Chiara Truppi

Chapter 5
Linguistic reconstruction and creole emergence: The case of Upper Guinea creoles

Abstract: The present chapter aims to discuss a combined methodology for linguistic reconstruction and to participate in the debate on the mechanisms involved in creole emergence. In particular, we will reconstruct the emergence of the paradigm of copulas (copularisation) in a group of Portuguese-related Atlantic creoles, i.e., Upper Guinea creoles, consisting of Capeverdean, (Bissau-)Guinean, and Casamancese. These creoles are said to share a common ancestor, i.e., a proto-creole. The reconstruction of the copular system of the proto-creole will be carried out on the basis of the comparison of synchronic data from these languages, also taking into account data from 19th-century sources. Insights from research on second language acquisition, grammaticalisation studies, and substrate, adstrate and areal influences will integrate our methodology. In doing so, this chapter will propose a combined strategy for the reconstruction of the proto-creole and will contribute to the debate on Upper Guinea creole emergence and, more broadly, language emergence.

Keywords: Portuguese-related creoles, proto-creole, copulas, grammaticalisation, L2 acquisition, contributing languages

1 Introduction

This chapter[1] focuses on a group of Portuguese-related creoles, i.e., Upper Guinea creoles (UGCs), consisting of Capeverdean (CV) and its varieties, (Bissau-)Guinean

1 This study was presented at the conference *New Issues in Language Contact Studies* – NILCS (L'Aquila, May 27–29, 2021), whose audience I wish to thank. I am grateful to Dominika Swolkien and Nicolas Quint for their precious comments on this work. Also thanks to Fernando Brissos for his help with historical and phonological questions on Portuguese, and to Barbara Hans-Bianchi, Barbara Vogt, and Valeria Lovato for further suggestions of improvement. This study was funded by the *Fundação para a Ciência e a Tecnologia* (SFRH/BPD/118401/2016) and the Centre of Linguistics of the University of Lisbon (UIDB/00214/2020).

Chiara Truppi, University of Lisbon, Portugal

https://doi.org/10.1515/9783111188348-005

(GB), Casamancese (CS), spoken in Cape Verde, Guinea-Bissau, and Ziguinchor – Lower Casamance region of Southern Senegal –, respectively.[2]

The African UGCs present a certain degree of intra-linguistic variation that, being the result of their linguistic history, will prove relevant to the reconstruction proposed in this chapter. As for continental UGCs, Wilson (1962) mentions the following historical varieties that correspond to the *praças* (fortified settlements): (i) the northern variety of Cacheu, including Ziguinchor – the latter, viz. CS, is a 17th-century offshoot of GB, possibly of the Cacheu variety; (ii) the central variety of Bissau; and (iii) the eastern variety of Geba. According to Wilson (1962) and Kihm (1994), the variety of Bissau has spread throughout the country and has (almost completely) replaced the other varieties, which are nearly extinct – some lexical and morpho-phonological variation is found in nowadays GB, though (see, e.g., Quint and Moreira 2019). With regard to CS, this language is said to be quite homogeneous (Biagui 2012). By contrast, CV has several diatopic varieties, i.e., the southern (Sotavento, including Santiago – ST – and Fogo – FG –, whose varieties will be taken into account in this study) and the northern (Barlavento). The southern islands of Santiago and Fogo were colonised during the first cycle of occupation (1460 to mid-17th century; see, e.g., Albuquerque and Madeira 1991) and the related varieties of ST and FG are said to be more ancient, with FG being an offshoot of ST.

There are few historical records of UGCs and that is the main reason why we need a combined methodology in order to reconstruct their emergence and to better understand their past. Most of the records focus on CV, for which some general descriptions and informative texts are available such as Costa and Duarte ([1886] 1967), Schuchardt (1887, 1888b), and Brito ([1887] 1967), among others. With regard to continental UGCs – formerly called Black Portuguese of Senegambia, Creole of the Province of Guinea or, in Barros' works, *Guineense* (Guinean) – historical records are mainly represented by Barros (1882, 1897–1899, 1900–1901), and Schuchardt (1888a). These records are quite recent, if we consider that the emergence of these languages – or, rather, of their common ancestor – presumably dates back to the second half of the 15th century. Nonetheless, based on a number of comparative studies on UGCs at both the lexical and structural level (see, e.g., Baptista, Mello, and Suzuki 2007; Quint and Moreira 2019), we are able to posit, with a very high probability, that the UGCs constitute a genetic unit and stem from a common ancestor, i.e., a proto-creole or proto-UGC. Moreover, the field of historical linguistics has provided us with the comparative method for linguistic reconstruc-

2 Papiamentu, a Spanish-related Caribbean creole spoken on the islands of Aruba, Bonaire, and Curaçao, is also considered as belonging to this group (see, e.g., Quint 2000): however, both lexical and grammatical studies place it as a clear outlier among the UGCs (see, e.g., Quint and Moreira 2019; Truppi 2021a).

tion (see, e.g., Anttila 1989): the comparison of languages allows to establish genetic relationships between them. Similarly, the comparison of related languages allows to reconstruct several aspects of their common ancestor.

With regard to languages for which hardly any historical records are available, such as Upper Guinea creoles, the reconstruction of their possible ancestor may benefit from insights from further perspectives such as grammaticalisation and L2 acquisition studies, along with the research on substrate, adstrate and/or areal influences: as we will discuss in §3, these fields have contributed relevant pieces of knowledge on the question of the emergence of pidgins and creoles. Drawing on these insights, this chapter wishes to propose a combined strategy for the reconstruction of proto-UGC, consisting of (i) the comparison of shared grammatical features, (ii) the study of grammaticalisation, (iii) insights from L2 acquisition studies, and (iv) possible substrate/adstrate influences and/or areal phenomena. The system of copulas in the UGCs will be the case study of this chapter, which will propose a reconstruction of their emergence (copularisation) in proto-UGC on the basis of the method outlined above. We will take into account both 19th-century data on UGCs and synchronic data from the varieties that lexical and grammatical studies have shown to be the closest: (i) ST – the oldest variety of CV, often considered as the closest to GB; (ii) FG – which may still contain some ancient ST items; (iii) GB – in the so-called variety of Bissau; (iv) CS – which may still present some ancient GB items (from the variety of Cacheu).

The chapter is organised as follows: §2 will provide an overview of UGCs as a genetic unit and will outline the main hypotheses regarding UGCs emergence. In §3, we will discuss a combined strategy for the reconstruction of proto-UGC. In §4, relevant data from the literature will be put in comparison: after a brief overview of copulas in the UGCs from 19th-century texts in §4.1, §4.2 will deal with synchronic data from the UGCs, while in §4.3 we will review some relevant data on copulas in their substrate and adstrate languages. In §5, a proposal for the reconstruction of copularisation in proto-UGC will be presented. Finally, §6 will summarise the main findings.

2 The emergence of Upper Guinea creoles: Genetic relationship and emergence

The UGCs constitute a genetic unit that most probably shares a common ancestor. With regard to CV and GB, a number of works have revealed striking similarities among them with regard to both their lexical inventories and grammati-

cal features.[3] As a consequence, it is widely accepted that these languages share a common history and genesis. From a lexical perspective, scholars such as Rougé (1999), Quint (2000), and Quint and Moreira (2019), among others, have studied the African-derived lexical items and contributed to show the degree of proximity of UGCs. More specifically, Rougé (1999) estimated that about 80% of the African-derived lexical items found in CV also occur in continental UGCs. Furthermore, it has been shown that most items come from Mandinka (Mande), followed by Wolof-derived lexical items (North Atlantic) and, to a lesser extent, Temne (Mel) and some other Atlantic languages (Quint and Moreira 2019).[4]

From a structural perspective, studies on shared grammatical features among UGCs are still scarce. The most relevant contribution is represented by Baptista, Mello, and Suzuki (2007), who compared a number of grammatical features of CV and GB and concluded that they share about 90% of their grammar. With respect to influences from African languages on UGC grammar, a number of works have contributed relevant pieces of knowledge (see, e.g., Quint 2000; Lang 2009; Kihm 2011; see Baptista 2004 as for similarities between copular clauses in CV and Wolof). More recently, Truppi (2019, 2021a) has shown striking similarities between the copular systems of CV, GB, and CS, along with possible substrate, adstrate and areal influences from languages such as Mandinka, Wolof, Temne, and a number of Atlantic languages spoken in the Upper Guinea linguistic area such as Fula, Byafada, Nyun (North), Manjaku, Mankanya, Pepel, Balanta, and Diola (Bak) – we will address these facts in more detail in §4.3. In conclusion, the most probable substrate languages for the UGCs are Mandinka and Wolof, while the role of Temne is still debated. The remaining Atlantic languages mentioned above are possible adstrate languages of continental UGCs.

