

# COPULA DELETION IN SAN ANDRESAN CREOLE<sup>1</sup>

## ELISIÓN DE LA CÓPULA VERBAL EN CRIOLLO SANANDRESANO

**CRISTINA SUÁREZ-GÓMEZ**

Universidad de las Islas Baleares  
cristina.suarez@uib.es

**MARGARITA MARÍA CHAMORRO-DÍAZ**

Centro Educativo y Cultural Español Reyes Católicos, Bogotá  
marchamo3@gmail.com

13

---

### Abstract

This paper deals with copula deletion in San Andresan Creole (SAC), an English-lexifier creole spoken in the Caribbean islands of Colombia. One of the most widely studied features of Caribbean creoles is the variable use of the verb BE (see Labov 1969; Holm 1976; Rickford 1996; Sharma and Rickford 2009; Michaelis et al. 2013, etc.). We aim to establish the linguistic and social determinants of observable variation in the copula system of SAC. To this end we will look primarily into BE presence (e.g. *dei waz der an di fishin graun* ‘they were there at the fishing ground’) vs BE deletion (e.g. *shi veri hongri* ‘she (was) very hungry’), and its distribution according to structural variables (e.g. grammatical context, grammatical category and grammatical person of the subject, and tense). The probabilistic analysis of the results shows that grammatical context and grammatical category of the subject determine variation in this domain of grammar, repeating a recurrent pattern shown by other Atlantic creoles. These findings provide a more complete picture of variation in the use of BE in SAC and offer valuable evidence regarding the vitality, unity and heterogeneity of this creole.

**Keywords:** San Andresan Creole, copula deletion, grammatical variation, grammatical context, subject category, grammatical person, tense.

## Resumen

Este artículo versa sobre el fenómeno de elisión de la copula verbal en criollo sanandresano, una lengua criolla de base léxica inglesa hablada en las islas caribeñas de Colombia. Uno de los fenómenos lingüísticos más estudiados en lenguas criollas del Caribe es la variación en el uso del verbo BE (véase Labov 1969; Holm 1976; Rickford 1996; Sharma y Rickford 2009; Michaelis et al. 2013, etc.). El objetivo de este trabajo es arrojar luz sobre los factores que condicionan la variación del sistema copulativo en criollo sanandresano. Para llevar esto a cabo analizamos los contextos en los que se usa BE de forma explícita (por ejemplo, *dei waz der an di fishin graun* ‘they were there at the fishing ground’) y aquellos en los que se elide (e.g. *shi veri hongri* ‘she (was) very hungry’), y prestamos atención a su distribución teniendo en cuenta variables estructurales tales como el contexto gramatical en el que se usa o elide BE, la categoría y persona gramatical del sujeto, y el tiempo verbal. El análisis probabilístico de los resultados muestra que el contexto y la categoría gramatical del sujeto determinan la variación de este fenómeno lingüístico, confirmando así un patrón de variación observado en otras lenguas criollas del Atlántico. Los resultados obtenidos en este estudio nos permiten obtener una descripción más completa del uso de BE en criollo sanandresano al tiempo que nos ofrecen información valiosa sobre la vitalidad, unidad y heterogeneidad de esta lengua criolla.

**Palabras clave:** criollo sanandresano, supresión del verbo copulativo, variación gramatical, contexto gramatical, categoría del sujeto, persona gramatical, tiempo verbal.

## 1. Introduction

The copular verb BE is one of the most widely studied linguistic phenomena in creoles. A vast number of research papers have been published on copula variation in creoles since early work by Labov (1969, 1972a) on African American Vernacular English (AAVE), with follow-ups on other well-known Atlantic Creoles, such as Jamaican Creole (Holm 1976; Rickford 1996, 1998; Deuber 2014), Gullah (Weldon 2003), Trinidadian Creole (Deuber 2014) and Guyanese Creole (Bickerton 1971), to cite the best known (see Section 2). Our aim in the present paper is to contribute to this topic by analyzing the phenomenon of copula variability in San Andresan Creole (henceforth SAC), a little-studied Caribbean creole. This will allow us to place SAC on the map of copula variability in Atlantic Creoles.

SAC is an English-based creole spoken in the islands of San Andrés, Providencia and Santa Catalina, located on the north-west Atlantic coast of Colombia (see

## Copula Deletion in San Andresan Creole

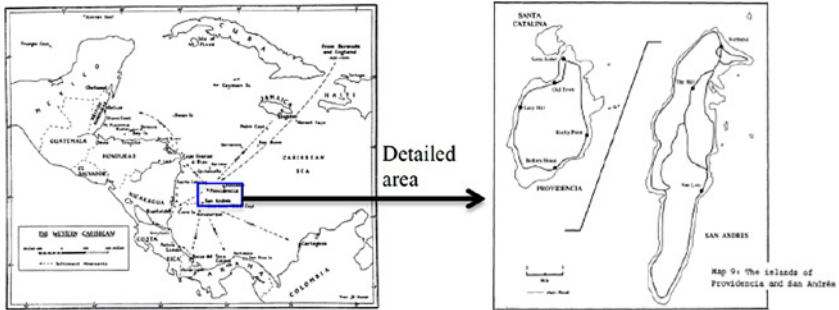


Figure 1. San Andrés and Providencia islands (Source: Ramírez-Cruz 2017: 1)

Figure 1). It developed in these islands during the seventeenth century, emerging through a process of language contact with British English (superstrate language) together with West African languages (from the Atlantic slave trade) and Spanish (substrate languages). The current linguistic situation of Colombia includes around 70 languages from different language families, two creole languages, plus Romani and Spanish, the latter divided into two main varieties: Coastal Spanish (regions of the Pacific and Atlantic coasts) and Andean Spanish (the rest of the country). Together with Spanish, SAC has been the official language of these territories since 1991 (Article 10, Constitution of Colombia). The dominant language in these islands today is Spanish, spoken by some 98% of the population, irrespective of context and age (Andrade Arbeláez 2012), although SAC remains the vehicular language in informal conversations, and the native language of the *raizal*<sup>2</sup> community, the local ethnic group.<sup>3</sup>

