good', in Romani spoken in the State of Veracruz, Mexico. The analysis of a 15-hour conversational corpus in Romani shows that the Romani copula *si* is used in variation with the subject clitic pronouns in *l-* in third person, affirmative clauses. In addition, sixty Romani-Spanish bilinguals from Veracruz responded to a contextualized copula choice task. Generalised linear mixed effect models were constructed to analyse the results. The analysis shows that the clitics are extremely dynamic in third person affirmative sentences and further reveals the linguistic variables that determine their use.

Keywords: Romani, Mexico, Spanish, variation, attributive clauses

1. Introduction

Adamou (2013) first reported that heritage speakers of Romani in Oaxaca, Mexico developed a distinction between attributive predications using the copula *si* 'to be', as in (1a), or the third person clitic pronouns in *l-*, as in (1b). In contrast, Romani speakers from Europe only use the copula *si* 'to be' (Matras 2002, Elšik and Matras 2006).

Romani spoken in the State of Oaxaca, Mexico

(1a) le šave muʁa bibiake **si** barbale
DEF.PL children POSS.1SG aunt.DAT be.3PL rich
'My aunt's children are rich.' (Adamou 2013:1085)

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

In addition to the Mexican data, Acuña and Adamou (2013) presented the results of a pilot study on similar uses in Romani spoken in Bogotá, Colombia, indicating that this innovation might be a widespread Romani feature in Latin America. This study aims to explore further the variation between the Romani copula and the innovative uses of the *l*-clitics in attributive clauses. It investigates copula choice in Romani in a larger group of Roma residing in the State of Veracruz, Mexico, where the same variation is encountered (see example 2), illustrating the variation between the Romani copula and the clitic in *l*-.

Romani spoken in the State of Veracruz, Mexico

(2)	o	mobili	**si**	kalo	aj	i	tapiceria	nevi=**la**
	DEF.SG.M	car	be.3SG	black	and	DEF.SG.F	upholstery	new-3SG.F

'The car is black and the upholstery is new'.

Section 2 provides some background information on Roma in the Americas and on the linguistic phenomenon under study. In Section 3, we present data from the conversational corpus and from the preference questionnaire in Section 4. Section 5 summarises the findings and offers some concluding remarks.

2. Background

2.1. Romani in the Americas and in Mexico

The Romani presence in the Americas is an under-studied topic that began to emerge in Romani studies mainly in the 2000s. In linguistics, alongside early work by Pickett (1962) and Pickett and Gonzalez (1964), there is only a recent paper by Adamou (2013) on Romani spoken in Mexico, an unpublished PhD dissertation by Deman (2005) on Bogotá Romani in Colombia, and a few studies on Chilean Romani by Salamanca and Rodríguez (2009) and Lizarralde and Salamanca (2010), and on Romani from Argentina (Bernal 1984).

It is estimated that there are between 1.5 and 3.5 million[1] Roma in the Americas. Most of these Roma probably migrated to the Americas along various routes in the nineteenth century as part of the more general European migration, but archival work is virtually non-existent (see Gomez Alfaro 1998, Pardo-Figueroa 2013, Sutre 2014). Some Romani communities were already settled in countries such as Brazil and Argentina following earlier migrations from Spain and Portugal starting in the sixteenth century, and historical documents mention the presence of the *gitanos* in eighteenth-century Mexico. Another important wave of migration took place after the Second World War and mobility to several countries in Northern, Central, and South America still continues today. Anthropological and linguistic research indicates that the Romani migration mainly concerned Kalderash Roma (from current-day Romania and Ukraine), but also the Xoraxane (Turkish) Roma from the Balkan countries.

In Mexico, Roma are settled in several cities, mainly in Mexico City, but also in Tuxtla Gutierrez (State of Chiapas), as well as in the outskirts of the cities of Oaxaca, Veracruz, Puebla, and Guadalajara (see Ripka 2007 and Muskus Guardia 2012 on the community in Guadalajara). The data presented in this paper were collected in 2016 and 2017 in the small city of La Rinconada in Veracruz State, Mexico (see map in Figure 1). Most Roma from 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

[1] The lower estimates are cited in the Revue des Etudes Tsiganes (2012) and the higher estimates are provided by the Romani organization SKOKRA.

Mexican Spanish. The Romani variety spoken in Veracruz shares several features with the south-eastern dialects of Europe and more specifically the Vlax dialects (dialect classification in Matras 2005), similar to the variety spoken in Oaxaca (Adamou 2013).

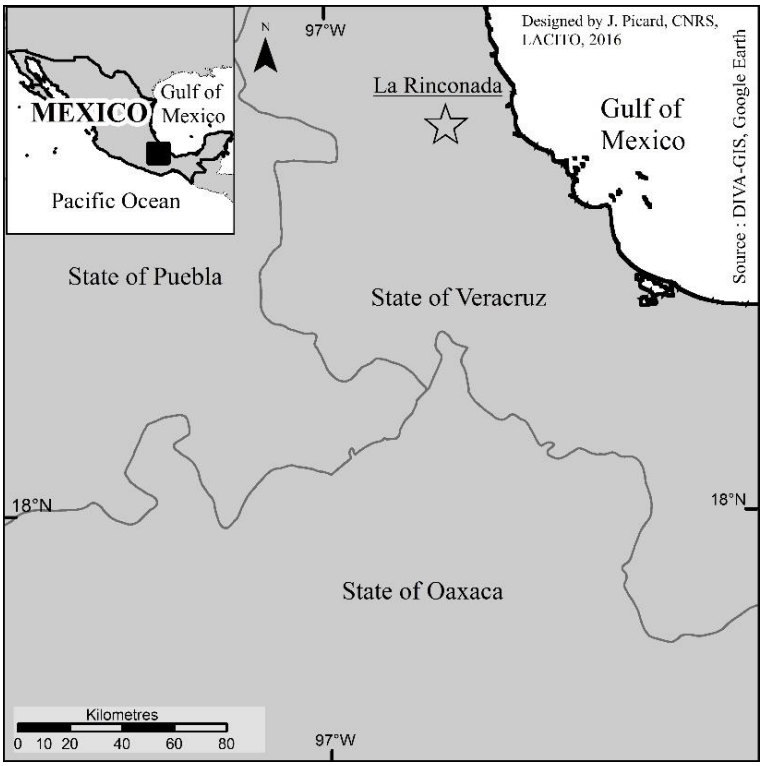

Figure 1. Map of Mexico. The study was conducted in the locality of La Rinconada in the Veracruz State.

2.2. The Romani clitic pronouns in *l-*

This study focuses on the third person clitic pronouns *lo, la, le*; see Table 1. These subject clitic pronouns should not be confused with the etymologically related and formally similar third person oblique pronouns (*le, la, le*). In Mexican Romani, the oblique pronouns are indeed used as direct objects with full lexical verbs and combine with the copula 'to be' to form the possessive construction e.g., *si la love* 'she has money'; *si le nonituria* 'he has children'. The *l-* pronouns we are focusing on are dubbed *clitics* because they always need to attach (*encliticise*) to other words and their position in the clause is not fixed. They are marked at the level of the transcription by the symbol = and they attach to their host.

Table 1. Third person clitic pronouns in Mexican Romani

Singular	Masculine	*lo*
	Feminine	*la*
Plural		*le*

Romani subject clitic pronouns in *l-* are an archaism (Matras 2002:102). Reconstruction of pre-European Romani, known as Proto-Romani, indicates the use of a set of demonstratives in *l-*, which developed first into full third person pronouns before they were used as clitics M.SG **-ta*>**-lo*; F.SG **-ti*>**-li*; PL **-te*>**-le* (Matras 2000:111). When Roma arrived in Asia Minor and the Balkans, a new set of demonstratives most likely developed as third person pronouns (M.SG **ov*, F.SG **oj*, PL **on*), first for emphasis, and then as the default pronouns which can be found nowadays in most Romani dialects (Matras 2000:111). This new development is

believed to have led to the loss of the *l-* subject pronouns in most present-day dialects spoken in Europe. Indeed, the clitics in *l-* have disappeared from the Northeastern and Northern Central Romani dialects (Matras 2002:102). In the Vlax Romani dialects, subject clitics can only be found in nonverbal predication with presentatives e.g., *eta lo* 'there he is!', place deictics e.g., *kate lo* 'here he is!', and interrogatives e.g., *kaj lo* 'where is he?' (Elšik and Matras 2006:316). In the North-western dialects, such as Finnish Romani, Welsh and English Romani, Slovene/Croatian Romani and partially in Caló (Spain) and in Romungro (Slovakia), the subject clitics in *l-* are used in the existential predications with the copula *si*, e.g., *si lo* 'he is' (Matras 2002:102). Finally, some dialects, such as Austrian Lovari and Klenovec Romungro (Slovakia), Sinti, and Roman (Austria), have maintained the use of the subject clitics with full lexical verbs (Elšik and Matras 2006:213).

The use of the subject clitics in *l-* in attributive predications e.g., *she is tall*, as those documented for Romani spoken in the Americas (Adamou 2013), are not encountered in European dialects. Adamou argued that the Mexican Romani speakers developed two conceptual representations of *being* to parallel those of Spanish, expressed by the copulas *ser* and *estar* 'to be', and then recruited obsolescent material in Romani i.e., the subject clitic pronouns, to replicate the uses of *estar*. Indeed, when two languages in contact have different conceptual representations encoded by distinct linguistic means, conceptual transfer is likely to occur from a dominant language to a heritage or minority language and from a first language (L1) to a second language (L2) (Jarvis and Pavlenko 2008 for an overview).