The question of the genesis of UGCs has long been the subject of debate and a number of hypotheses have been proposed. We will provide here a brief overview of the main ones, namely the insular hypothesis, the continental one, and the hypoth-

3 Until Biagui's (2012) dissertation, CS was always treated together with GB, highlighting some lexical and morpho-phonological variation (see, e.g., Kihm 1994).

4 The role of Mandinka as for ST is still a matter of debate. Rougé (2019: 61–62) points at an apparent problem with Mandinka-derived lexical items present in GB but not in CV. Although we cannot discuss this question in more detail in this chapter, it is important to stress the importance of Mandinka in Upper Guinea at the time of the arrival of the Portuguese, i.e., mid-15th century: this was the time of the Mandinka Empire, the Mandinkas were the greatest merchants in the Senegambian region (Donelha 1625 in Mota and Hair 1977: 46), they did their trades with the Portuguese and their language was used as a lingua franca in Upper Guinea, along with the creole (Havik 2007). Based on these facts, a strong adstratic influence from Mandinka on continental UGCs is very likely and explains the presence of these Mandinka-derived items in GB: they may have entered GB on the continent and were not present in proto-UGC.

esis of *Língua de Preto* (LdP; lit. 'Language of the black') as ancestor of the UGCs. The insular hypothesis was proposed in the version of *Santiago birth* by Jacobs (2010): the common ancestor of UGCs emerged and became L1 of a first generation between the end of the 15th and the beginning of the 16th century on the island of Santiago and was transferred to the continent, i.e., to Cacheu, in present-day Guinea-Bissau, in the late 16th century. According to Biagui, Nunez, and Quint (to appear), at the basis of the proto-creole emerged in Santiago there would be a pidgin – either based on a variety spoken along the western coast of Africa or on LdP. The opposite perspective is adopted by the continental hypothesis (see, e.g., Rougé 1986): the common ancestor emerged on the continent, around the praças, which developed between the 16th and the 17th century; later, it was transferred to the archipelago of Cape Verde. By contrast, the hypothesis of *Língua de preto* by Kihm and Rougé (2013) attributes the origin of West African Portuguese-related creoles (UGCs and Gulf of Guinea creoles) to a basic variety (BV) of Portuguese spoken by West African slaves in Lisbon, whose reconstruction is based on the language that the characters of the 16th-century playwrights of authors such as Gil Vicente spoke acting the part of the slaves. Such BV would have been brought to the West African coast by slaves who had gained their freedom and returned to Africa, where the BV developed into a pidgin and later into UGCs and Gulf of Guinea creoles. All these hypotheses build on certain linguistic and historical evidence and/or reconstruction. Although UGC origin is still debated, historical and linguistic data – along with the fact that insularity represents a suitable condition for creole emergence given the break of the contact with the slaves' L1(s) – point towards an insular origin.

Social groups such as the *lançados* ("Portuguese marginals [. . .] with a criminal record or Jewish parentage"; Kihm 1994: 4) and the *grumetes* 'shipboys' are commonly assumed to have had a crucial role in the transfer of the emerging (or newly emerged) variety – i.e., proto-UGC – from its birth place and its subsequent spread. The former, who were traders based in Santiago and had commercial relationships with the coast of Guinea, would be responsible for the transfer of proto-UGC to the continent, where many of them remained because of more advantageous commercial opportunities (see, e.g., Baleno 1991). Here, they turned into *tangomaos* (so were called the lançados who established themselves on the continent and frequently integrated themselves among local populations) and founded *aldeias de tangomaos* 'tangomaos' villages' – one such village was located by the river of São Domingos, northern region of present-day Guinea Bissau (Donelha 1625 in Mota and Hair 1977: 108; see also Baleno 1991: 152). The grumetes – Christianised Africans who acted as intermediaries between the Portuguese and local African groups (Kihm 1994: 4) – and the *linguas* – African slaves serving the Portuguese as interpreters – also contributed to the spread of the proto-creole. According to Kihm (1994: 4), GB was fully formed by the beginning of the 17th century: his assumption

is based on the presence in GB of certain Portuguese items and specific phonological features which were common in 16th-century Portuguese but not afterwards.

3 Instruments for the reconstruction of the proto-creole

In order to reconstruct copularisation in proto-UGC, we will use a combined strategy consisting of (i) the comparative method for linguistic reconstruction, (ii) the study of grammaticalisation, (iii) insights from L2 acquisition studies, and (iv) the investigation of influences from the contributing languages. By doing so, the chapter aims to contribute novel insights into the emergence of this genetic unit. The need mainly to rely on the comparative method derives from the fact that historical records on continental UGCs are scarce, while only a few are available for CV. As a consequence, we will compare synchronic data from the literature on individual-level and stage-level copulas of the UGCs, also taking into account data from 19th-century texts on CV and GB.

With regard to studies on grammaticalisation, they have contributed relevant insights into creole emergence. In Heine and Kuteva's words (2005: 242), "contact-induced grammaticalisation is a major driving force in the grammatical development of pidgins and creoles". However, whether the trigger of grammaticalisation is language-internal or contact-induced is a matter of debate (see, e.g., Mufwene 2006), along with the question whether this process works differently in creoles with respect to non-creoles (see, e.g., Plag 2002; Bruyn 2009; Michaelis and Haspelmath 2020) or is just the same (see, e.g., Mufwene 2001, 2006; Heine 2003). We cannot discuss these questions in more detail here, but we refer the reader to Hans-Bianchi, Truppi, and Vogt (this volume; see the discussion below and the contributions by Rabanus and by Melissaropoulou, this volume, for cases of structural change where language contact is not the trigger). Given that grammaticalisation is a process that allows languages to derive new grammatical forms and categories and contributes, thus, to the shaping of a language (Heine 2003: 583–584), its study may be useful to the reconstruction of diachronic stages of a language (in our case, proto-UGC) and to understanding language change in settings characterised by intensive language contact such as the contexts where pidgins and creoles emerged. Ambiguity is a crucial factor, since it triggers grammaticalisation and, in a broader sense, contributes to language change (see, e.g., Heine 2003; see Truppi 2021b for the specific case of the copula *i* in GB). In the whole process, speakers play an active role by introducing innovations, triggering language change, and – (at least to a certain extent) consciously – reanalysing ambiguous structures. As we will see in

§5 with regard to the grammaticalisation of the individual-level copulas *e*/*i* in UGCs, our results support the perspective that this process in creoles works just the same as in non-creoles. The copulas *e*/*i* grammaticalised from the 3SG pronoun and this kind of grammaticalisation is well attested both in creoles and non-creoles – see, e.g., Sranan (Arends 1989; McWhorter 1997), Mandarin Chinese (Li and Thompson 1977), and Hebrew (Katz 1996).