Unlike other Caribbean creoles, there is little linguistic research into SAC. Most studies deal with the sociolinguistic situation of these islands (Hooker et al. 2002; Moya-Chaves 2010; García León and García León 2012) and the vitality of SAC in different contexts, such as the educational system (Morren 2001; Hooker et al. 2002; Bowie and Dittmann 2007; Guerrero 2008; Moya-Chaves 2010) or the media (Sanmiguel 2007). There are some linguistic descriptions of SAC, but they tend to be of qualitative nature, as is the case with the studies by Dittmann (1992), O'Flynn de Chaves (1990, 2002) and Bartens (2013). In themselves, these are very useful in that they provide a detailed and complete picture of what SAC is like, focusing particularly on pronunciation and grammar. A recent PhD by Ramírez-Cruz (2017) provides a more systematic account of SAC, both linguistically (especially at the level of lexis and morphosyntax) and ethnolinguistically.

In terms of morphosyntax, SAC is very similar to other Caribbean English-based creoles. The following are among the most notable features reported by Bartens (2013) (see also Dittmann (1992) and Chamorro-Díaz and Suárez-Gómez (2019)) in SAC:

- unmarked SVO clausal word-order (Dittmann 1992: 64; Bartens 2013: Feature 1);
- the use of particles marking the grammatical categories Tense-Aspect-Mood (TAM system) preceding the verb (Bartens 2013: Feature 43) (e.g. *de* and *wen de* for the progressive or *go* and *gwain* for the future; see Dittmann 1992: 67-70);
- lack of inflectional endings to indicate the present-past tense distinction of lexical verbs (Bartens 2013: Feature 49);
- lack of inflectional endings to mark number in nouns (Dittmann 1992: 75; Bartens 2013: Feature 22; Chamorro-Díaz and Suárez-Gómez 2019: 140);
- case syncretism in the pronominal system (Chamorro-Díaz and Suárez-Gómez 2019: 139-140);
- the use of invariable negators *no* or *never* preceding the lexical verb (Dittmann 1992: 71; Bartens 2013: Feature 101; Chamorro-Díaz and Suárez-Gómez 2019: 141-142);
- copula deletion (Dittmann 1992: 67-69; Bartens 2013: Features 73-76).

16

The paper is structured as follows. Section 2 provides an overview of the existing literature on copula deletion. In Section 3 we describe our methodology, including the source of texts used for the analysis and a detailed description of the dependent variable. Section 4 contains a description of the corpus, according to different predictors which have been reported as relevant in the distribution of the copular verb BE. Section 5 describes the probabilistic analysis applied to the variables and provides an interpretation and discussion of the results. Finally, in Section 6 we offer a summary and the main conclusions.

## 2. Copula Deletion

The copula is one of the most widely studied linguistic variables in Atlantic Creoles, as reflected in the number of features analysed in *The Atlas of Pidgin and Creole Language Structures Online* (APiCS online) (Michaelis et al. 2013: Features 73-76). This is also reflected in the number of research papers published on copula variation in Atlantic Creoles since early work by Labov (1969, 1972a) on AAVE. Labov's studies were followed by similar studies of the same linguistic feature in other Atlantic Creoles, as detailed in Section 1.

Copula deletion refers to the omission of a form of the verb BE (in both its copulative or auxiliary uses) in contexts where its presence is required in Standard English. As is common in studies of copula variability, the term ‘copula’ is the conventional way of referring both to the copulative and to the auxiliary uses of the verb TO BE. This will be maintained in the present study, although reference to the specific forms will be made where necessary. Some accounts of the omission of the copular BE explain it as a result of imperfect second language learning (Winford 1998; McWhorter 2000; see Sharma and Rickford 2009 for a critique of the Imperfect Learning Hypothesis). However, a greater number of studies consider the omission of the copula to be due to the influence of the substrate (Holm 1976; Sharma and Rickford 2009).

From Labov (1969, 1972a) onwards, it has been observed that the omission of the copula (marked in the examples with the empty set symbol  $\emptyset$ ) is not a random phenomenon, but rather it shows an ordered patterning, which is largely repeated in most of the Atlantic Creoles analyzed. The general observation is that the distribution of the copula is determined by grammatical environment. Labov’s general findings in his studies are that the verb BE is almost systematically omitted with *gonna* (example (1)) and very frequently with *-ing* forms (example (2)) while it is more frequently used if it is followed by a noun phrase (example (3)); and that there is more variation when it is followed by an adjective phrase (example (4)) or a locative complement (example (5)). Labov’s studies led to parallel work in other creoles with similar results regarding the distribution of the copula.

- (1) She  $\emptyset$  gon tell him.
  - (2) She  $\emptyset$  walking.
  - (3) He  $\emptyset$  a man.
  - (4) She  $\emptyset$  happy.
  - (5) He  $\emptyset$  in the car.
- (Examples from Sharma and Rickford 2009: 53)

One of the first of these studies was Bickerton (1973) on Guyanese Creole, who found a similar pattern of copula distribution to that in AAVE. Weldon (2003) made a similar analysis of Gullah, also finding results consistent with AAVE in relation to the grammatical environment factor. Additionally, she observed that other linguistic factors, such as the grammatical person of the subject and the phonological environment surrounding the gap of the copula, are also strong predictors of variability. In a comparative work on Atlantic Creoles, Holm (1999: 99) found that with very few exceptions, most of these varieties require an expressed copula with noun phrases, which is frequently used but can be deleted before locatives, and which is frequently deleted with adjective phrases. Rickford also discovered similar quantitative patterns of copula absence between Jamaican Creole and AAVE: more

copula absence before adjectives than before locatives and noun phrases, and also the absence of a copula before *-ing* forms and *gonna* (1996: 358), rank-ordered as follows, from more deletion to less deletion: V + *ing* > V + *ed* > Adjective > Loc > NP.