3. The conversational corpus

We follow a community-based data collection method in accordance with the variationist tradition in linguistics. In particular, we arrange the analysis of spontaneous speech recorded by trained fieldworkers through informal interviews that target the narration of personal experiences or topics specific to the local community (see Labov 1969, 1984).

3.1. Method

Participants. The sample comprises 19 Romani-Spanish bilinguals from the State of Veracruz, Mexico (4 women; age range 30-90). They were all residents of the community of La Rinconada. They were informed of the broader goals of the study, that is, to describe the Romani variety they speak and learn about their customs, but they did not have any information about the specific phenomenon that was under study. All participants gave written consent in Spanish and received no compensation for their participation in the study.

Corpus. The corpus is composed of interviews of a total of 15 hours and 40 minutes in Romani. Each interview lasted an average of 48 minutes.

Procedure. The Roma participants were interviewed in 2016 in their homes by the first author of the paper who is a Romani native speaker of a similar dialect from Europe. As the interviewer is not an in-group community member and speaks a different Vlax dialect, he was particularly careful to also employ the *l*-clitics in his speech. The interviews were recorded with a Marantz recorder and an external microphone. Interviewees would discuss their way of life in Mexico and their Romani traditions in the community. The relevant structures were transcribed in an Excel file and coded for speaker, gender, the variant used (copula or clitic), as well as for person number, adjective class, predicate type, frame of reference, experience with referent, change, and animacy. The data were anonymised.

3.2. Results

The analysis of the corpus reveals the use of 116 affirmative attributive clauses in third person. Fifty attributive clauses are constructed with the Romani copula and 66 with the innovative clitics. In attributive clauses in the third person, the copula si 'to be' was found in the speech of altogether 13 speakers while the *l*-clitics were found among 18 speakers. These overall rates confirm that the clitics in *l*- are used in the community and that they are dynamic.

Examples in (3) illustrate the two variations. In these examples, the copula *si* 'to be' is used with an animate referent and the clitic *lo* with an inanimate, but all combinations are found in the corpus in similar proportions.

(3a)	vo	motholas	ke	**si**	čoʁo
	3SG.M.NOM	screamed.3SG	that	be.3SG	poor
	'He screamed that he is poor.'				(SP28; 1:15:42)

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

Examples in (4) illustrate the use of the copula and the clitics in third person plural affirmative clauses. Again, there are no clear tendencies in the corpus and plural is less common with both the copula (N=12) and the clitics (N=18).

(4a)	kadia	si	džene	kaj	**si**	uče
	this	be.3PL	person.PL	that	be.3PL	tall.PL

aj	**si**	teloʁe	**si**	čule
and	be.3PL	small.PL	be.3PL	fat.PL

aj	**si**	ezlabi
and	be.3PL	thin.PL

'The people are like this, they are tall and small, they are fat and thin.' (SP28; 21:5)

(4b)	bešen	ando	boš	pues	melale=**le**
	sit.3PL	in.DEF.SG.M	forest	therefore	dirty-3PL
	'They live in the forest and therefore they are dirty.'				(SP28; 0.18)

The analysis of the spontaneous speech of the Roma speakers in Veracruz confirms the existence of variation between the copula and the clitics in attributive affirmative clauses in the third person. However, due to the small number of occurrences and the variability of the lexical items and contexts, the conversational data do not allow for a clear identification of the variables that favour the use of the copula or the clitics. For this reason, we turn to a structured elicitation task that allows for analysis that is more systematic.

4. The preference task

4.1.Method

Participants. Sixty Romani-Spanish bilinguals from the State of Veracruz, Mexico, participated in this study, including all 19 speakers that were interviewed for the corpus study (48 male, 12 female; the sample is skewed toward male speakers because of local norms discouraging sessions between male outsiders and female members of the community). Thirty-two participants were early simultaneous bilinguals, i.e., they acquired both languages before the age of 3, 27 were early sequential bilinguals, i.e., they acquired Romani before the age of 3 and Spanish after the age of 3, and one had acquired Romani after the age of 18. 57 participants had less than 12 years of education. Age of participants ranged from 17 to 90 (M = 37.08; SD = 18.86). Participants were all residents of the community of La Rinconada. All the men worked in the car trade and the women often did house-work. Two participants were attending high school at the time of the study. All participants gave written consent and received no compensation for their participation in the study. A celebration dinner was organised by the interviewer to thank the community for their participation in this study.

Materials. We used the contextualised copula choice task, which was developed in Spanish by Geeslin and Guijarro-Fuentes (2008). This task comprises 28 pre-constructed clauses with *ser* or *estar*, introduced by a paragraph-long context that forms a coherent narration; see Appendix A.

Procedure. The study was conducted in 2016. The participants were tested in their homes. The testing was conducted by the first author who is a Romani native speaker of a similar Vlax dialect from Europe. After giving their written consent, participants heard a recording of the instructions in Spanish and each of the 28 clauses in Spanish introduced by a paragraph-long context. The recording was made by a native Mexican Spanish speaker. For each clause, participants were asked to choose between the copulas *ser* and *estar* or to indicate when both were applicable. The participants were then immediately asked by the interviewer in Romani to translate the target clauses into Romani, i.e., 'How would you say this in Romani?'. In total, each participant responded to 56 questions.

Analysis. 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.[2] The analyses aimed at modelling Romani copula choice in the bilingual group. Since the copula alternation in Romani only appears in affirmative clauses and in the third plural and singular person, only the clauses with these features were retained for the analyses from the Geeslin and Guijarro-Fuentes (2008) set. Clauses 1, 3, 8, 9, 10, 12, 13, 24, 25, 27, 28 (either negative or first and second person) were therefore not considered; see Appendix A. Furthermore, we discarded clauses in which respondents considered both copulas appropriate or in which the Romani translation did not feature a copula at all. Consequently about 51% of all results collected were retained for further analysis [final size = 858 data points].

We coded extra-linguistic variables "generation" (three balanced age groups of 20 participants each, i.e., young for ages 17-22, middle for ages 23-45, old for ages 48-90);

[2] Packages in R are units of shareable statistical material developed and issued by R-community members. They bundle code, data and documentation, often coherently revolving around specific statistical techniques (clustering, regression analyses, multidimensional scaling etc.). In this respect, the lme4 package offers many functions specifically designed to perform linear and generalized linear mixed-effects regression and analyse its output. For this study, we used the function glmer, which allows us to fit a regression model with both fixed effects and random effects on a binary dependent variable, that is, copula variants in either Romani or Spanish.

"gender" (male vs. female), as it is widely shown in the variationist literature that social characteristics of the speaker may correlate with the linguistic variations (see an overview in Labov 2001). As we do not have a good understanding of the variation in Romani, we coded the linguistic variables that have been widely discussed in the literature on Spanish copula variation (as coded in the questionnaire by Geeslin and Guijarro-Fuentes 2008). More specifically, we coded "frame of reference" (class, that is, when referents are compared to a set of referents that share the same property, vs. individual, when comparison for a given referent is made between two points in time), "experience with referent" (immediate i.e., direct contact with the referent or surprise that results from the contrast between the speaker's observation in a given situation and her expectations, vs. ongoing that is, when the speaker has continuous experience with the referent), "change" (no vs. yes), and "animacy" (no vs. yes). We added the variables "person number" (third singular vs. third plural) and "Spanish copula choice" (*ser* vs. *estar*) which might be relevant for Romani. Fixed effects are shown in Table 2.[3] In addition, random intercepts were considered for "participant" and "experimental item".

We conducted generalised linear mixed-effects regressions, with "Romani copula" as response variable with two levels, *si* and *lo*. There are at least two advantages that favour this type of statistical analysis. First, it is able to explain the observed variation in the use of a binary response variable, such as this copula alternation with two levels. By constructing a regression model with several predictors, shown in Table 2, it can simultaneously be determined which predictors have the largest greatest on the choice of one of the two variants and which have none. These predictors are dubbed "fixed effects": they can be repeated in subsequent studies and their levels are supposed to exhaust all possible levels in the population (e.g., gender: male vs. female). Second, mixed-effects regressions allow the consideration of variability at the level of individual participants and experimental items. To cope with this variation we can include participants and items as "random effects": such effects are non-repeatable, since each new study will require new participants and items, and the levels thus constitute a random sample of its population. Furthermore, the addition of a participant random effect helps to better generalise up to the whole population of Romani speakers. In this study, we included formulation random effects, or more precisely random intercepts, in our initial model.[4]

Table 2. The fixed effects that were used in the statistical analyses

Fixed effects	**Levels**
extra-linguistic	
generation	*young, mid, old*
gender	*male, female*
linguistic	
frame of reference (abbr. referent)	*class, individual*
animacy	*animate, inanimate*
change	*change, no change*
experience with referent	*immediate, ongoing*
copula choice in Spanish	*ser, estar*
person number	*third singular, third plural*

[3] Frame of reference and predicate type are two variables that completely overlap in the questionnaire and therefore only frame of reference was maintained.

[4] When adding a random effect varying intercept and/or slopes can be specified. With these varying intercepts, we are basically acknowledging that some speakers have a baseline preference for using the *l*-clitic more often than the *si* copula, and that some items on average are more likely to be translated with the *l*-clitic than with the *si* copula.