Insights from studies on L2 acquisition and BVs are relevant to our reconstruction insofar as they have contributed relevant pieces of knowledge on creole emergence. BVs as discussed in Klein and Perdue (1997) represent a step in L2 acquisition that may further evolve and lead to more or less successful language acquisition or may fossilise. With regard to copulas and copular clauses, works on BVs and on foreigner talk varieties (see, e.g., Klein and Perdue 1997) have shown the possibility of a copulaless initial stage of an emerging variety. According to Klein and Perdue (1997: 16), "[i]n copular constructions, and for verbs which take only one argument, NP position depends on the way in which information is distributed over an utterance in context, that is, by pragmatic factors". Within the set of constraints they identified as guiding the organisation of an utterance, there are in particular two pragmatic ones that regulate information structure in BVs: (i) a constraint on what is new and what has already been said in discourse; (ii) a constraint on the structure of topic and focus. A topic (or a focus) may be indicated through intonation, clefting or by specific markers in certain languages; as a difference, in BVs this is done through word order (Klein and Perdue 1997). These findings bring further support to our claims with regard to the emergence of the individual-level copula in proto-UGC from 3SG pronoun used as resumptive to topic-comment structures (see §5). As we will see below, individual-level copular clauses with *e*/*i* (in CV and GB/CS, respectively) resemble a basic pattern of organisation of information structure: NP1 is the topic (or what is already given in the context), the copula *e*/*i* is its resumptive pronoun, while NP2 (or an adjective) represents the focus (or the new information; see Truppi 2021b).

Finally, UGCs contributor languages have also taken part in this process of copularisation. In order to discuss this, we will rely on data published in Truppi (2019, 2021a) and on her assessment with regard to substrate, adstrate, and possible areal influences. In order to check for possible influences from the lexifier, i.e., Portuguese, we will also discuss regional historical data from 15th–16th century Portuguese: as a matter of fact, actual lexifiers of creoles were nonstandard varieties (see, e.g., Mufwene 2006; also see Jacobs and Quint 2016 as to the relevance of data from 15th–16th century Portuguese with regard to Atlantic creoles). The results will show that, although the contributing languages may have influenced, to a certain extent, the copularisation process and the codification of predication in proto-UGC, they have not triggered either this structural change (or emergence of new categories in proto-UGC) or the grammaticalisation of the individual-level copula.

4 Copulas in the Upper Guinea creoles and their contributing languages

The high rate of structural proximity of the UGCs (§2) and the comparative method discussed as one of the instruments integrating our combined method for linguistic reconstruction (§3) will be at the base of the comparison of data from the UGCs as presented in this section.

With regard to our case study, i.e., copularisation in proto-UGC, recent works such as Truppi (2019, 2021a) will prove crucial to our discussion: they have shown the striking similarity of the system of copulas in the UGCs and the fact that they share three relevant features, i.e., (i) the predicational split, i.e. distinguished copulas for individual-level and stage-level (and locative) predication (see Stassen 2013), (ii) non-verbal predication, and (iii) the possibility of copulaless structures in individual-level predication.

Furthermore, Truppi (2021a) has contributed in distinguishing between substrate and adstrate influences in UGCs and shown the possibility of areal influences. More specifically, while the copular items clearly stem from Portuguese lexicon, Mandinka and Wolof may have influenced the emergence of (i) and (ii). As for (iii), i.e., copulaless predication, this represents in Mandinka a marginal phenomenon (Creissels, to appear); as for Wolof, it depends upon whether we consider the item *-a* as a copula or a focus marker (see footnote 14 below).

The same split encoding as in the UGCs and their substrate is also shared by Portuguese and Atlantic languages such as Fula, Nyun, and Diola. However, given the few lexical items that these Atlantic languages have contributed to the African lexical core of the UGCs (see Quint and Moreira 2019), it would be more accurate to consider them as adstrate of continental UGCs: indeed, GB and CS were always in contact with these local languages which may have further influenced their grammars. On the other hand, non-verbal and copulaless predication are features shared by all Atlantic and Mel languages considered in the study, i.e., Fula, Biafada, Nyun (North), Manjaku, Mankanya, Pepel, Balanta, Diola (Bak), and Temne (Mel).[5] Moreover, the features of non-verbal and copulaless predication in CV are found under certain conditions and are not as frequent as in continental UGCs. As we will see in §4.2, non-verbal predication with individual-level predicates is not fully maintained in CV: the copula *e* in ST and FG shows semi-verbal behaviour, while it

5 No relevant data on non-verbal predication in Biafada and Diola (Banjal) and copulaless predication in Mankanya were found in the literature (Truppi 2021a). This does not exclude the possibility of the existence of these features in these languages; on the contrary, this possibility is expected given that all other related languages have it.

is fully non-verbal in continental UGCs. Also, copulaless predication in CV varieties is allowed in limited contexts only. These facts suggest two further scenarios: (i) Atlantic languages – except the substrate Wolof – possibly had a role as adstrate of continental UGCs; (ii) the two features and their maintenance in the continental UGCs are due to areal influences. To better assess these possibilities, a more detailed study of Upper Guinea linguistic area is needed.

In what follows, we will discuss data from the UGCs. In §4.1, we will provide an overview of copulas in UGCs as found in late 19th-century texts on these languages. In §4.2 we will present a selection of relevant synchronic data from the UGCs and their contributing languages, as discussed in Truppi (2021a) in order to show the syntactic behaviour of copulas in these languages and their structural proximity. Also, we find useful to insert historical data from 15th–16th century Portuguese (see §3), that is when the proto-creole probably began to emerge.

4.1 Copulas from 19th-century records on Upper Guinea creoles

The copulas found in data from 19th-century records on UGCs do not diverge significantly from synchronic data. The predicational system was a split one in both insular and continental varieties. In the case of CV, data from Costa and Duarte (1967) and Brito (1967) show items such as the individual-level copula *é* in the present perfective and the stage-level copula (*i*)*stâ* (1a), *era* in the past (1b), and *sér* used with aspect markers or bare in the imperative and subjunctive mood (1c). The sentence in (1d) shows the non-verbal behaviour of the copula *é* given that the negation – preverbal in all UGCs – is post-copular (see §4.2). This is one of the classical diagnostic tests for detecting the non-verbal behaviour of copulas.[6]

(1) a. *É* (*a*)*li* *qu'el* *'stâ.*
COP here REL+3SG.NCL COP
'This is where he is'.
(adapted, Costa and Duarte 1967: 292)

b. *Ĩ* *era.*
1SG.CL COP.PST.IPFV
'I was'.
(adapted, Brito 1967: 372)

6 For further syntactic tests for the non-verbal behaviour of *e*/*i*, see, e.g., Ichinose (1993), Kihm (1994, 2007), Baptista (2002, 2004), and Truppi (2019, 2021a,b).

c. *Ĩ* *al-sér*.
1SG.CL A-COP
'I will be'.
(adapted, Brito 1967: 372)

d. *É* *ka* *di* *bu* *kõta*.
COP NEG of POSS.2SG count
'It is not your business'.
(adapted, Brito 1967: 399)

With regard to the continental UGCs, we found some relevant differences between the copulas from late 19th-century data (mainly from Barros 1897–1899; 1900–1901 and Schuchardt 1888a, the latter based on Barros' data) and the ones observed in present-day continental UGCs. First, we notice the presence of the individual-level copula *ê*, which is completely absent from synchronic data from both GB and CS. Also, we find the item *sedo* instead of *sedu/sedi* of present-day GB/CS. In particular, Barros (1897–1899: 289–290) presents the following items: *ê* (2a), *sedo* (2b) – bare or with aspect markers –, and *era* (2c). Finally, like (1d) above, the second instance of the copula *ê* in (2a) precedes the negation: this is evidence of non-verbal predication (see §4.2).