The “following grammatical environment” is a factor widely repeated to account for copula deletion, but some other factors were also seen to be relevant, particularly the grammatical category of the subject, with noun phrases favoring the explicit use of BE. The tense of the copula was also relevant; the present tense deletes the copula more often than past tense. By contrast, other factors, such as grammatical person of the subject and text type, were not determining in the use of the copula in Rickford’s analysis.

A quantitative analysis of copula forms in Jamaican Creole is also provided by Deuber (2014), with data from the Jamaican subcorpus of the *International Corpus of English* (ICE-Jamaica). Unlike Rickford’s study (1996), whose data were based on a set of Jamaican Creole stories with basilectal traces, ICE-Jamaica contains speech from upper-mesolectal or acrolectal speakers. Therefore, the prestige of the creole (whether it is a basilect, mesolect or acrolect) also yields significant results. If Rickford’s and Deuber’s results are compared, we see that the closer the speech to the most prestigious variety, the higher the frequency of BE, especially in those grammatical environments which favor copula omission, as is the case with adjective phrases and the auxiliary use (Deuber 2014: 87). In Trinidadian Creole, also studied by Deuber (2014) based on data from the ICE-Trinidad and Tobago, a similar distribution is shown: BE is almost consistently used when followed by a Noun Phrase (NP), as opposed to adjective phrases, where zero copula is recorded; zero copula is also more frequently used with progressives, aligning with the characteristic pattern of copula absence described elsewhere for Caribbean Creoles. In Trinidad the only exception is the use of the copula with locative predicates, scoring higher than in other creoles, which Deuber argues can be attributed to the limitations of the sample (2014: 145).

Sharma and Rickford (2009) compare the phenomenon of copula absence in AAVE/Creole data and L2 English data from speakers of New Englishes (Indian English, South African Indian English and Singapore English) and learners of English as a foreign language (more specifically, Spanish learners). Their results show a different patterning between the different data sets, which leads them to conclude that the ‘imperfect learner hypothesis’ (Winford 1998) as a possible justification for the omission of the copula cannot be supported. Sharma and Rickford’s results provide further evidence for the substrate hypothesis, proposing as a likely source the substrate influence of West African languages, which would also justify the shared patterning in most Caribbean Creoles. The fact that copula absence is more frequently found in the verbal environment, i.e. BE as an auxiliary

verb, irrespective of the data set, should not be taken as very strong evidence for the imperfect learning hypothesis, since:

this may be attributed in part to the perceived redundancy, on the part of the learner, in using the auxiliary with a verbal predicate. Auxiliary uses of *be* with *V-ing* and *gonna* involve verbal content or inflection (including suppletion of *is* and *are*) at a minimum of two points in the clause —the auxiliary and the progressive verb— whereas copular sentences with non-verbal predicates require morphological inflection at only one point. (Sharma and Rickford 2009: 84-85)

Ramírez-Cruz (2017) also includes an analysis of copula variability. The author only analyzes real examples of copulative structures, excluding the use of BE as an auxiliary verb, and bases his results on production and translation tasks designed specifically for the purpose. Although both the object of study and the type of data analyzed are different from the study presented here, he also concludes that variation in the copula choice is consistent with what has been found in other creoles. Predicate type also becomes a significant predictor of variation: a following NP frequently includes the presence of BE (243), there is more competition between BE and copula absence with a following adjective phrase (244), and there is inter-speaker variation in the case of locatives (245).

19

The phenomenon of copula deletion in SAC has only been mentioned briefly in descriptive catalogues (Dittmann 1992: 67-69; Bartens 2013: Features 73-76), and to the best of our knowledge has never been studied in detail with real data. Inspired by earlier studies of copula deletion in different Atlantic Creoles, this variable has been selected because, as seen in this section, it is one of the most widely researched variables in these creoles and therefore allows for comparisons between them. This also allows us to place SAC on the map of copula variability in Atlantic Creoles.

### 3. Methodology

The data for this study comes from a selection of texts available at the Instituto Caro y Cuervo in Bogotá (Colombia), a local institution linked to the Ministry of Culture specialised in promoting the local languages and literatures of Colombia, and keeping its local traditions alive.

The texts are recorded samples of natural speech, whose transcripts are publicly available for inspection and analysis. Unfortunately, we did not have access to the recordings, and thus we relied exclusively on the available transcripts. The texts are transcriptions of recordings of conversations from three local women, between 55 and 70 years old, considered to be permanent residents of the island of San Andrés,

and who had received a limited education. From the content of the transcribed recordings, it can be appreciated that they are strongly identified with the *raizal* community (see Section 1). We can consider them mesolectal speakers, according to the creole continuum, although on some occasions there are some basilectal features interspersed, as in the use of the aspectual marker of the progressive *de* in example (6):

- (6) Ai onli **de** len yo di dres yo gou daans wit  
 I only PRG lend you the dress you go dance with  
 “I was only lending you the dress you are going to dance with”.<sup>4</sup>

The data-set contains texts of three different types: (i) local folk stories derived from the Afro-Anglo-Caribbean tradition (Anaansi Stories); (ii) personal stories of the informants about their life in the “old days”, and (iii) local practices, a dominant topic here being the therapeutic use of local plants. They are all topics with which the speakers were familiar, and with which they felt comfortable. This is an important issue in sociolinguistics, both as a means of obtaining an authentically vernacular variety of the language and also to minimise the Observer’s Paradox (Labov 1972b). The data-set amounts to c. 5000 words of transcriptions, with nothing excluded from the analysis. We are aware that the data-set is very small and the number of informants is limited, and the results derived from this study must remain tentative in nature.