4.2. Results

In this section, we present the results of the Romani translations (see Appendix B for the descriptive results). Before discussing the generalised linear mixed-effects model, we need to point out to some unexpected issues that arose from data collection. Figure 2 plots the distribution of the Romani copula for the different experimental items in the study. It can be seen that for approximately half of the items, participants translated almost categorically using the innovative *l*-clitics (i.e., percentages above 90%). In the case of Item 22 there was no sign of variation at all. We therefore continue the analysis by only focusing on experimental items in which either of the two Romani variants is chosen less than 98% of the time (in absolute numbers: more than 1). This results in the deletion of 20, 26, 18 and 22; see Appendix A. This new dataset contains 641 data points.

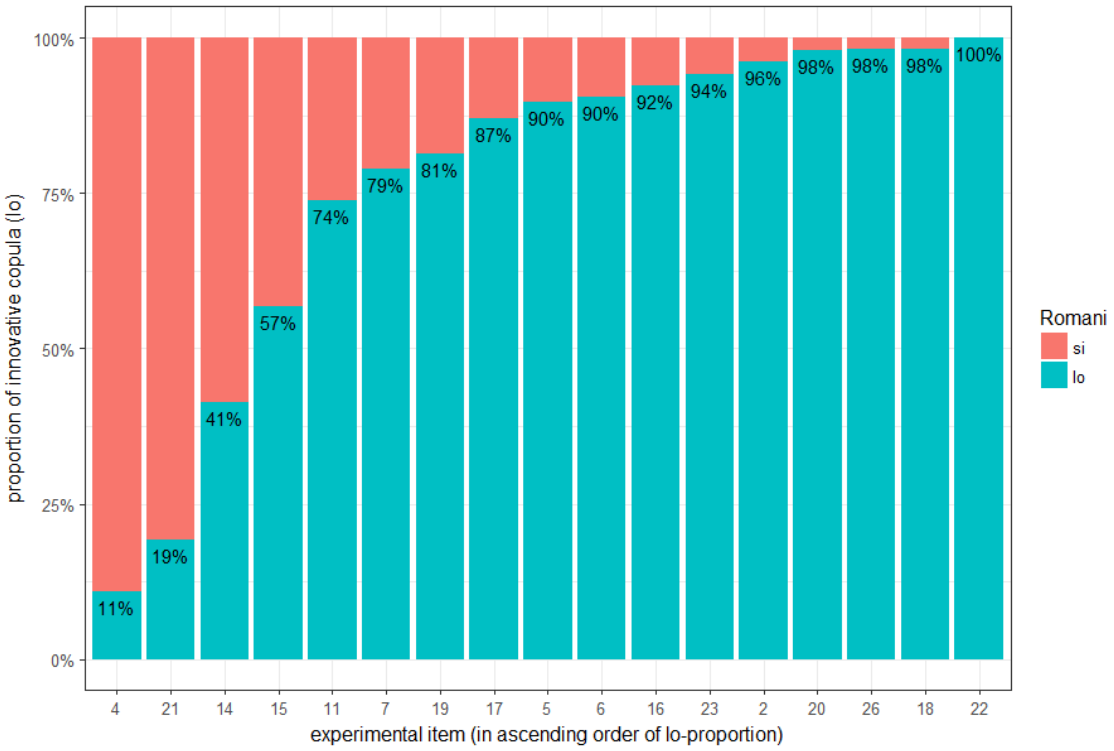

Figure 2. The Romani translations in the third person sentences, using either the copula *si* or the *l*-clitics (abbreviated as *lo*)

Generalised linear mixed-effects models could suffer when data are unbalanced, which often leads to data sparsity in one of the levels of the binary response variable and causes estimation problems. As plotted in Figure 2, the Romani data for third person are unbalanced in that our participants prefer to translate the items much more often with the *l-clitic* (77%) than with the *si copula* (23%). Popularised in linguistics by Tagliamonte and Baayen (2012), a Random Forests analysis pursues the same goal as the regression analysis, but as a nonparametric method it is robust against correlated predictors, i.e., predictors with many levels, and data sparsity, both of which we observe in the Romani data (for an introduction see Strobl, Malley, and Tutz 2009). In a nutshell, the technique works as follows: it first performs independence tests between each predictor and the response variable and selects the predictor whose levels have the strongest discriminative power on the response variable. In the following step it splits the data according to this selected predictor and tests all the remaining predictors as before on these separate partitions of the data. The result is a Conditional Inference Tree, where the strongest discriminative predictor is at the top of the tree, and recursive splitting by other predictors generates a hierarchy of interacting predictors influencing the response variable. A Random

Forests analysis generates a large number of trees and evaluates the importance of the predictors based on their average accuracy.[5] We first performed a Random Forests analysis to determine a potentially significant fixed-effect structure (see Figure 3), and then inspected the best Conditional Inference Tree (see Figure 4).

The Random Forests analysis, plotted in Figure 3, shows that the predictors "change", "gender" and "generation" are not significant (their variable importance factors are all indistinguishable from zero). In contrast, the predictors "frame of reference (abbr. referent)" and "experience with referent" turn out to be very significant, while "animacy" and "person number" were significant to a lesser extent.

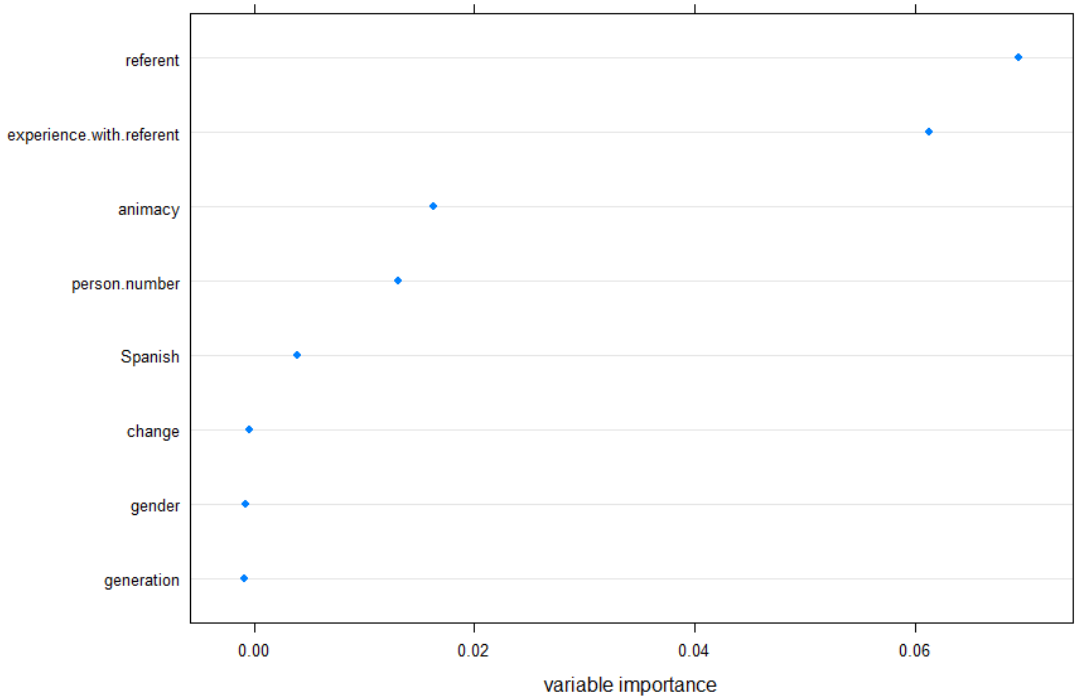

Figure 3. A Random Forests analysis for the choice of the clitics in Romani. The "variable importance" score is a measure of the average accuracy of a predictor over all grown Conditional Inference Trees of the forest (in our case over 500 trees). The higher the score, the more influential the predictor.

The clearest finding from the Conditional Inference Tree in Figure 4, is that while the Romani clitics in *l-* are the preferred option for class referents when the experience with the referent is ongoing, it is the less likely alternative in sentences with individual referents in that same kind of experience. A discrepancy between the output of the Random Forests and the Conditional Inference Tree is observed regarding the inclusion of the predictor "person number", which is not present in the conditional inference tree whereas it is in the Random Forests analysis.

[5] As previously, the R software was used to perform the analyses. The packages 'party' and 'partykit' were used for the Random Forests and Conditional Inference Tree calculations, specifically with the functions 'cforest' and 'ctree' respectively.