(2) a. *Ê* *bo?* *Nãu,* *ê* *ca* *mi!*
COP 2SG.NCL no COP NEG 1SG.NCL
'Was it you? No, it wasn't me'.
(Barros 1900–1901: 310)

b. *Lion* *co'* *onça* *é* *sedo* *comedor* *de* *carna.*
lion with jaguar 3PL.CL COP eater of meat
'The lion and the jaguar are flesh-eater'.
(Barros 1900–1901: 301)

c. *əm* *era-ba /* *əm* *sedo-ba.*
1SG.CL COP.PST.IPFV-PST / 1SG.CL COP-PST
'I was'.
(Barros 1897–1899: 289)

It is not clear from the literature to which variety the data in (2a)–(2c) belong. Still, Barros (1897–1899: 180) points at some phonological differences between the variety of Cacheu, on the one hand, and of Bissau (and Geba), on the other: the vowels *e* and *o* would be realised as *i* and *u*, respectively, in the latter variety (as for phonological differences such as /e/ in Cacheu and /i/ in Bissau, also see Wilson 1962: 35). The presence of final *-o* instead of *-u*/*-i* in items such as the copula *sedo* (cf. *sedu*/*sedi* in GB/CS in §4.2) makes it very likely that Barros used data from Cacheu.

4.2 Portuguese-related Upper Guinea creoles

Individual-level copulas are CV *e* and GB/CS *i*. They are used with nominal and adjectival predicates and occur in perfective contexts, resembling the basic distinction in UGCs between perfective aspect (expressed by the bare verb) and imperfective aspect (expressed through preverbal markers). *E*/*i* never occur in the presence of imperfective aspect markers (see examples 3a, 3b, and 3c). In the continental UGCs and in FG, the individual-level copula is sometimes omitted: these varieties allow copulaless predication (see 3b and 3c).[7]

(3) a. Fogo
Zingi e kel lata artu, e ka sima
Zingi COP DEM.DIST can tall COP NEG like
es bardi di oji [. . .].
DEM.PROX bucket of today
'The *zingi* is a tall can, it is not like present-day buckets [. . .]'.
(adapted, Moreira 2020: 154)

b. Fogo
Si povu pergunta=bu, ka fra karru di bo.
if people ask=2SG.CL NEG say car of 2SG.NCL
'If people ask, do not tell that the car is yours'.
(adapted, Moreira 2020: 204)

c. Guinean
Kil omi-s (*i*) *piskadur(-is).*
DEM.DIST man-PL COP fisherman(-PL)
'Those men are fishermen'.
(adapted, Truppi 2019: 93)

In the case of negated copular clauses, the negation *ka* – which is preverbal in all the varieties considered here – occurs after the copulas *e*/*i*. While the negation in the continental UGCs is always post-copular (4a), *e* in ST shows a semi-verbal behaviour insofar as the negation may also be pre-copular (4b). As a difference, the copula *e* of FG seems to have a behaviour closer to continental UGCs: it always follows the negation *ka* and it can be omitted (see 3a and 3b above, respectively).

7 According to Rougé (2013), in present-day CS, the copula is used only with nouns, while adjectives are still used without a copula.

(4) a. Casamancese
Abo i ka fiju di labrador.
2SG.NCL COP NEG son of farmer
'You are not a farmer's son'.
(adapted, Biagui 2012: 182)

b. Santiago
[. . .] *es kusa ka e dretu.*
DEM.PROX thing NEG COP good
[. . .] this thing is not good'.
(adapted, Veiga 2000: 157 in Baptista 2004: 101)

In the presence of aspect markers, the copulas CV *ser*, GB *sedu*, and CS *sedi* are selected (5a and 5b). GB/CS *sedu*/*sedi* may be used as bare verbs with a resultative interpretation (5c).[8] In the case of CV, the use of bare *ser* is possible in negated contexts such as the negative imperative in (5d).[9]

(5) a. Santiago
Joao ka ta ser profesor.
João NEG A COP professor
'João will not be a professor'.
(adapted, Baptista 2002: 106)

b. Guinean
Bu na sedu pursor.
2SG.CL A COP teacher
'You will be a teacher'.
(adapted, Ichinose 1993: 24, in Truppi 2019: 95)

c. Casamancese
Pidru sedi boŋ soldadi.
Pidru COP good soldier
'Pidru has become a good soldier'.
(adapted, Biagui 2012: 188)

8 The resultative meaning is not the only value associated with bare *sedu* in GB (and possibly with bare *sedi* in CS). However, the semantics of *sedu* in GB needs to be better studied (cf. Ichinose 1993; Kihm 1994).

9 Thanks to Dominika Swolkien and Nicolas Quint for making me aware of this. Interestingly, bare *sedu* is compatible with negative imperatives also in GB: *Ka bu sedu kabalidu*, NEG 2SG.CL COP bad, 'Don't be bad' (elicited data). See also footnote 10 for bare *ser* in ST rural varieties.

d. Capeverdean
Ka *bu* *ser* *fastentu!*
NEG 2SG.CL COP annoying
'Don't be annoying!'
(Dominika Swolkien, p.c.)

In the case of past tense, the situation is more complex. All UGCs have the suppletive forms *era* and *foi* (see 6a and 6b) from Portuguese 3SG imperfective *era* and perfective *foi* – with the exception of CS that has only *yera*.[10] In GB, *era* also occurs in the variants *yera* and, very seldom, *yara*. In continental UGCs, all copulas may be followed by the past marker *ba*(*ŋ*), while this is not possible in CV. In ST, *ba* may occur with *ser* either in combination with aspect markers, similarly to the other UGCs, or with complex VPs (counterfactual reading). In addition, *ser* in FG may be followed by the past marker *-ba* also in conjunction with aspect markers (6c).[11] In continental UGCs, further strategies are (i) *i* or its null counterpart + complement + *ba*(*ŋ*) – (6d), and (ii) *sedu*/*sedi* + *ba*(*ŋ*) – (6e).

(6) a. Santiago
Bu/bo *era* *trabadjadera.*
2SG.CL/2SG.NCL COP.PST.IPFV worker
'You were a hard worker'.
(adapted, Baptista 2002: 107)

b. Fogo
El *foi* *nos* *amigu.*
3SG.CL COP.PST.PFV POSS.1PL friend
'He was our friend'.
(adapted, Moreira 2020: 258)

c. Fogo
Si *n* *studa-ba,* *n* *ta* *ser-ba* *dotor.*
if 1SG.CL study-PST 1SG.CL A COP-PST doctor
'If I studied, I would be a doctor'.
(Moreira 2020: 206)

10 According to Quint (2012: 169), in rural ST *foi* is rarely uttered; the most usual expression of past perfective in this variety is the bare verb *ser*.
11 The past form *sere-ba* is also attested, while **sere* is not (Moreira 2020: 205).

d. Guinean
Abo (***i***) *bon alunu ba.*
2SG.NCL COP good student PST
'You were/had been a good student'.
(adapted, Truppi 2019: 95)

e. Casamancese
Pidru sedi baŋ boŋ soldadi.
Pidru COP PST good soldier
'Pidru had been a good soldier'.
(adapted, Biagui 2012: 189)

Stage-level predication and locative predication are expressed in UGCs through a specific copula: *sta* in ST, FG, and GB and *sa* in CS.[12] Interestingly, *sa* is also attested in FG. These forms most probably come from 3SG Portuguese stage-level and locative copula *está*.[13] The copulas *sta*/*sa* may occur with locatives (7a), adjectives (7b), and nouns (7c) in all UGCs, except GB, where the occurrence of *sta* is not allowed with nouns.

(7) a. Guinean
Si kuku sta dentru di ki-la.
POSS.3SG kernel COP inside of DEM.DIST-there
'The kernel is inside it [the fruit]'.
(Truppi 2019: 101)

b. Fogo
Kabu sta mau.
place COP bad
'Things are not going well (lit. The place is bad)'.
(adapted, Moreira 2020: 186)

c. Casamancese
Pidru sa piskador.
Pidru COP fisherman
'Pidru is a fisherman'.
(adapted, Biagui 2012: 191)

12 According to Rougé (2013: 203), *sta* is also present in CS as a variant of *sa*.

13 Quint (2015) argues that most Portuguese-derived ST verbs derive from infinitival forms. However, the frequency of occurrence of a specific form is one of the main factors that trigger the selection of a certain form in an emerging language (see, e.g., Mufwene 1996). If we consider that, among the forms of Portuguese *estar*, 3SG present indicative *está* is by far the most frequent (Clements 2022), then *está* is the most plausible candidate for UGC *s*(*t*)*a*.