The analysis of the copula was carried out manually, in order to select relevant examples and to discard invalid instances. We distinguished between the full form of the copula (example (7)) and copula deletion (example (8)), both in the copulative (example (8)) and in the auxiliary (example (7)) uses.

- (7) Ai **waz** livin in a ranch  
 I was living in a ranch  
 “I was living in a ranch”.
- (8) Ih Ø gud fa wen yu hav shuga in di blod  
 It good for when you have sugar in the blood  
 “It is good when you have sugar in the blood”.

We also included contracted forms (example (9)), but we finally decided to exclude them from the analysis because there were only four examples, three with the structure *dats wai* (‘that’s why’) and one introduced by the existential *derz* (‘there’s’), behaving as a sort of lexicalised structure which shows no variation in the sample:

- (9) **Dats** wai shi did nou so moch abaut di bosh  
 that’s why she did know so much about the bush  
 “That’s why she knew so much about medicinal plants”.



We also excluded cases of BE used as a modal verb (example (10)), because in these contexts BE is systematically used:

- (10) Wen **waz** tu stap grain, hi kudn stap ih!  
 when was to stop grind he couldn't stop it  
 "When he had to stop (the mill) grinding, he couldn't stop it".

All examples were entered into a SPSS database, and were coded for:

- (i) DEPENDENT VARIABLE: presence (1) vs absence (0) of the verb BE;
- (ii) PREDICATE TYPE, which identifies the contexts in which BE occurs: clause (0), Noun Phrase (NP) (1), Adjective Phrase (AdjPh) (2), adverbial (which includes both locative and temporal constituents) (3), *-ing* form (4), *gwan* ('gonna') (5), and past participle (PPLE) (6);
- (iii) SUBJECT CATEGORY: omitted (0), NP (1), personal pronoun (2), and other pronouns (3);
- (iv) SUBJECT GRAMMATICAL PERSON: first (1), second (2) and third (3);
- (v) TENSE: present (1) and past (2).

The next section is devoted to an investigation of the variable contexts and the factors that govern the variation of the variable. Since different variables were analyzed, the individual searches are detailed in the analysis of each variable.

## 4. Data Description

### 4.1. General Overview of the Data

Table 1 below provides the raw numbers and percentages of tokens showing variation in the use of the copula, either present or absent:

Copula variability	Tokens and frequency
Presence	86 (57.3 %)
Absence	64 (42.7 %)
TOTAL	150

Table 1. Overall distribution of copula variability in SAC

The analysis yielded a total of 150 examples of copula deleted forms (example (8)) and copula present forms (example (7)). Table 1 confirms that the copula BE represents a clear case of morphosyntactic variation in SAC; although BE is present more frequently in the relevant contexts, a rate of absence of almost 43% in the examples confirms that it can be regarded as a case of language variation.

#### 4.2. Contextual Factors: Grammatical Environment, Grammatical Category of the Subject, Grammatical Person of the Subject and Tense

Grammatical environment is the most pertinent factor in accounting for the distribution of BE, and has been studied repeatedly in previous research. In order to circumscribe the variable and to identify variants, two steps were followed: (1) we reviewed previous analyses (see the relevant references in Section 2) and the list of related features provided in APiCS online; (2) we complemented the list of variants through a careful reading of the texts, which allowed us to identify variants not previously mentioned in the literature. The complete set of variants included in the analysis is listed below; variants (i)-(v) were drawn from previous studies, and variants (vi)-(vii) arose from an analysis of the data-set itself.<sup>5</sup>

(i) Predicative noun phrases (Michaelis et al. 2013: Feature 73, Feature 76), as illustrated in example (11), which shows both a case of copula deletion and copula presence:

(11) Dis wan **iz** klat a mai dres, dis wan Ø di riil kola a mai dres  
 this one is cloth of my dress this one the real color of my dress  
 “This is the cloth of my dress; this is the original color of my dress”.

(ii) Predicative adjectives (Michaelis et al. 2013: Feature 74), as in example (12) (copula deletion) and example (13) (explicit copula):

(12) Maibi di presha Ø hai  
 maybe the pressure high  
 “Maybe the (blood) pressure is high”.

(13) Dis **iz** veri dilishos  
 this is very delicious  
 “This is very delicious”.

(iii) Predicative locative phrases (Michaelis et al.: Feature 75, Feature 76) (examples (14) and (15)):

(14) Dei kil aut al di monki **waz** in di haus  
 they killed out all the monkeys was in the house  
 “They killed all the monkeys which were in the house”.

(15) Two a dem dai, som a dem Ø in San Andres rait nau  
 two of them died some of them in San Andres right now  
 “Two of them (sons) died, some of them are in San Andrés right now”.

(iv) Progressive construction: BE + *-ing* (Sharma and Rickford 2009) (examples (16) and (17)):

(16) Wen yu get a gud kot, yu Ø **bliidin** a lat an yu kyaan get ih stap  
 when you get a gut cut you bleeding a lot and you cannot get it stop  
 “When you cut yourself, you bleed a lot and you cannot stop it (the hemorrhage)”.

## Copula Deletion in San Andresan Creole

- (17) Wan sekan **waz warkin** along wit him, sou him tiich di wan sekan  
 one second was working along with him, so him teach the one second  
 “One assistant was working with him, so he taught the assistant (all he knew)”.
- (v) *Gwain* (‘going to’, ‘gonna’) as a marker of future (Sharma and Rickford 2009), as in example (18):
- (18) Ai Ø **gwain** put yu rait hir in di bed  
 I going-to put you right here in the bed  
 “I am going to put you right here in bed”.
- (vi) Passive constructions, as in examples (19) and (20):
- (19) Ai Ø **baan** in di 1900’s bot shi Ø baan in di 1800’s  
 I born in the 90s but she born in the 80s  
 “I was born in the 90s, but she was born in the 80s”.
- (20) Ai **waz baarn** in San Andres  
 I was born in San Andres  
 “I was born in San Andrés”.
- (vii) Predicative structures in which the copula is followed by a clause, as in examples (21) and (22):
- (21) Di almon liif Ø fa<sup>6</sup> wen yu hav “swing hed”  
 the almond leaf for when you have swing head  
 “Almond leaves are used when you feel dizzy”.
- (22) Shi **waz** hu yus tu bail di bush  
 She was who used to boil the bush  
 “She was the one who would boil the (medicine) plants”.