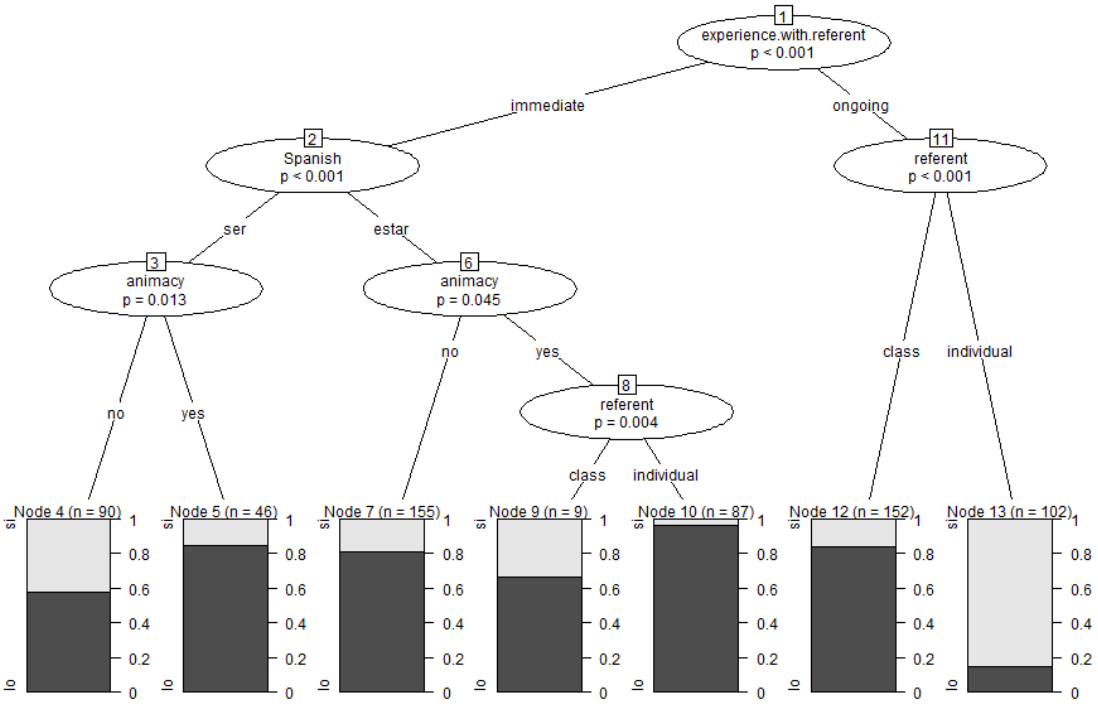

Figure 4. A Conditional Inference Tree for the choice of the clitics in Romani

Now that we have explored the intricate relationships between possible explanatory variables and the Romani copula alternation in our dataset, we can turn to the generalised linear mixed-effects regression. After a stepwise backward model selection procedure on a maximal model with the abovementioned predictors (except for "change", "gender" and "generation", which were not significant) and their interaction, we retained only "participant" ($p < 0.0001$) as random effect. This means that now, after removal of the skewed sentences shown in Figure 2, our fixed-effects structure and the random effect "participant" fits the variability of the remaining items perfectly. For this generalised linear mixed-effects regression on the bilinguals' choice of Romani copula, we were able to reveal highly significant interaction effect of "frame of reference" and "experience with referent" ($\chi^2 = 45.8159$, df = 1, $p < 0.0001$); "person number" and "frame of reference" ($\chi^2 = 24.3541$, df = 1, $p < 0.0001$); "frame of reference" and "animacy" ($\chi^2 = 24.1256$ 1, df = 1, $p < 0.0001$). Other significant interaction effects are those between "person number" and "Spanish copula" ($\chi^2 = 8.6847$, df = 1, $p < 0.001$) and "person number" and "animacy" ($\chi^2 = 6.6287$, df = 1, $p < 0.001$). The discovery of such significant interactions proves that it is relevant to add interaction terms to the model formula from the very start of the computation. These interactions show that the behaviour of one predictor is jointly influenced by the behaviour of another predictor (note that, in the interpretation of the results, these predictors should not be analysed on their own as relevant). The model has furthermore a very good predictive power (C = 0.934) and classification accuracy is well above chance level, i.e., 89% compared to 70% for always choosing the most frequent class i.e., the clitics in *l*-.

In what follows, the importance of the joint influence of predictors will become clearer when we visualise the effects on Romani copula alternation in different plots. These plots show on the y-axis the fitted probability for the "success" level of the response variable in the model i.e.,

the Romani clitics in *l*-.[6] In other words, as is customary during the interpretation of the regression results, we express the relationship between two categorical choices, *l*- vs. *si*, by means of a proportion of only one of the two variants, in this case the *l*-clitic. Proportions higher than 50% indicate a strong association between a certain predictor level (e.g., class referents) and the *l*-clitic. Proportions lower than 50% indicate a strong dissociation between a specific level of the predictor (e.g., individual referents) and the *l*-clitic. Alternatively, this dissociation can be more intuitively understood as a strong preference for the *si* copula. If the confidence bars do not include the 50% threshold, shown as a dashed red line, we can be confident that the effect we have observed is a true effect in the population and therefore not due to some sampling chance. The interacting behaviour of the predictors in the following plots is represented by means of a combination of colour-coding and the information on the x-axis: the first predictor and its levels are shown on the x-axis, and within each level on the x-axis we look at the (colour-coded) levels of the second predictor. As a concrete example, let us consider the left blue bar in Figure 5. It can be seen that in sentences with immediate experience with the referent (left position at the bottom) and an individual referent (blue colour) the participants overwhelmingly use the *l*-clitic (the top of the bar is well above the 80%, see below).

More specifically, Figure 5 shows that in cases of immediate experience with the referent (e.g., surprise or direct contact), Mexican Roma significantly prefer the *l*- clitic, regardless of the frame of reference. This amounts to 78% of the cases with class referents (i.e., a referent compared to a class of referents) and 91% of the cases with individual referents (i.e., a referent compared to itself). However, when confronted with clauses with ongoing experience with the referent, the frame of reference dramatically influences the choice of the Romani copula: participants overwhelmingly use *l*-clitics for class referents (87%), but completely avoid using the clitics for individual referents (only 1% of the cases, which boils down to an almost categorical choice for the *si* copula).

Example in (5a) illustrates the preferred clitic choice for immediate experience with what was termed an "individual" referent, that is, when the speaker talks about a property regarding a referent after direct observation and compares it to itself at some earlier stage: in this example the attribute "being angry" contrasts with other moments in life when Pablo was not angry. Example in (5b) illustrates the preferred use of the copula *si* when the referent, Raúl's friend with whom he has ongoing experience, is compared to another point in time when she was not a Catholic.

Individual referent, immediate experience with the referent
CONTEXT: [Paula says thank you and asks if their friend Pablo will be joining them for dinner. Paula wants to talk to him about their math class. Raúl says that Pablo isn't coming and Paula wants to know why:
Paula: Why isn't Pablo coming?
Raúl: Because I didn't call him earlier and *now he is mad*.]

(5a) akana xuljariko=**lo**
now angry-3SG.M
'Now he is mad.' (Preference task, item 2; SP 18 years old)

Individual referent, ongoing experience with the referent

[6] The use of the words "failure" and "success" outcome levels are naming conventions linked to the automatic conversion of the binary response variable (*si*, *lo*) to an internal "dummy variable" (with values 0 and 1). These wordings by no means imply that the use of the *l*-clitic is more successful than the use of the *si* copula in any other way than the technical manner used here.

CONTEXT: [Since Raúl is curious about Paula's boyfriend, they discuss the possibility of a wedding. Paula is not sure about marrying her boyfriend because she is Catholic and her boyfriend is Protestant. She knows Raúl had a friend who married a man with a different religion and she asks what happened with her:

Paula: What religion does your friend practice now?

Raúl: *Now she is Catholic, too.*]

(5b)	akana	bi	voj	**si**	katolika
	now	also	3SG.F.NOM	be.3SG	Catholic

'Now she is Catholic, too.' (Preference task, item 21; SP 50 years old)

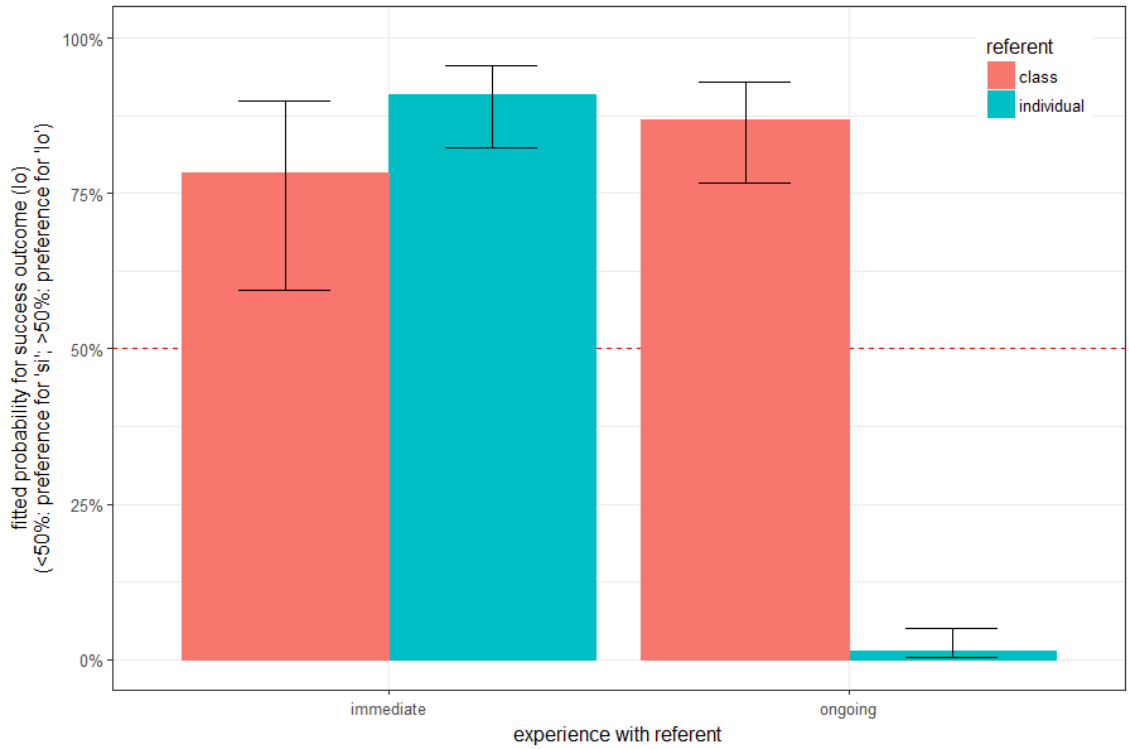

Figure 5. The selection of the Romani *l*-clitics with respect to the variable "frame of reference" (class vs. individual) and "experience with the referent" (immediate vs. ongoing).