Table 1 and 2 below (adapted from Truppi 2019: 104) provide a summary of the findings and allow better to appreciate the proximity of the predicational systems of the UGCs.

Table 1: Individual-level predication in UGCs.

	Copulas	Aspect marking	Tense marking (past)	Predicate type
ST	*e/ø*	no	no	N/ADJ
	ser	yes	yes	N/ADJ
	era/foi	no	no	N/ADJ
FG	*e/ø*	no	no	N/ADJ
	ser	yes	yes	N/ADJ
	era/foi	no	no	N/ADJ
GB	*i/ø*	no	yes	N/ADJ
	sedu	yes	yes	N/ADJ
	(y)era/yara/foi	no	yes	N/ADJ
CS	*i/ø*	no	yes	N/ADJ
	sedi	yes	yes	N/ADJ
	yera	no	yes	N/ADJ

Table 2: Stage-level predication in UGCs.

	Copulas	Aspect marking	Tense marking (past)	Predicate type
ST	*sta*	yes	yes	N/ADJ/LOC
FG	*sta/sá*	yes	yes	N/ADJ/LOC
GB	*sta*	yes	yes	ADJ/LOC
CS	*sá/(sta?)*	yes	yes	N/ADJ/LOC

4.3 Copulas in UGCs contributing languages

The lexifier Portuguese shares with the UGCs the predicative split-encoding. It uses the verbal copula *ser* for individual-level predication (8a) and *estar* for stage-level and locative predication (8b).

(8) a. *Ele é professor.*
'He is a teacher'.
b. *Ele está aqui.*
'He is here'.
(Truppi 2021a: 193)

If we take a look at Portuguese examples from 15th and 16th centuries, we can see that no relevant difference is visible in the choice of the copulas. In other words, the split between individual-level (9a) and stage-level (9b) predication was part of the system.

(9) a. *Esta* *he* *aquella* *que*
DEM.PROX.FEM COP.PRS.3SG DEM.DIST.FEM REL
aos *boos* *homees* *faz* *cair* *em* *pecado* [. . .].
to.DET.PL good.PL man.PL make fall in sin
'This is what makes the good men fall into sin [. . .]'.
(1430–1455, *Livro das Tres Vertudes*, CIPM)

b. [. . .] *com*(*m*)*o* *om*(*m*)*e* *q*(*ue*) *está* *en* *seu*
like man REL COP.PRS.3SG in POSS.3SG
p(*ro*)*pio* *aco*(*r*)*do.*
own agreement
' [. . .] like a man who is of sound mind (lit. like a man who is in his own agreement)'.
(1450, HGP044, CIPM)

With regard to possible influences from the substrate and adstrate languages, they may have had a certain role in the emergence of the copular system of proto-UGC. The substrate languages Mandinka and Wolof may have influenced the split encoding of the predicational system of proto-UGC, along with Portuguese. Furthermore, a number of Atlantic languages such as Fula, Nyun, and Diola, which also share the same encoding as the UGCs and their substrate languages, may have contributed either to the maintenance of this feature in the continental UGCs or to its setting in proto-UGC as an instance of areal influence. For convenience's sake, we re-propose here some of the examples discussed in Truppi (2021a). The sentences below represent individual-level and stage-level predication in Mandinka (10a and 10b), Wolof (10c and 10d), Nyun Gubëeher (10e) and Nyun Gujaxer (10f), respectively.[14]

(10) a. *À-té* *lè* *mú* *màns-ôo* *tí.*
3SG-EMPH FOC COP king-D POSTP
'He is the king'.
(adapted, Creissels, to appear, p. 24)

14 Notice that there is no agreement among scholars as to the status of Wolof items *-a*/*la,* alternatively described as copulas or as focus markers, among other definitions (see Truppi 2019, 2021a and references therein). That is why we did not attribute a label to these items in the glosses of (10c) and (12b). If we consider *-a* as a focus marker, the sentence in (10c) is a copulaless clause.

b. *Díndíŋ-ò bé búŋ-ò kónò.*
child-D COP house-D in
'The child is in the house'.
(adapted, Creissels, to appear, p. 24)

c. *Xale yi nàppkat-a.*
child D.PL D.PL fisherman-*a*
'The children are fishermen'.
(adapted, Torrence 2005: 226)

d. *Móódu mu-ng-i ca ja ba.*
Moodu 3SG-LOC.CL P market D
'Moodu is at the market'.
(Torrence 2005: 255)

e. *Me u-saw.*
1SG C-hunt
'I am a hunter'.
(adapted, Cobbinah 2013: 260)

f. *Agu Bisaw.*
COP Bissau
'He is in Bissau'.
(Wilson 2007: 90)

With regard to non-verbal and copulaless predication, all African contributing languages may have influenced the emergence of the predicative system of proto-UGC, since all share these features. The sentence above in (10e) is a copulaless clause in Nyun Gubëeher, while (11) represents a case of non-verbal predication through noun class agreement in Manjaku.

(11) *Ukam baneki.*
War last.year
'Last year (was the time of) the war'.
(adapted, Karlik 1972: 488)

With regard to the marking of past tense, the past marker *ba*(*ŋ*) in continental UGCs can co-occur in a copular clause with any copula (§4.2). Crucially, the same is true for UGC substrate and adstrate languages (Truppi 2019, 2021a). The past marker may occur in non-verbal and copulaless clauses in Mandinka (12a) and Wolof (12b). Furthermore, it can occur with nouns also in copulaless clauses in both North and Bak Atlantic languages in general (Cobbinah, p.c., in Truppi 2021a: 211); as an example, see the case of Fula Gombe (12c). This points towards possible influences from the African contributing languages in the marking of the past tense by inde-

pendent monosyllabic markers in peripheral positions, i.e., after a verb, a noun, an adjective, an adverbial or in sentence-final position (Truppi and Hagemeijer 2018; see also Truppi 2019).

(12) a. *Mandiŋk-oo-lu, wo-lu le mu nǔŋ jǎŋ*
Mandinga-D-PL DEM-PL FOC COP PST here
karammoo-lu ti.
marabout-D-PL OBL
'The Mandingas, it's they who were the marabouts here'.
(adapted, Creissels and Sambou 2013: 83)

b. *Sidi (l-)a woon.*
Sidi *la* PST
'It was Sidi'.
(adapted, Torrence 2005: 264)

c. *Ali ɗon(-no) ɗo'o.*
'Ali is (was) here'.
(Arnott 1970: 32)

The findings are summarised in Table 3.

Table 3: Possible influences from UGCs contributors (adapted from Truppi 2021a: 212).

	Split encoding	**Non-verbal predication**	**Copulaless predication**
Portuguese	yes	no	no
Mandinka	yes	yes	unknown
Wolof	yes	yes	unknown
Temne	no	yes	yes
Further Atlantic[15]	Fula, Nyun, Diola	all (unknown in Biafada and Diola)	all (unknown in Mankanya)

15 Under this label, we include the following Atlantic languages, as discussed in Truppi (2021a) and in §2 and §4.3: Fula, Byafada, Nyun (North), Manjaku, Mankanya, Pepel, Balanta, and Diola (Bak).