The distribution of all these is set out in Table 2:

	Absence	Presence	TOTAL
<i>Gwain</i> (‘gonna’)	13 (100%)	-	13
Passive (BE + pple)	9 (69.2%)	4 (30.8%)	13
AdjPh	26 (59.1%)	18 (40.9%)	44
Clause (CL)	5 (38.5%)	8 (61.5%)	13
Progressive (BE + <i>-ing</i> )	4 (36.4%)	7 (63.6%)	11
Adverbial	5 (20%)	20 (80%)	25
NP	2 (6.5%)	29 (93.5%)	31
TOTAL	64	86	150

Table 2. Distribution of copula variability according to grammatical environment

The data in Table 2 reflect the distribution of copula variability in the seven relevant contexts. As can be seen, with the exception of *gwain* ‘gonna’, in all these contexts there is alternation between deletion and the explicit use of the copula, with different frequencies, as expected. If we rank-order them from the lowest to the highest frequency of use of BE, we see the following hierarchy:

(23) *gwain* > BE + pple > AdjPh > CL > BE -*ing* > Adverbial > NP

Another linguistic predictor of copula variability is the grammatical category of the subject (Labov 1972a; Rickford 1996; Weldon 2003). The variants distinguished here are the following:

- (i) Noun Phrase (see examples (17) and (21) above, and (24) below);
- (ii) Personal Pronoun (examples (16), (18), (19), (20) and (22) above);
- (iii) Other pronouns, which include demonstrative pronouns (example (13)) or existential pronouns (example (25) below) (see Rickford 1996 for a similar classification):

(24) **Di tingz** wazn rili ekspensiv bot doz deiz a peso an fifty sens  
 The things weren’t really expensive but those days a peso and fifty cents  
 kud bai three yaad a klat  
 could buy three yards of cloth  
 “Things weren’t really expensive, but in those days a peso and fifty cents  
 could buy three yards of cloth”.

(25) Wan die him luking **der** waz no salt  
 one day him looking there was no salt  
 “One day, he (the assistant) saw there was no salt”.

- (iv) Omitted subject (mostly *it*) (example (26)):

(26) yu kyan evn bail di hol plaant. **Iz** gud fa ches kol  
 you can even boil the whole plant is good for chest cold  
 “You can even boil the whole plant. It is good for chest colds”.

Table 3 presents the results from this analysis:

	Absence	Presence	TOTAL
NP	19 (32.8%)	39 (67.2%)	58 (38.7%)
Personal pronouns	35 (57.4%)	26 (42.6%)	61 (40.7%)
Other pronouns ( <i>der, dis, dat</i> )	9 (39.1%)	14 (60.9%)	23 (15.3%)
Omitted	1 (12.5%)	7 (87.5%)	8 (5.3%)
TOTAL	64	86	150

Table 3. Distribution of copula variation according to subject category

Table 3 shows that the category of the subject is a relevant variable in the use of the copula in SAC. As expected, the most frequent subjects are those realised by personal pronouns (40.7%) and noun phrases (38.7%), which together amount to almost 80% of the tokens. Noun phrases favor the use of explicit BE, as opposed to personal pronouns, which are more frequently used with copula deletion, as is the case of other pronouns, such as the demonstratives *dis* and *dat* and the existential *der*.<sup>7</sup> Finally, omitted pronouns tend to use the full form.

A third linguistically relevant variable in the description of copula variability is the grammatical person of the subject (Rickford 1996; Weldon 2003). In the database, both singular and plural forms were distinguished, but this distinction of grammatical person in terms of number was discarded because the tendencies in the singular and the plural were very similar. Additionally, it is not always easy to determine whether the grammatical subject is singular or plural, as in example (21) above, where *di almon liif* ‘almond leaves/the almond leaf’ is clearly a 3<sup>rd</sup> person subject, yet it is not possible to determine whether it is singular or plural, since unmarked plural nouns exist in SAC (Dittmann 1992: 75; Bartens 2013: Feature 22; Chamorro-Díaz and Suárez-Gómez 2019) and the form of the definite article *di* has been found both with grammatically singular (example (27)) and plural nouns (example (24)) (see also Dittmann 1992: 74; Chamorro-Díaz and Suárez-Gómez 2019: 140).

- (27) Maibi **di** presha  $\emptyset$  hai  
 Maybe the pressure high  
 ‘Maybe the pressure is high’.

Therefore, in order to avoid very low numbers in some of the cells, and even empty cells, we decided to conflate the distinction of grammatical number. In the analysis, the following variants were distinguished:

- (i) 1<sup>st</sup> person (singular and plural), as in examples (18), (19) and (20) above;
- (ii) 2<sup>nd</sup> person (singular and plural) (see example (26) above);
- (iii) 3<sup>rd</sup> person (singular and plural) (see examples (22), (24), (26) or (27) above).

Table 4 sets out the results of the use of the verb BE according to grammatical person:

	Absence	Presence	TOTAL
1st person	11 (61.1%)	7 (38.9%)	18
2nd person	10 (100%)	-	10
3rd person	43 (35.2%)	79 (64.8%)	122
TOTAL	64	86	150

Table 4. Distribution of copula variation according to subject person

The results from Table 4 show different distributions according to the grammatical person of the subject, as has also been shown by Weldon (2003) for Gullah, but not confirmed by Rickford (1996) for Jamaican Creole. The 2<sup>nd</sup> person invariably deletes the copula, as opposed to the 1<sup>st</sup> and 3<sup>rd</sup> persons, which show variation and opposing tendencies. In the 1<sup>st</sup> person, deletion is favored, as opposed to the 3<sup>rd</sup> person, the most frequent grammatical person (it comprises both pronominal forms and noun phrases) which resorts more frequently to the presence of the copula.