Analysis of results shows that the predictor "frame of reference", that is, whether one compares the referent to a class of referents (class) or to itself (individual), further interacts with "animacy", as illustrated in Figure 6. While the *l*-clitic is used in 89% of the sentences with inanimate class referents, i.e., inanimate referents that are compared to a class of referents that share the same property, it is only used in 10% of the sentences with inanimate individual referents, i.e., when an inanimate referent is compared to itself. Regarding animate referents, we observe that the *l*-clitic is preferred across the board, both in sentences with class referents (57%), i.e., with animate referents compared to a class of referents who possess the same attribute, as in those with individual referents (79%), i.e., animate referents compared to themselves, although this preference only reaches significance for the latter.

Example in (6a) illustrates the preferred use of the clitic with an inanimate referent, i.e., the new dish that Raúl has ordered, which is compared to a class of referents, that is, other good dishes. Example in (6b) illustrates the preferred use of the clitic with an individual animate referent, i.e., Pablo, where the property 'mad' is compared to different states of mind of the same person, that is, those moments when Pablo is not upset, but may be sad, happy, etc.

Class, inanimate
CONTEXT: [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?
Raúl: Yes, *dinner is good*.]

(6a) o xabe lašo=**lo**
DEF.SG.M food good-3SG.M
'The dinner is good.' (Preference task, item 5; SP 20 years old)

Individual, animate
CONTEXT: [Paula says thank you and asks if their friend Pablo will be joining them for dinner. Paula wants to talk to him about their math class. Raúl says that Pablo isn't coming and Paula wants to know why:
Paula: Why isn't Pablo coming?
Raúl: Because I didn't call him and *now he is mad*.]

(6b) akana xuljariko=**lo**
now angry-3SG.M
'Now he is mad.' (Preference task, item 2; SP 18 years old)

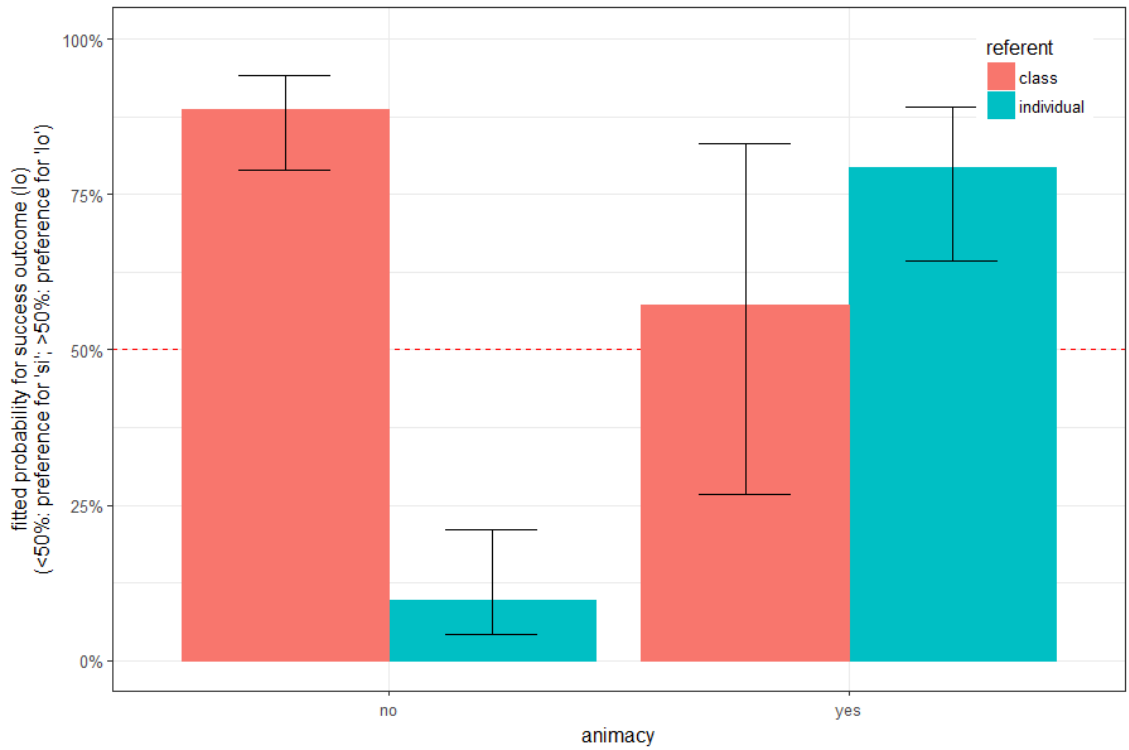

Figure 6. The selection of the Romani *l*- clitics with respect to the variable "frame of reference" (class vs. individual) and "animacy" (no vs. yes)

In Figure 7 it can be seen that participants fluctuate between the *l*-clitic and the *si* copula in sentences with third person plural subjects and that the frame of reference does not influence

that choice. *L*-clitics seem to be dispreferred with third person plural, but that does not reach significance (confidence intervals include 50%, in case of class referents [33%] as well as individual referents [46%]). In contrast, in sentences with third person singular subjects, Mexican Roma prefer overwhelmingly the use of the *l*-clitics for class referents (91%), but significantly disprefer the clitic in clauses with individual referents (28%).

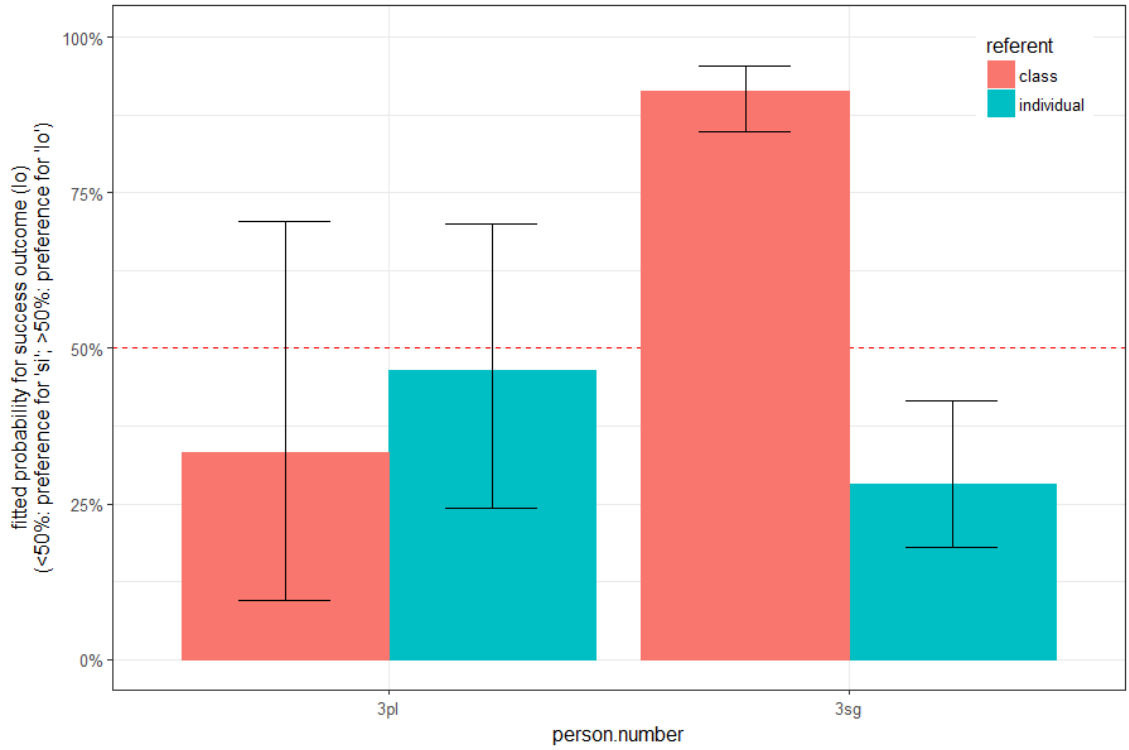

Figure 7. The selection of the Romani *l*- clitics with respect to the variable "frame of reference" (class vs. individual) and "person number" (third singular vs. third plural)

In Figure 8 we plot the interaction between the chosen "Spanish copula" and the "person number" of the subject of the clause. We observe that sentences that had previously triggered the choice of *estar* in Spanish are significantly more frequently translated with the *l*-clitic, whether the number of the subject in the clause is singular or plural (both in 75% of the cases). However, clauses for which the Mexican Roma participants had chosen *ser* in their Spanish responses are split up by person number: the respondents prefer the *l*-clitics over the *si* copula in third person singular clauses (75%), whereas they tend to prefer the *si* copula in third person plural clauses (100-37% = 63%).