5 Reconstructing copularisation in the proto-creole

Based on the results of the comparison and on our combined methodology for linguistic reconstruction, in the present section we will provide a reconstruction of copularisation in proto-UGC. We argue that proto-UGC, at a very early stage of its emergence, was a copulaless language, at least with regard to individual-level predication. On the basis of the BV constraints discussed in §3 and of similar instances of copula emergence from pronominal forms in other languages, we propose that proto-UGC speakers derived a copula from 3SG clitic pronoun **e** as resumptive to topic-comment structures of the kind shown in Figure 1(a).[16]

(a) **(rapas) ele ø piskador*	>	(b) **rapas/el* e_{3SG} *piskador*	>	(c) *rapas/el* e_{COP} *piskador*
'(boy) he ø fisherman'	>	'boy/he he fisherman'	>	'boy/he is (a) fisherman'

Figure 1: From topic-comment structures to copular clauses in proto-UGC (adapted from Truppi 2021b: 19).[17]

The structure (b) in Figure 1 was ambiguous between a topic-comment and a predicative reading. This ambiguity has triggered the reanalysis of the pronoun into a copula and of the whole structure into a copular clause (c); this reanalysis was probably concomitant with the emergence of strong and weak pronouns in proto-UGC (Truppi 2021b: 19). The copulas *e*/*i* in CV/GB emerged through the grammaticalisation into copula of proto-UGC 3SG clitic pronoun **e*. The derivation of the pronominal forms is shown in Figure 2: the grammaticalisation into strong and weak pronouns in proto-UGC was most probably influenced by substrate languages, given that both Mandinka and Wolof have strong and weak pronouns. Finally, Figure 3 summarises the path of grammaticalisation of proto-UGC non-verbal copula from which CV *e* and GB/CS *i* derive.[18]

16 For a pronominal origin of GB *i*, see also Ichinose (1993) and Kihm (2007). With regard to ST copula *e*, Baptista (2002) also defends a pronominal origin, while Salas Barrena (2006) claims that the two items have a separate origin and converged later, probably due to substrate influence.

17 The proposed form in Truppi (2021b) was *piskadur* from the GB word for 'fisherman'. Since CS and CV display the item *piskador*, the most plausible form for proto-UGC is the latter. Thanks to Nicolas Quint for making me aware of this point.

18 Although in modern Portuguese the final segment /e/ in *ele* is typically realised either as /ə/ or deleted, the realisation of final /e/ was still a possibility in 15th–16th century Portuguese – and still is, in some dialectal varieties; also, this segment is raised to /i/ in some dialects of Portuguese (Fernando Brissos, p.c.).

Port. 3SG.MASC *ele* > proto-UGC **el-e* > CV 3SG.NCL *el* + 3SG.CL *e*
Port. 3SG.MASC *ele* > proto-UGC **el-e* > GB (> CS) 3SG.NCL *el* + 3SG.CL *i*

Figure 2: Emergence of 3SG pronouns in the UGCs (adapted from Truppi 2021b: 19, adapted there from Kihm 2007: 293).

proto-UGC $*\varnothing_{COP}$ > proto-UGC $*ele_{3SG}$ > proto-UGC $*el_{3SG.NCL} + e_{3SG.CL}$ > proto-UGC $*el_{3SG.NCL} + e_{COP}$

Figure 3: Emergence of non-verbal copula in proto-UGC (adapted from Truppi 2021b: 20, adapted there from Kihm 2007: 293).

As the reader may recall, in §4.1 we discussed Barros' (1897–1899; 1900–1901) data on GB – very likely, the variety of Cacheu: it included the copula *ê*.[19] This same item, however, does not appear in present-day continental UGCs. This could suggest that *ê* was the perfective individual-level copula of the 19th-century variety of Cacheu, which is nearly extinct nowadays. In contemporary GB and CS, we only have the form *i*: in the case of GB, this is explained by the fact that what we call GB is the Bissau variety which has spread throughout the country and almost completely replaced pre-existing varieties. In the case of CS, this is quite odd, if we consider that CS is a later offshoot of the variety of Cacheu. Based on these observations, the most plausible scenario is that the original copula in the Cacheu variety was *i* and that CS inherited this item. Later, *i* was replaced by *ê* in Cacheu (but not in Ziguinchor or in Bissau), probably due to CV influence. This hypothesis is supported by the fact that contacts between the islands and the coast of Guinea continued along the centuries due to commercial reasons (see, e.g., Albuquerque and Madeira 1991), in particular due to the presence of social groups such as the lançados, tangomaos, grumetes, and African slaves trained as interpreters (see §2). On the other hand, there was limited communication between the Portuguese settlements on the continent (see Wilson 1962: vii), probably until the War of Independence in Guinea-Bissau (1963–1974), when the creole was used as means of national unity. It is also worth noting that the 3SG clitic pronoun in Barros' data is *e* and not *i* as in present-day GB and CS. This could confirm either a certain influence from CV or alternatively a later development of initial *i* into *e* in the Cacheu variety – surely after the split of CS from the variety of Cacheu. Another perspective is that Cacheu maintained initial *e*

19 With regard to 19th–century data on UGCs, both Brito (1967) and Barros (1897–1899; 1900–1901) graphically distinguish between the copula and the 3SG subject clitic pronoun (ST/GB *é*/*ê* for the copula and ST/GB *ê*/*e*' for 3SG pronoun): this could suggest that, over time, the two items diverged (phonetically) and were not homophonous in late 19th–century UGCs (as for the case of CV, thanks to Dominika Swolkien for making me aware of this).

from proto-UGC like CV and it raised to *i* in Bissau and Ziguinchor due to the influence of the languages spoken by the local populations: to evaluate this hypothesis, we would need a more in-depth study of the morpho-phonological influences from African languages on continental UGCs.

Evidence of this kind of grammaticalisation, i.e., from a personal pronoun into a copula, are (i) the striking similarity of topic-comment structures and *i*-copular clauses in GB (Ichinose 1993; Truppi 2021b), (ii) the intimate link between topic-comment/given-new information and predicative structures (information structure), as discussed above with regard to BVs, and (iii) the fact that the same kind of copula emergence has already been documented in languages such as Mandarin Chinese, Hebrew, and Sranan, among others. As mentioned, this supports the view that grammaticalisation in creoles does not show relevant differences from non-creole languages. Moreover, this reconstruction also shows that grammaticalisation is a diachronic process that takes place over time and is visible, on the synchronic axis, in a sort of continuum represented by the possibility of both individual-level copulaless and copular expressions (as for the synchronic footprints of grammaticalisation, see Hans-Bianchi, Truppi, and Vogt, this volume). This also suggests that creole emergence is a gradual process (see, e.g., Mufwene 2001). Furthermore, ambiguity was a relevant driving force throughout the process: the interpretation of structures like (b) in Figure 1 was ambiguous between a copular and a topic-comment interpretation (see Truppi 2021b). Incidentally, we may sometimes find this same kind of ambiguity in *i*-copular clauses (at least, in GB). With regard to the timeline of this process in proto-UGC, we suggest that the grammaticalisation of the 3SG subject clitic pronoun into an individual-level non-verbal copula happened at a very early stage in proto-UGC formation. Moreover, Lang (2009) argues that ST verbal pair *ten*/*tene* 'have' – which expresses the distinction between individual-level vs. stage-level possession (Lang 2018) and is presumably an old feature of the language – emerged due to Wolof influence and was modelled on the distinction between individual-level and stage-level copulas *e* and *sta* in ST.[20] If so, this provides further evidence in favour of a very early grammaticalisation of copulas in proto-UGC.