The last structural variable to account for the variation in the use of BE is tense. Previous studies circumscribe the context of variation to present tense cases (Weldon 2003); other studies (e.g. Rickford 1996) include present and past tense forms and treat tense as an independent variable. In our analysis we decided to include all forms, irrespective of tense, and classify them in the database. Two variants were distinguished:

- (i) present tense forms, illustrated by example (26);
- (ii) past tense forms, as in example (25).

From Table 5 we can confirm the usefulness of including past tense examples in the global count because they show variation. These results agree with Rickford's results for Jamaican Creole, in that present tense forms favor deletion over past tense forms.

26

	Absence	Presence	TOTAL
Present	28 (48.3%)	30 (51.7%)	58
Past	36 (39.1%)	56 (60.9%)	92
TOTAL	64	86	150

Table 5. Distribution of copula variation according to tense

## 5. Data Analysis

### 5.1. Analysis

A multivariate approach via a regression analysis was used to predict the presence/absence of BE in SAC adjusting for potential covariables. Analyses were performed using the Generalised Additive Mixed Models (GAMM) (Wood 2006a, 2006b). In comparison with classical linear regression models, GAMM models offer more flexibility, since they allow non-Gaussian responses to be considered and the effect of covariates to be estimated in a flexible manner. Following Wood (2006b), we conducted basic model checking plots for model fitting and no convergence problems were detected, which proves the validity and congruence of the model.

Thus, the GAMM regression model described was used considering a binomial distribution for the response (BE\_Form) and four categorical covariates (*Grammatical*

## Copula Deletion in San Andresan Creole

*Environment, Subject Category, Subject Grammatical Person, Tense*). Additionally, the covariable *Speaker* was also taken into account in order to check for random effects introduced by the different individuals, but this variable (*Speaker*) did not turn out to be statistically significant and was left out from the model. The distribution of copula deletion found in this study, therefore, cannot be attributed to idiolectal preferences.

Statistical analyses were performed using the *mgv*-package (Wood 2006b) of the open-source R statistical software. All significance levels were established at 0.05. The results obtained respective to the effect of the relevant covariates are summarised in Table 6 below:

Predictor	Estimate	Std. Error	z-value	P-value
Intercept	2.3195	1.4821	1.565	0.118
<b>Grammatical Environment</b> (Reference Level: <i>Clause</i> )				
<i>Noun Phrase</i>	2.2163	0.9445	2.347	<b>0.019</b>
<i>Adjective Phrase</i>	-0.674	0.708	-0.951	0.342
<i>Adverbial</i>	1.583	0.840	1.89	0.059
<i>Present Participle (-ing)</i>	1.317	1.041	1.265	0.201
<i>Gwain</i>	-2.8354	1.297	-2.187	<b>0.029</b>
<i>Past Participle</i>	-1.0148	0.987	-1.029	0.304
<b>Tense</b> (Reference Level: <i>Present</i> )				
<i>Past</i>	0.874	0.512	1.686	0.0919
<b>Subject Category</b> (Reference Level: <i>No subject</i> )				
<i>Noun Phrase</i>	1.655	1.037	-1.600	0.110
<i>Personal Pronouns</i>	-2.942	1.167	-2.521	<b>0.012</b>
<i>Other pronouns</i>	-1.6014	1.215	-1.318	0.187
<b>Grammatical person</b> (Reference Level: <i>First</i> )				
<i>Second</i>	-16.4187	631.490	-0.026	0.979
<i>Third</i>	-0.524	0.817	-0.642	0.521

Table 6. Summary of the estimated linear effects for the binomial regression model (*p*-values < 0.05 in bold type)

Of the variables analyzed, *Grammatical Environment* and *Subject Category* have a significant effect on the choice between absence and presence of the verb *BE*. Starting with *Grammatical Environment*, SAC seems to show a significantly

higher probability of using the verb *BE* when followed by an NP in comparison with the reference variant ‘Clause’; by contrast, the probability of using it if followed by the vernacular form marking the progressing *gwain* is significantly lower, also in comparison with the reference variant ‘Clause’.

The covariate *Subject Category* is also statistically significant. More specifically, the presence of the verb *BE* shows a lower probability if the subject is realised by a personal pronoun, in comparison with the reference variant ‘No Subject’.

In the regression model, the variables *Tense* and *Grammatical Person* do not have a significant effect on the absence/presence of the verb *BE*. Therefore, the tense of the verb *BE*, either present or past, and the grammatical person of the subject (first, second or third) do not seem to be responsible for the selection of the copular verb.

## 5.2. Discussion

The results of copula variability show notable similarities between SAC and other Caribbean Creoles. The results show that the use of the verb *BE* is largely conditioned by the grammatical environment in which it occurs, in agreement with Sharma and Rickford’s (2009) findings for AAVE and Rickford’s (1996) for Jamaican Creole. The predicate type which most frequently selects the use of *BE* is the NP, which seems to be the favorite locus of explicit *BE*; at the opposite end, the predicate type which opts for omitting the copula systematically is *gwain* (‘gonna’). These two variable contexts, as also found in previous literature, are contexts of systematic variation. The remaining environments tend towards one option or the other: adverbial and *-ing* favour the explicit use of the verb *BE*, as opposed past participle and adverbial, which opt more frequently for an absent copula. However, the non-significant results prevent us from reaching definite conclusions. These tendencies may be conditioned by other factors or by the limitations of the sample.