These choices are illustrated in examples in (7). The example in (7a) illustrates the preferred use of the Romani clitic as the translation of a sentence with the Spanish copula *estar* in third person singular (we note that they would have chosen the clitic in third person plural too). The example in (7b) shows the preferred use of the Romani clitic in third person singular as the translation of a sentence with the Spanish copula *ser* (they would have chosen the copula *si* in third person plural).

Third person singular Romani clitic translating *estar*
CONTEXT: [Paula recalls that Raúl did not have class today. Since she can't remember why, she asks him.
Paula: Raúl, Why aren't you going to science class today?
Raúl: *Because the professor is sick*.]

(7a) o školari nasfalo=**lo**
DEF.SG.M professor sick-3SG.M
SPANISH: Porque el profesor está enfermo.
'Because the professor is sick.' (Preference task, item 19; SP 28 years old)

Third person singular Romani clitic translating *ser*
CONTEXT: [Paula and Raúl also get a chance to catch up on the events of the day. Raúl mentions that his sister called and said she might like to come visit and stay at the apartment for a while. Paula didn't know Raúl had a sister so she asks how old she is:
Paula: How old is your sister?
Raúl: *Like us, she is young.*]

(7b) sar amende voj desa terni=**la**
like 1PL.LOC 3SG.F.NOM very young-3SF.F
SPANISH: Como nosotros, ella es bastante joven.
'Like us, she is young too.' (Preference task, item 17; SP 34 years old)

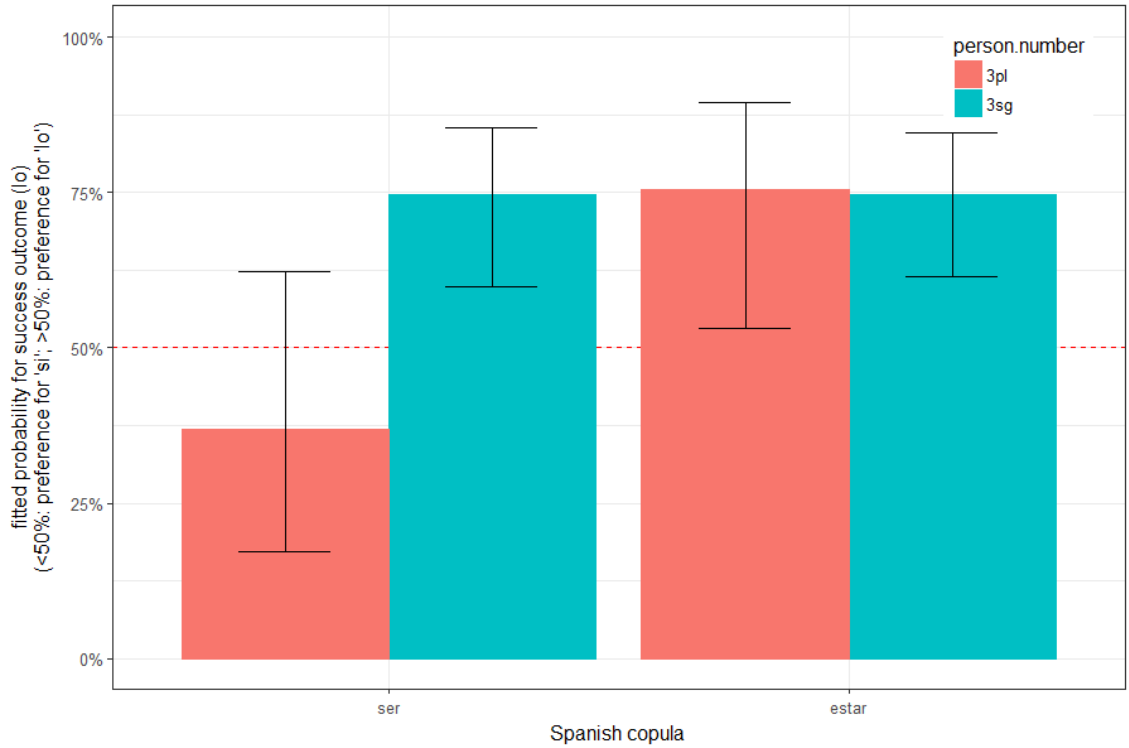

Figure 8. The selection of the Romani *l-* clitics with respect to the variable "person number" (third plural vs. third singular) and "Spanish copula" (*ser* vs. *estar*)

Finally, Figure 9 shows the last interaction between "animacy" and "person number". It can be seen that whereas third person plural clitics are used with animates and inanimates almost as frequently as the copula *si* 'to be', third person singular clitics are used almost categorically with animates.

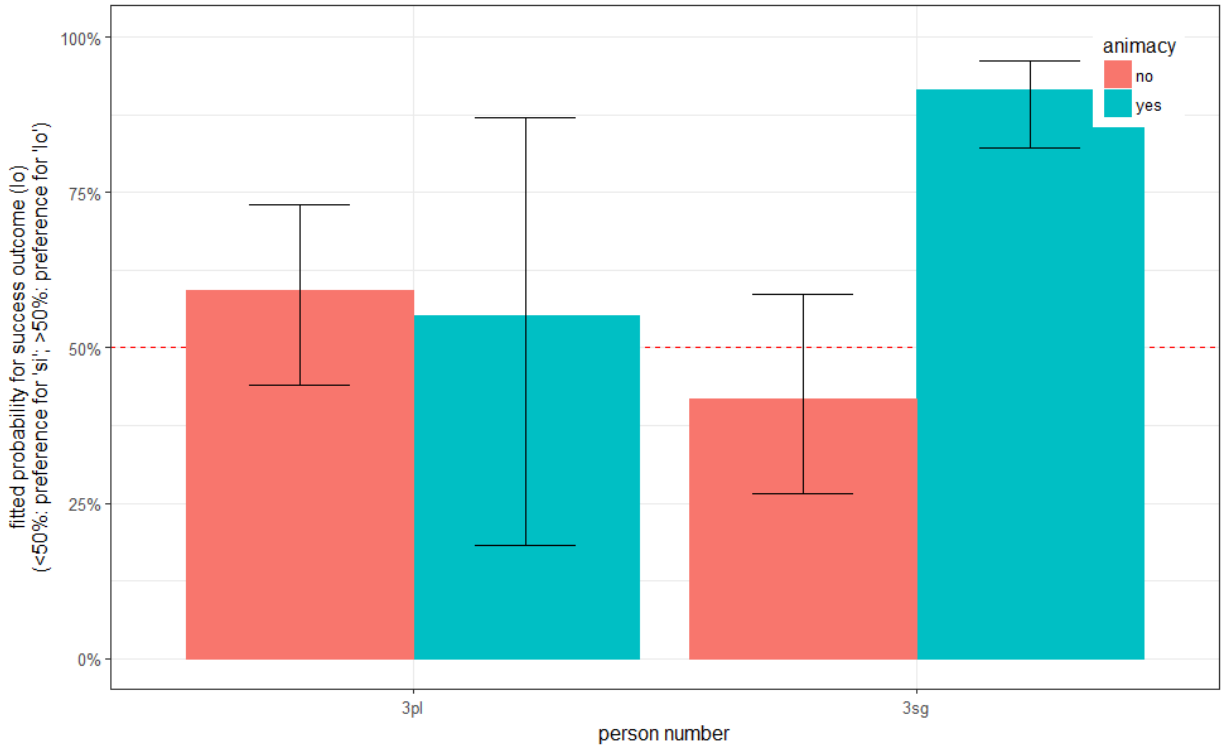

Figure 9. The selection of the Romani *l*- clitics with respect to the variable "animacy" (no vs. yes) and "person number" (third plural vs. third singular)

5. General discussion

This study confirms the preliminary observations in Adamou (2013) by examining a new, large corpus of conversational speech in Romani from the community of La Rinconada in the State of Veracruz, Mexico. In addition, the quantitative approach based on elicited data from 60 Romani-Spanish bilinguals residing in the same community allows for a much better understanding of the variation between the traditional Romani copula and the innovative clitics in *l*-.

In sum, the various statistical analyses reveal that for the choice of the *l*-clitics the linguistic variable "frame of reference" is crucial, that is, whether a referent in an attributive predication is compared to other similar referents (dubbed "class") or whether it is merely compared to itself (dubbed "individual"). The variable "frame of reference" has two significant interactions. The first one with the variable "experience with the referent" (whether the speaker has "immediate" experience with the referent, resulting from direct contact and eventually linked to surprise, or whether the speaker has "ongoing" experience with the referent, that is, when the observation can have been made at several points of time). The second interaction is with the variable "animacy" (when the referent is a person, dubbed "animate", or an object, dubbed "inanimate"). More specifically, it appears that in Romani from Veracruz when the speaker has an immediate experience with a referent he/she can use the innovative *l*-clitics whether the referent is compared to a class of referents or to itself. However, when the speaker has ongoing, continuous experience with the referent he/she will opt for the *l*-clitics to compare the referent to a class of referents that share the same property, but not to compare the referent to itself. In addition, Mexican Roma favour the innovative *l*-clitics for objects whose attributes are compared to the attributes of other similar objects (inanimate and class referents) and for people whose attributes are compared to themselves at some earlier stage (animate and individual referents). These linguistic variables are also considered as relevant in the extensive literature

on copula choice in Spanish, in particular, frame of reference and, its equivalent in this study, predicate type. These studies, however, show that linguistic variables interact with one another in a complex manner preventing researchers from isolating a single variable or a single set of variables to predict copula choice among all Spanish-speaking communities and individual speakers (see the results from several Spanish-speaking communities in the Iberian Peninsula in Geeslin and Guijarro-Fuentes 2008).