With regard to the individual-level copulas *ser*/*sedu*/*sedi*, from Portuguese *ser,* it is interesting to notice that they are the only copular forms coming from an infinitival form, and not from 3SG present indicative. The presence of the past form *sere-ba* in FG suggests the existence of a previous form **sere* which would date back to proto-UGC (Moreira 2020: 205). A further change would have taken place on the continent: /r/ turned into /d/ and /u/ or /i/ were inserted through deltacism

20 Thanks to Dominika Swolkien for making me aware of this point.

and final epenthesis, respectively (Kihm 1994: 272). We have shown in §4.1 that the form *sedo* is probably from the variety of Cacheu. The fact that the variety of Bissau has almost completely replaced the other varieties explains why we have in GB the form *sedu* and not *sedo*. The presence of *sedi* in CS does not come as a surprise: the ending *-i* is probably older than *-u* in GB varieties and remnants of it are visible in GB words such as *kasamentu*/*kasamenti* 'wedding'. Moreover, *sedi* is often indicated by GB speakers as an older item which may be still heard in elders' speech. Furthermore, in his brief sketch of GB grammar (variety of Bissau), Wilson (1962: 23) reports the form *sede*, which is not to be found in any of the diachronic and synchronic works available. This could mean either that the development into /-u/ in GB could be more recent or that the two forms co-existed, just like the forms *kasamentu*/*kasamenti* above.[21] On the basis of the (former) presence of *sede* in GB, we propose the proto-UGC form **sere* as ancestor of *ser*/*sedu*/*sedi* (also see Moreira 2020: 205 for the postulation of the form **sere* in proto-UGC).[22] ST present-day form *ser* would be due to renewed Portuguese influence.[23] If we consider that the form **sere* emerged from the need of an infinitival form for marking TAM properties other than the present perfective, then the use of bare *sedu*/*sedi* in GB/CS results as an innovation of continental UGCs, while it is found in more restricted (i.e. negated) contexts in the insular varieties.[24] A more in-depth study of the semantic properties of this copula is still needed in order to complete the picture (§4.2). From 19th-century data (§4.1), we know that *sedo* could be used either bare (with a present indicative interpretation) or with aspect markers. This, along with remnant uses of bare *ser* in rural varieties of ST (Quint 2012; see also footnote 10), and the regular use of bare *sedu*/*sedi* in GB/CS suggest that bare **sere* in proto-UGC was used as a variant of *e*, in addition to its use in conjunction with TAM markers.

21 The former hypothesis is the most plausible: given that forms in *-i* are pervasive in CS (e.g., *sedi*, *kasamenti*) and still possible in ST (Nicolas Quint, p.c.), whereas both forms in *-i* and *-u* coexist in GB, it is likely that the former are older.

22 Another possibility is that an epenthetic *-e* was added to an initial form **ser* (thanks to Nicolas Quint for suggesting this alternative). However, since all the corresponding continental forms discussed in this chapter, namely *sedo*, *sede*, *sedu*, and *sedi*, present a final vowel, we tend to consider **sere* as the most likely original form.

23 If we consider that final mid-vowels in GB are usually close (see Kihm 1994: 17), what nicely explains the form *sedi*, the item *sede* /sɛdɛ/ in Wilson (1962: 23) looks like an exception in this regard. We could hypothesise that final /ɛ/ was inserted due to vowel harmony.

24 In rural varieties of ST, bare *ser* is used instead of the past perfective *foi* (footnote 10): this could represent evidence that the bare use of this pseudo-infinitival form was already present in proto-UGC and survived in some ST rural varieties. At present, we do not have sufficient data in order to corroborate this hypothesis.

With regard to the expression of the past tense in proto-UGC, we suggest that the original past perfective strategy was the non-verbal one: **e* + complement + *ba* (also see Ichinose 1993 as for GB). Evidence is represented by the fact that the substrate Mandinka and Wolof – and Atlantic languages in general – display a non-verbal strategy for the expression of past tense in copular clauses, where the non-verbal copula (or copula-like item) is followed by the past marker (see example 6d and 12a, 12b, and12c for continental UGCs and their African contributors, respectively). The past marker in these languages always occupies a peripheral position (§4.3): it may either attach to the right of a verb or occur, as a free morpheme, after a non-verbal item (noun, adjective or adverb) or in sentence-final position (Truppi and Hagemeijer 2018; Truppi 2019 and references therein). This strategy survived on the continent, where GB/CS *ba*(*ŋ*) can show up after verbs, after nominals/adverbials and in sentence-final position as a free morpheme, whereas in CV *-ba* became a suffix and is used only with verbs (Truppi and Hagemeijer 2018). With regard to the suppletive form *era*, since this form is shared by all UGCs, we argue that it dates back to proto-UGC. It comes from Portuguese 3SG past imperfective form *era* 'was' (similarly to the case of *sta*; see footnote 13). It may have emerged due to the convergence of two factors: (i) the need to express the past imperfective, and (ii) the frequency and, therefore, availability of the Portuguese form (see Clements 2022). The form *yera* on the continent represents a phonological change through insertion of initial semivowel /y/ in analogy to forms such as *yanda* 'to walk' and *yagu* 'water' (from Portuguese *andar* and *água*, respectively). Finally, with regard to *foi*, it was not present in proto-UGC; it was adopted more recently from CV and GB due to Portuguese influence. This is supported by the fact that *foi* is not present in CS and is felt as more "Portuguese" by GB speakers. Also, Quint (2012: 169) observes that this form is rarely uttered in rural varieties of ST.

Finally, with regard to the stage-level and locative copula *sa*/*sta*, as already mentioned, these forms most probably come from Portuguese 3SG present indicative *está* (stage-level copula). We have two possibilities: (i) both *sa* and *sta* were present in proto-UGC, but the former survived only in peripheral varieties (in CS and, along with *sta*, in FG); (ii) *sa* was the stage-level/locative copula in proto-UGC and was later substituted by *sta* in CV and GB due to Portuguese influence, while it survived as a variant of *sta* in FG and as the unique form for stage-level and locative predication in CS.[25] The latter may be due to the fact that CS was not in contact with Portuguese since the end of the 19th century. The form *sta* is found in Schuchardt's works (1887; 1888a,b): thus the substitution of *sa* by *sta* must have happened before

25 A further possibility is that *sa* in CS is a variant of *sta* (Rougé 2013; see footnote 12). Unfortunately, no data on this form was found in the literature on CS.

that period.[26] Before concluding and looking at Figure 4, which summarises our reconstruction, it is worth mentioning that the form *sa* resembles the LdP copula *sa*, used in both individual-level and stage-level predication (see Kihm and Rougé 2013). However, if one assumes LdP at the basis of the Portuguese-related pidgin from which proto-UGC possibly emerged, it would be hard to explain why LdP *sa* would lose its individual-level function and keep only the stage-level (and locative) meaning. Moreover, according to the reconstruction of LdP by the authors, there also are seldom occurrences of the copulas *he*/*é* and *estae* (Portuguese *é* 'is' and *está* 'is, stays', respectively). If so, we also need to explain why *sa* would be both individual-level and stage-level, whereas further specialised forms were present in LdP.

a.	Split-encoding	
b.	Individual-level (perfective)	*ø/ *e
c.	Individual-level (perfective/non-finite)	*sere
d.	Individual-level (past perfective)	*ø/*e + *ba
e.	Individual-level (past imperfective)	*era
f.	Stage-level and locative	*sa/*(sta)

Figure 4: The system of copulas in proto-UGC.

6 Conclusions

The emergence of copulas in proto-UGC is the result of a complex process of language emergence in a context of extreme language contact. L2 acquisition in a multilingual context where the social groups involved had no common language is involved in the process, along with basic patterns of organisation of information structure – acting in a context of L2 acquisition – and grammaticalisation. Our reconstruction shows that the paradigm of copulas of proto-UGC was a split system, i.e., it had one copula specified for individual-level predication and one copula for stage-level (and locative) predication. With regard to individual-level predication, proto-UGC was initially copulaless and, subsequently, derived a copula from 3SG clitic pronoun through grammaticalisation, similarly to what happens in several

26 It is possible that *sa* was present in the GB variety of Cacheu and, after the foundation of Ziguinchor, it survived in CS because of the lack of contact with Portuguese (starting from the Berlin Conference, when Ziguinchor became a French dominion), while, in all other varieties, *sa* was substituted by *sta* due to the spread of the Bissau variety and to Portuguese influence. Unfortunately, we do not have data from the historical variety of Cacheu previous to the late 19th-century in order to evaluate this hypothesis.

languages from around the world. In the present case – and in virtually any case of language change (see, e.g., Mufwene 2001) – both universal constraints acting in L2 acquisition and the influence of the languages in contact were involved in the process.