Another finding which agrees with previous research on copula variability in Atlantic Creoles is the grammatical category of the subject. As was the case in AAVE (Labov 1972a) and in Gullah (Weldon 2003), personal pronouns are significantly more frequently used with copula deletion. Other pronouns, such as the demonstratives *dis* and *dat* and the existential *der*, also disfavor the deletion of the copula, something which was also noted by Rickford (1996), although here the results are not significant, probably because of the scarcity of examples within this category. Regarding omitted subjects (‘No subject’), the tendency is to use the full form for the purposes of transparency, since it facilitates the processing of information, as in example (26), repeated here for convenience as (28), but we cannot confirm this result because the number of examples with omitted subjects is very low.



- (28) yu kyan evn bail di hol plaant. **Iz** gud fa ches kol  
you can even boil the whole plant is good for chest cold  
“You can even boil the whole plant. It is good for chest colds”.

Another predictor which was considered to play a role in the use of the copular verb BE was its tense. These results agree with Rickford’s (1996) results for Jamaican Creole, in that present tense forms favor deletion over past tense forms. However, these results are not significant at 0.05 level (*p-value* 0.09), and therefore we cannot consider this predictor as a determining variable in the distribution of copula variability. Further research on the effect of tense on copula variability is necessary with a larger sample.

The tendencies observed in Section 4.2 to demonstrate the hypothesis that the grammatical person of the subject conditions the selection of the copula cannot be confirmed, in agreement with Rickford (1996) for Jamaican Creole. The cause of variation in this realm is found in contextual factors, such as the grammatical category of the subject or the predicate type following the verb BE.

## 6. Conclusion

29

This paper presents a preliminary analysis of copula variability in San Andresan Creole, an English-based creole spoken in the Caribbean islands of San Andrés, Providencia and Santa Catalina, Colombia.

Research on SAC is very scarce, mostly because of the lack of available data, and the few existing studies tend to deal with the ethnolinguistic vitality of the creole and with the language policies that have been applied. There are also some linguistic descriptions, but these are mostly of a qualitative nature. Despite the paucity of the data available, we have attempted to provide a preliminary analysis of copula variability in SAC from a qualitative perspective, testing the resulting claims with various quantitative statistical tests to prove their validity and increase, to some extent, their credibility, and thus make up for the low number of data handled. Copula variability has been selected because it is one of the most widely researched variables in these creoles and therefore allows for comparisons between them. It refers to the alternation between the deletion and the explicit presence of the verb BE in copulative and in auxiliary contexts (progressive and passive). One of the main conclusions of studies on copula deletion is that the phenomenon is not random, and that there is rather a systematic pattern, governed mainly by predicate type. This patterning is very consistently shown in the Atlantic Creoles so far analyzed, and the results from our own analysis show that SAC seems to be no exception in this respect.

In order to test the validity of the results obtained from a qualitative data analysis, we studied the effect of contextual factors using a regression model with a binomial response distribution (see Table 6). The structural contextual factors here show that rates of copula presence are higher when followed by an NP and lower in contexts of *gwain* ('gonna'); there is more variability if the verb TO BE is followed by an adjective, an adverbial or *V-ing*. This distribution reflects the hierarchy (*gwain* > *V-ing* > Adj > Loc > NP) that applies to AAVE and Caribbean Creoles (Sharma and Rickford 2009), especially at the poles, thus supporting the substrate hypothesis proposed by these authors.

Another factor which potentially motivates the distribution of the verb TO BE in the corpus is the category of the grammatical subject, especially if the subject is realised by a personal pronoun, which disfavors the use of BE. As mentioned, the regression analysis confirmed that our corpus sample is amenable to the drawing of conclusions regarding grammatical environment and subject category. However, the influence of other factors such as grammatical person of the subject and tense cannot be confirmed as predictors of copula variability in SAC and will need to be tested on larger samples.

30 This study has shown that there is a complex interplay of structural factors interacting to shape grammatical variation. Copula variability does not seem to be a random phenomenon in SAC, but rather places SAC on the map of Atlantic Creoles since it shows the ordered patterning largely detected in most of the Atlantic Creoles analyzed to date. The conclusions shown, however, must remain tentative in nature since the sample used for the study is very small. Future analyses of a larger sample will shed light on this still underexplored English-based creole.

## Notes

---

1. We are grateful to Elena Seoane and to two anonymous reviewers for their valuable comments. Any errors remain our sole responsibility. For support with this study, our gratitude goes to the Spanish Ministry of Science and Innovation and the European Regional Development Fund (Grant No. FFI 2017-82162-P) and the Instituto Caro y Cuervo (Ministry of Culture, Colombia).

2. The term *raizal* comes from Spanish *raíz-* ('root') + *-al* (a derivative suffix meaning 'relating to') and makes reference to the indigenous community of San Andrés, Providencia and Santa Catalina.

3. The *raizal* community is identified by the following traits: their islander ancestry, the fact that they were born in the islands, the creole language, and their cultural practices (see Ramírez-Cruz 2017: 6-31 for detailed information on the *raizal* community).

4. All examples from SAC are glossed and translated. Translations are our own.

5. Due to the size of the sample and the limited number of speakers, we cannot discard the possibility that these unrecorded variants in previous studies

represent idiolectal tendencies. This is taken into account in section 5, which deals with the data analysis.

6. According to Washabaugh (1975), the particle *fi/fo/fa* is considered to be a complementiser in SAC when followed by a clause.

7. Findings for contracted forms, although scarce, seem to point in the same direction. The only contexts where contracted forms appear in our data-set are with demonstrative and existential pronouns (see Rickford 1996: 369) (e.g. **Dats wai shi did nou so moch about di bosh**); contracted forms, although reduced forms, will also represent an explicit copula.