In addition, the Romani data from Veracruz confirm the importance of person number in the innovative uses of the clitics, as already noted in Adamou (2013), but they further highlight the dynamism of the third person singular clitics, in particular for animate subjects. In other words, it is extremely likely for a Roma speaker from Veracruz to use the *l*-clitics when describing another person rather than when describing a group of people or objects.

Finally, the current study furthers our understanding of the uses of the Romani copula *si* and the *l*-clitics and their correspondence to the Spanish copulas *ser* and *estar*. It appears that while Roma from Veracruz preferably translate sentences with *estar* using the Romani clitics in third person, both singular and plural, they translate those with *ser* using the Romani copula *si* only for third person plural and using the Romani clitics for third person singular. That is, while there is clear evidence of the correspondence between *estar* and the clitics, there is no evidence for the correspondence of Spanish *ser* and the Romani copula *si*. Rather, what we observe is the generalisation of the clitics in third person singular, a pattern that we can relate to the general trend in Spanish to extend *estar* in contexts previously occupied by *ser* (see for Mexican Spanish Gutiérrez 1994). This means that although Spanish has served as the model for the uses of the Romani clitics in attributive clauses and although the same linguistic variables as those found in Spanish are also encountered in Romani, the development of the Romani distribution between the clitics and the copula can differ from that of Spanish copula choice. In order to more accurately compare the Romani linguistic variables to the Mexican Spanish variables, we need to compare the Spanish responses of the Roma to those of the monolingual Mexican Spanish speakers of similar socio-economic groups, an endeavour that is the object of a different study.

To conclude, the contemporary Romani data from Mexico illustrate the long-term effects of partial conceptual equivalence encoded by distinct linguistic means in a bilingual context characterised by low normative pressure. Beyond the significance of the Romani data for contact linguistics, we should also stress their importance in that they offer a description of the ways in which Romani is spoken in the Americas. In a cross-disciplinary perspective, the exploration of this linguistic innovation among Roma residing in other Latin American countries could serve as a solid cue for reconstructing the Romani networks across the continent during the twentieth century in that the communities that share this innovation were most likely in contact with one another. More research needs to be done to document this innovation in other Spanish-speaking countries and check for their use in the communities settled in English-speaking countries such as the United States of America and Canada.

Abbreviations

1, 2, 3	first, second, third person
DAT	dative
DEF	definite
F	feminine
LOC	locative
M	masculine
NOM	nominative

PL	plural
POSS	possessive
SG	singular

Author Contributions

Conception of the study: EA; acquisition, annotation, and coding of the data: CP; statistical analysis of the data: SDP; drafting of the manuscript: EA; revising of the manuscript: EA, SDP; final approval of the content: CP, SDP, EA; supervision of all stages of the project: EA.

Acknowledgments

We are indebted to the Romani people of the Rinconada who kindly participated in this study. Data collection was supported by the French National Centre for Scientific Research (CNRS) and the INALCO, via the laboratory Oral Tradition Languages and Civilizations (LACITO UMR 7107), awarded to Cristian Padure. This research was partly 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 doctoral visiting fellow Stefano De Pascale.

References

Acuña, Esteban, and Adamou, Evangelia. 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, University of Strathclyde, Glasgow, September 12–14, 2013.

Adamou, Evangelia. 2013. Replicating Spanish estar in Mexican Romani. Linguistics 51(6):1075–1105.

Bates, Douglas, Maechler, Martin, Bolker, Ben, and Walker, Steve. 2015. Fitting Linear Mixed-Effects Models using lme4. Journal of Statistical Software 67:1–48.

Bernal, Jorge M. 1984. Influences of the Spanish language in the Gypsy dialects of Danubian or Romanian type in Argentina-South America. QSG 14:21–26.

Deman, Katharina. 2005. Untersuchung zur Grammatik der Romani-Varietät der Kumpanja in Bogotá/Kolumbien. PhD dissertation at the Universität Graz, Austria.

Elšík, Viktor, and Matras, Yaron. 2006. Markedness and language change. Berlin and New York: Mouton de Gruyter.

Etudes Tsiganes. 2012. Special issue Les tsiganes en Amérique du Sud, 51.

Geeslin, Kimberly L., and Guijarro-Fuentes, Pedro. 2008. Variation in contemporary Spanish: Linguistic predictors of estar in four cases of language contact. Bilingualism: Language and Cognition 11(3):365–380.

Gomez Alfaro, Antonio. 1982. La polémica sobre la deportación de los gitanos a las colonias de América. Cuadernos Hispanoamericanos 386:308–336.

Gutiérrez, Manuel J. 1994. 'Ser' y 'estar' en el habla de Michoacán, México. Mexico: UNAM.

Jarvis, Scott, and Pavlenko, Aneta. 2008. Cross-linguistic influence in language and cognition. New York and London: Routledge.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. Language 45(4): 715–762.

Labov, William. 1984. Field methods of the project on linguistic change and variation. In: Baugh, John, and Sherzer, Joel, eds. Language in use: Readings in sociolinguistics. Englewood Cliffs: Prentice Hall. 28–53.
Labov, William. 2001. Principles of linguistic change. Vol. 2 Social factors. Malden, MA: Wiley-Blackwell.
Lizarralde, Diego, and Salamanca, Gaston. 2010. Morfología de los adjetivos demostrativos en el rromané jorajané de Chile. Literatura y Lingüística 21:109–126.
Matras, Yaron. 2000. The structural and functional composition of Romani demonstratives. In: Elšík, Viktor, and Matras, Yaron, eds. Grammatical relations in Romani: The noun phrase. Amsterdam and Philadelphia: John Benjamins. 95–122.
Matras, Yaron. 2002. Romani: A linguistic introduction. Cambridge: Cambridge University Press.
Matras, Yaron. 2005. The classification of Romani dialects. A geographic-historical perspective. In: Halwachs, Dieter, and Schrammel, Barbara, eds. General and applied Romani linguistics. Munich: Lincom Europa. 7–26.
Muskus Guardia, Hector. 2012. Los Roma del Centro Occidente de México: Religión pentecostal y organización social. Unpublished MA thesis at the Collegio de San Luis, Mexico.
Pardo-Figueroa Thays, Carlos. 2013. Gitanos en Lima: Historia, cultura e imágenes de los rom, los ludar y los calé peruanos. Instituto Riva-Agüero, Pontificia Universidad Católica del Peru.
Pickett, David W. 1962. Prolegomena to the study of Gypsies of Mexico. M.A. thesis at the Syracuse University.
Pickett, David W., and Gonzalez, Rafael. 1964. A Mexican Gypsy legend. Journal of the Gypsy Lore Society 43:3–11.
R Core Team. 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Ripka, Štepan. 2007. Informe de avances de investigación: Gitanos pentecostales en México. Unpublished M.A. thesis at the Charles University of Prague.
Salamanca, Gaston, and Rodríguez, Ursula. 2009. Vitalidad del rromané medida de acuerdo a parámetros fonológicos: Diseño de un instrumento y resultados de una primera aplicación. Universum 24(1):183–205.
Strobl, Carolin, Malley, James, and Tutz, Gerhard. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14(4):323–348.
Sutre, Adele. 2014. “They give a history of wandering over the world”. A Romani clan’s transnational movement in the early 20th century. Quaderni Storici 49(2):471–498.
Tagliamonte, Sali A., and Baayen, Harald. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. Language Variation and Change 24(2):135–178.

Appendix A. The contextualised preference questionnaire

This Appendix contains the contextualized preference questionnaire in Spanish designed by Geeslin and Guijarro-Fuentes (2008).

1. Paula y Raúl van a un restaurante esta noche. Paula habla desde su habitación mientras se viste y hace los planes con Raúl, quien está en la sala. Cuando sale de la habitación le pregunta a Raúl:
Paula: ¿Quieres que vayamos en mi coche?

A. Raúl: ¡Ay! ¡Qué bonita estás!
B. Raúl: ¡Ay! ¡Qué bonita eres!

2. Paula le agradece el cumplido y le pregunta si viene su amigo Pablo al restaurante. Paula quiere discutir algo sobre la clase de matemáticas. Raúl le dice que Pablo no vendrá y Paula quiere saber por qué:
Paula: ¿Por qué no viene Pablo?

A. Raúl: Porque no le llamé antes y ahora está enojado.
B. Raúl: Porque no le llamé antes y ahora es enojado.

3. Esto le sorprende a Paula porque Raúl y Pablo son buenos amigos. Paula sabía que Raúl pensaba llamar a Pablo y le pregunta qué pasó:
Paula: ¿Por qué no llamaste a tu mejor amigo?

A. Raúl: Porque este año no es amable conmigo.
B. Raúl: Porque este año no está amable conmigo.

4. La situación le interesa a Paula porque Raúl no parece preocupado por el comportamiento de Pablo. Paula quiere saber si Pablo normalmente se comporta así:
Paula: ¿Pablo se comporta así frecuentemente?

A. Raúl: No, me trata bien cuando está alegre.
B. Raúl: No, me trata bien cuando es alegre.

5. Paula y Raúl salen del apartamento y van al restaurante. Comen allá 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?