## Works Cited

- ANDRADE ARBELÁEZ, Juliana. 2012. "Estudio sociolingüístico de San Andrés, isla: un aporte a la cultura sandresana". *Cuadernos del Caribe* 8: 42-55.
- BARTENS, Angela. 2013. "San Andres Creole English Structure Dataset". In Michaelis, Susanne Maria, Philippe Maurer, Marin Haspelmath and Magnus Huber (eds.): contribution n° 10.
- BICKERTON, Derek. 1971. "Inherent Variability and Variable Rules". *Foundations of Language* 7: 457-492.
- BICKERTON, Derek. 1973. "The Nature of a Creole Continuum". *Language* 49: 640-669.
- BOWIE, Pauline and Marcia DITTMANN. 2007. "El proyecto piloto trilingüe de la Universidad Cristiana de San Andrés en las escuelas Emanuel Bautista y Bautista Central en la isla de San Andrés". In de Truscott de Mejía, Anne-Marie and Sol Colmenares (eds.) *Bialfabetismo: lectura y escritura en dos lenguas de Colombia*. Cali: Universidad del Valle: 67-87.
- CHAMORRO-DÍAZ, Margarita M. and Cristina SUÁREZ-GÓMEZ. 2019. "Language Contact in Colombia: A Pilot Study of Criollo Sanandresano". *Athens Journal of Philology* 6 (2): 131-146.
- DEUBER, Dagmar. 2014. *English in the Caribbean. Variation, Style and Standards in Jamaica and Trinidad*. Cambridge: Cambridge U.P.
- DITTMANN, Marcia. 1992. *El criollo sanandresano*. Cali: Universidad del Valle.
- GARCÍA LEÓN, Javier and David GARCÍA LEÓN. 2012. "Políticas lingüísticas en Colombia: tensiones entre políticas para lenguas mayoritarias y lenguas minoritarias". *Boletín de Filología* 47 (2): 47-70.
- GUERRERO, Carmen Helena. 2008. "Bilingual Colombia: What does it Mean to Be Bilingual within the Framework of the National Plan of Bilingualism?". *PROFILE* 10: 27-45.
- HOLM, John. 1976. "Copula Variability on the Afro-American Continuum". Paper presented to the Society for Caribbean Linguistics, University of Guyana, Georgetown.
- HOLM, John. 1999. "Copula Patterns in Atlantic and non-Atlantic Creoles". In Rickford, John R. and Suzanne Romaine (eds.) *Creole Genesis, Attitudes and Discourse*. Amsterdam: John Benjamins: 97-119.
- HOOKE, Yolanda Yolette, Beatriz ROBINSON and Alberto ABOUCHAAR. 2002. "Estudio lingüístico para la implementación del programa de educación bilingüe en el municipio de Providencia y Santa Catalina". *Cuadernos del Caribe* 2 (3): 60-101.
- LABOV, William. 1969. "Contraction, Deletion and Inherent Variability of the English Copula". *Language* 45: 715-762.
- LABOV, William. 1972a. *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- LABOV, William. 1972b. *Sociolinguistic Patterns*. Philadelphia: Philadelphia U.P.

- McWHORTER, John. 2000. "Strange Bedfellows. Recovering the Origins of Black English" *Diachronica* 17: 389-432.
- MICHAELIS, Susanne Maria, Philippe MAURER, Martin HASPELMATH and Magnus HUBER (eds.) 2013. *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info>>. Accessed March 25, 2019.
- MORREN, Ronald E. 2001. "Creole-based Trilingual Education in the Caribbean Archipelago of San Andres, Providence and Santa Catalina" *Journal of Multilingual and Multicultural Development* 22 (3): 227-241.
- MOYA-CHAVES, Deyanira. 2010. "Situación sociolingüística de la lengua creole de San Andrés: el caso de San Luis" BA Thesis, Universidad Nacional de Colombia.
- O'FLYNN DE CHAVES, Carol. 1990. *Tiempo, aspecto y modalidad en el criollo sanandresano*. Bogotá: Universidad de los Andes.
- O'FLYNN DE CHAVES, Carol. 2002. "Una descripción lingüística del criollo de San Andrés" *Cuadernos del Caribe* 3: 19-22.
- RAMÍREZ-CRUZ, Héctor. 2017. *Ethnolinguistic Vitality in a Creole Ecology: San Andrés and Providencia*. PhD Dissertation, University of Pittsburg.
- RICKFORD, John R. 1996. "Copula Variability in Jamaican Creole and African American Vernacular English: A reanalysis of DeCamp's texts". In Guy, Gregory R., Crawford Feagin, Deborah Schiffrin and John Baugh (eds.) *Towards Social Science of Language. Papers in Honor of William Labov. Vol. 1: Variation and Change in Language and Society*. Amsterdam: John Benjamins: 357-372.
- RICKFORD, John R. 1998. "The Creole Origins of African-American Vernacular English: Evidence from Copula Absence". In Mufwene, Salikoko S., John R. Rickford, Guy Bailey and John Baugh (eds.) *African American English*. London: Routledge.
- SANMIGUEL, Raquel. 2007. "El bilingüismo en el Archipiélago de San Andrés, Providencia y Santa Catalina". *Revista Internacional Magisterio* 25: 28-31.
- SHARMA, Devyani and John R. RICKFORD. 2009. "AAVE Creole Copula Absence: A Critique of Imperfect Learning Hypothesis". *Journal of Pidgin and Creole Languages* 24: 53-90.
- WASHABAUGH, William. 1975. "On the Development of Complementizers in Creolization". *Working Papers in Language Universals* 17: 109-140.
- WELDON, Tracy. 2003. "Copula Variability in Gullah" *Language Variation and Change* 15: 37-72.
- WINFORD, Donald. 1998. "On the Origins of African American Vernacular English. A Creolist Perspective. Part II: Linguistic Features". *Diachronica* 15: 99-154.
- WOOD, Simon N. 2006a. "Low Rank Scale Invariant Tensor Product Smooths for Generalized Additive Mixed Models". *Biometrics* 62: 1025-1036.
- WOOD, Simon N. 2006b. *Generalized Additive Models: An Introduction with R*. New York: Chapman and Hall/CRC Press.

Received: 1 April 2019  
Accepted: 27 March 2020