A. Raúl: Sí, la cena es buena.
B. Raúl: Sí, la cena está buena.

6. Durante la cena, Paula y Raúl tienen la oportunidad de charlar sobre muchos asuntos diarios. El apartamento en el que viven tiene un alquiler muy alto y por eso discuten la posibilidad de cambiar de sitio. Paula quiere saber lo que piensa Raúl sobre el apartamento:
Paula: ¿Te gusta nuestro apartamento?

A. Raúl: Sí, nuestro apartamento es grande.
B. Raúl: Sí, nuestro apartamento está grande.

7. A Paula le gustaría vivir cerca de la universidad. Le pregunta a Raúl sobre un apartamento que visitaron ayer:
Paula: ¿Qué piensas del apartamento que vimos ayer?

A. Raúl: No me gustó el dueño del apartamento, está desagradable.
B. Raúl: No me gustó el dueño del apartamento, es desagradable.

8. Paula está de acuerdo y ella también reaccionó de la misma forma. Ella no entiende por qué el dueño los trató tan mal.
Paula: ¿Crees que nos trató mal por no tener dinero?

A. Raúl: No, no somos pobres.
B. Raúl: No, no estamos pobres.

9. Raúl piensa que el dueño los trató mal por su edad. Paula no está de acuerdo y le explica que ella no parece tan joven porque tiene unos años más que los otros estudiantes.
Paula: No puede ser por eso, voy a cumplir 23 años mañana.

A. Raúl: Ah, ¡qué vieja estás!
B. Raúl: Ah ¡qué vieja eres!

10. Paula sabe que Raúl saca muy malas notas en la universidad. El padre de Raúl trabaja mucho para pagar los gastos de la universidad y Raúl tiene miedo de decirle que va mal con las clases. Paula le pregunta si puede evitar hablar con su papá sobre las notas.
Paula: ¿Tienes que hablar con tu papá?

A. Raúl: Sí, claro, mi papá no está estúpido.
B. Raúl: Sí, claro, mi papá no es estúpido.

11. Paula le pregunta más a Raúl la razón de sus malas notas.
Paula: Raúl, ¿Por qué recibes tantas malas notas?

A. Raúl: Mis profesores dicen que en comparación con el año pasado mis ensayos son peores.
B. Raúl: Mis profesores dicen que en comparación con el año pasado mis ensayos están peores.

12. Paula tiene una teoría sobre las dificultades que Raúl tiene este año. Ella sugiere que le cuesta adaptarse a los nuevos profesores.
Paula: ¿Crees que las dificultades resultan por el cambio de metodología de la clase?

A. Raúl: No, soy acostumbrado a la metodología.
B. Raúl: No, estoy acostumbrado a la metodología.

13. Paula y Raúl han tenido problemas con sus horarios. Como sólo Paula tiene coche, Raúl tiene que saber qué días él tiene que ir en autobús. Durante el día, Paula encontró una solución para ahorrar dinero:
Paula: Puedes venir conmigo por la mañana y volver con Juan por la tarde.

A. Raúl: Ay, ¡Qué inteligente estás!
B. Raúl: Ay, ¡Qué inteligente eres!

14. A Raúl le gusta la idea pero Paula cree que hay un posible problema. A veces, Juan tiene dificultades con la policía porque conduce demasiado rápido. Raúl dice que no le asusta la velocidad, pero sí le importa la apariencia de Juan. Paula no entiende lo que quiere decir:
Paula: ¿Por qué no te gusta su apariencia?

A. Raúl: ¡Esta semana su pelo es azul!
B. Raúl: ¡Esta semana su pelo está azul!

15. Este comentario le parece a Paula bastante gracioso. Ella no sabía que Raúl tuviera ideas tan tradicionales. Ella le muestra las uñas que acaba de pintarse.
Paula: Entonces, ¿qué piensas de las uñas?

A. Raúl: ¡Las uñas están azules también!
B. Raúl: ¡Las uñas son azules también!

16. Paula ve que Raúl no tiene una reacción tan fuerte como esperaba. Ella decide preguntarle otra vez qué opina de sus uñas azules.
Paula: En serio Raúl, ¿qué piensas de las uñas?

A. Raúl: La verdad es que hoy tus uñas son bonitas.

B. Raúl: La verdad es que hoy tus uñas están bonitas.

17. Paula y Raúl también tienen la oportunidad de hablar sobre cómo les fue el día. Raúl menciona que su hermana lo llamó y que ella quiere visitarlos y quedarse en el apartamento unos días. Paula no sabía que Raúl tuviera una hermana y quiere saber más de ella:
Paula: ¿Cuántos años tiene tu hermana?

A. Raúl: Como nosotros, ella está bastante joven.
B. Raúl: Como nosotros, ella es bastante joven.

18. Se le ocurre a Paula que Raúl no va a clase hoy. No sabe el porqué y le pregunta:
Paula: Raúl, ¿Por qué no vas a la clase de ciencia hoy día?

A. Raúl: Porque el profesor es enfermo.
B. Raúl: Porque el profesor está enfermo.

19. Durante la cena Raúl nota que Paula parece un poco diferente. Paula se da cuenta de que la mira y le pregunta qué le pasa:
Paula: Raúl, ¿Qué te preocupa?

A. Raúl: No sé por qué, pero tus ojos son azules esta noche.
B. Raúl: No sé por qué, pero tus ojos están azules esta noche.

20. Paula le dice a Raúl que ella había pensado que él se había dado cuenta de que el novio de ella acababa de darle un anillo nuevo. Con orgullo, ella le muestra el anillo a Raúl:
Paula: ¿Qué piensas de mi anillo nuevo?

A. Raúl: ¡Qué bonito está!
B. Raúl: ¡Qué bonito es!

21. A Raúl le interesa el novio de Paula, y ellos discuten la posibilidad de que ella se case. Paula tiene dudas porque ella practica la religión católica y su novio practica la religión protestante. Ella sabe que una amiga de Raúl se casó con un hombre de otra religión y le pregunta como resolvieron el problema:
Paula: ¿Qué religión practica tu amiga ahora?

A. Raúl: Ahora ella es católica también.
B. Raúl: Ahora ella está católica también

22. Mientras el camarero quita la mesa, Paula se da cuenta de que Raúl no se comió la sopa. Como Raúl siempre come lo que pide, ella quiere saber qué pasó:
Paula: Raúl, ¿Por qué no te comes la sopa?

A. Raúl: La sopa es muy fría.
B. Raúl: La sopa está muy fría.

23. Paula se ríe porque Raúl pidió el Gazpacho, una sopa que se sirve fría. Paula nota que a los tomates en el gazpacho les falta madurar y ella le pregunta otra vez:
Paula: ¿Por qué no te comes el gazpacho?

A. Raúl: Los tomates están verdes.
B. Raúl: Los tomates son verdes.

24. Paula acaba de encontrar un trabajo en la universidad. Cuando el camarero trae la cuenta, Raúl dice que ella debe pagar la cuenta. Paula no tiene la misma idea:
Paula: ¿Por qué crees que yo debo pagar la cuenta?

A. Raúl: Porque ahora estás rica.
B. Raúl: Porque ahora eres rica

25. Al salir del restaurante, Paula y Raúl ven a unos amigos suyos que quieren ir al cine. Ellos le piden a Paula que los lleve en coche al cine. Apenas hay espacio en el coche para todos y Paula les dice:
Paula: Hay mucha gente en mi coche.

A. Raúl: Sí, ahora tu coche no es muy grande.
B. Raúl: Sí, ahora tu coche no está muy grande.

26. Después de despedirse de los otros amigos, Raúl le dice a Paula que se alegra de que ella no haya querido ir al cine con ellos. Paula le pregunta a Raúl por qué él no quería ir y Raúl le explica que no le gusta Alicia:
Paula: ¿Por qué no te cae bien Alicia?

A. Raúl: Alicia está muy triste.
B. Raúl: Alicia es muy triste.

27. Paula le dice a Raúl que su reacción le sorprende. Ella no sabía que Raúl conocía a Alicia. Raúl empieza a reírse de Paula porque Paula no se ha dado cuenta de que Alicia y Raúl son del mismo pueblo. Paula no entiende:
Paula: ¿Por qué te ríes?

A. Raúl: ¡Nunca eres despierta!
B. Raúl: ¡Nunca estás despierta!

28. Paula y Raúl llegan al apartamento otra vez. Paula nota que comió mucho y tendrá que comer menos mañana.
Paula: Comí tanto que voy a empezar a engordar.

A. Raúl: Pero Paula, tú estás muy delgada.
B. Raúl: Pero Paula, tú eres muy delgada.

Appendix B. Romani responses to the preference questionnaire (third person affirmative clauses only)

Choice	Copula 199		Clitics 675	
Predicate type	Individual 40	Stage 159	Individual 311	Stage 364
Referent	Individual 159	Class 40	Individual 364	Class 311
Experience with the referent	Immediate 85	Ongoing 114	Immediate 475	Ongoing 200
Change	Yes 145	No 54	Yes 529	No 146
Animacy	Yes 112	No 87	Yes 303	No 372

Status	9
Size	47
Sensory characteristic	102
Physical state	104
Physical appearance	50
Personality	31
Mental state	111
Description	115
Colour	58
Age	48

Adjective class	
Status	39
Size	5
Sensory characteristic	5
Physical state	4
Physical appearance	4
Personality	9
Mental state	53
Description	35
Colour	38
Age	7