Speaking of grammaticalisation, we did not refer to contact-induced grammaticalisation due to two reasons: (i) we agree with Mufwene (2001, 2006) that whether the trigger is internal or external does not yield any structural difference in the process itself; (ii) the speakers who created proto-UGC did not transfer any pattern (or matter) from their native languages with regard to the emergence of the individual-level copula. As a matter of fact, no substrate/adstrate language analysed so far presents non-verbal pronominal copulas of the type discussed here. With regard to (ii), we referred to the speakers' role in the process of grammaticalisation in §3. Speakers actively participate in language change: according to Mufwene (2001), they recreate certain aspects of their native language(s) and innovate the grammar of the emerging variety. With regard to our case study, the speakers creating proto-UGC innovated with respect to their native languages: on the basis of constraints during the process of L2 acquisition and following a basic pattern of discourse organisation they created individual-level copular clauses replicating the structure of topic-comment expressions through grammaticalisation. As a consequence, a non-verbal copula emerged from a pronoun: this represents an innovation with respect to UGCs contributing languages. Unfortunately, in this chapter we could not discuss in more detail this fascinating topic; we wish that this research has given some insights with respect to the speakers' active role in language change and UGC emergence, and that it may represent a hint for future research.

Finally, with regard to influences from African languages, our reconstruction – together with the findings in Truppi (2021a) – has helped assessing what is substrate (Mandinka and Wolof) and what is possibly adstrate (the Atlantic languages mentioned above other than Wolof), while the role of Temne remains quite marginal. We also mentioned possible areal influences acting in Upper Guinea: however, this possibility needs to be further explored.

Abbreviations

1,2,3	person
A	aspect
BV	basic variety
C	classifier
CL	clitic
COP	copula

CS	Casamancese
CV	Capeverdean
D	definite
DEM	demonstrative
DIST	distal
EMPH	emphatic
FEM	feminine
FOC	focus
GB	(Bissau-)Guinean
IPFV	imperfective
LdP	Língua de preto
NCL	non-clitic
NEG	negation
NP	noun phrase
OBL	oblique
P	preposition
PFV	perfective
PL	plural
POSS	possessive
POSTP	post-position
PROX	proximal
PRS	present
PST	past
SG	singular
TAM	tense-aspect-mood

References

Albuquerque, Luís de & Maria E. Madeira Santos (eds.). 1991. *História geral de Cabo Verde* [General history of Cape Verde]. Lisbon/Praia: Instituto de Investigação Científica Tropical/Direcção Geral do Património Cultural de Cabo Verde.

Anttila, Raimo. 1989. *Historical and comparative linguistics.* Amsterdam: John Benjamins.

Arends, Jacques. 1989. *Syntactic developments in Sranan*. Nijmegen: University of Nijmegen dissertation.

Arnott, David W. 1970. *The nominal and verbal systems of Fula*. Oxford: Oxford University Press.

Baleno, Ilídio Cabral. 1991. Povoamento e formação da sociedade [Settlement and formation of the society]. In Luís Albuquerque & Maria E. Madeira Santos (eds.), *História geral de Cabo Verde* [General history of Cape Verde]. Lisbon/Praia: Instituto de Investigação Científica Tropical/ Direcção Geral do Património Cultural de Cabo Verde.

Baptista, Marlyse. 2002. *The syntax of Cape Verdean Creole: The Sotavento varieties*. Amsterdam/ Philadelphia: John Benjamins.

Baptista, Marlyse. 2004. A cross-linguistic comparison of copular predication: Some basic assumptions revisited. *Journal of Portuguese Linguistics* 3. 97–113.

Baptista, Marlyse, Heliana Ribeiro de Mello & Miki Suzuki. 2007. The syntax of Cape Verdean Creole and Guinea-Bissau Creole. In John Holm & Peter Patrick (eds.), *Comparative creole syntax*, 53–82. London: Battlebridge.

Barros, Marcelino Marques de. 1897–1899. O Guinèense [Guinean]. *Revista Lusitana* 5. 174–181; 271–298.

Barros, Marcelino Marques de. 1900–1901. O Guinèense [Guinean]. *Revista Lusitana* 6. 300–317.

Biagui, Noël B. 2012. *Description générale du créole afro-portugais parlé à Ziguinchor (Sénégal)*. Dakar/Paris: Université Cheikh Anta Diop de Dakar (UCAD)/Institut National des Langues et Civilisations Orientales (INaLCO) dissertation.

Biagui, Noël B., Joseph J.-F. Nunez & Nicolas Quint. Forthcoming. *Casamance Creole*. In Friederike Lüpke (ed.), *Oxford guide to the world's languages: Atlantic.* Oxford: Oxford University Press.

Brito, António de Paula. 1967 [1887]. Apontamentos para a gramática do crioulo que se fala na ilha de Santiago de Cabo Verde [Notes for the grammar of the creole spoken on the island of Santiago of Cape Verde]. In Joaquim Morais-Barbosa (ed.), *Estudos linguísticos crioulos*, 329–404. Lisbon: Academia Internacional de Cultura Portuguesa.

Bruyn, Adrienne. 2009. Grammaticalization in creoles: Ordinary and not-so-ordinary cases. In Clancy J. Clements & Shelome Gooden (eds.), *Language change in contact languages: Grammatical and prosodic considerations*, 312–337. Amsterdam: Benjamins.

Clements, Clancy. 2022. Paradigmatic restructuring: The case of northern Indo-Portuguese creoles. In Bettina Migge & Shelome Gooden (eds.), *Social and structural aspects of language contact and change*, 157–182. (Contact and Multilingualism 5). Berlin: Language Science Press.

Cobbinah, Alexander Y. 2013. *Nominal classification and verbal nouns in Baïnounk Gubëeher*. London: SOAS dissertation.

Corpus informatizado do português medieval – CIPM. Centro de Linguística da Universidade Nova de Lisboa. https://cipm.fcsh.unl.pt (accessed 28 May 2023).

Costa, Joaquim Vieira Botelho da & Custódio José Duarte. [1886] 1967. O crioulo de Cabo Verde. Breves estudos sobre o crioulo das ilhas de Cabo Verde oferecido ao Dr. Hugo Schuchardt [The creole of Cape Verde. Brief studies on the creole of the islands of Cape Verde offered to Dr. Hugo Schuchardt]. In Joaquim Morais-Barbosa (ed.), *Estudos linguísticos crioulos*, 235–327. Lisbon: Academia Internacional de Cultura Portuguesa.

Creissels, Denis. Forthcoming. A sketch of Mandinka. In Friederike Lüpke (ed.), *The Oxford guide to the Atlantic languages of West Africa*. Oxford: Oxford University Press.

Creissels, Denis & Pierre Sambou. 2013. *Le Mandinka: Phonologie, grammaire, textes*. Paris: Karthala.

Donelha, André. 1625. *Descrição da Serra Leoa e dos Rios da Guiné do Cabo Verde/Description de la Serra Leoa et des Rios de Guiné du Cabo Verde*, edited by Avelino T. da Mota with notes by P.E.H. Hair. Lisbon: Junta de Investigações Científicas do Ultramar, 1977.

Hans-Bianchi, Barbara, Chiara Truppi & Barbara Vogt. This volume. Contact-related variation and change: crossing the boundaries within and between languages. In Barbara Hans-Bianchi, Chiara Truppi & Barbara Vogt (eds.), *Speakers and structures in language contact: pluralistic approaches to change and variation.* (Language Contact and Bilingualism 31). Berlin/Boston: Mouton De Gruyter.

Havik, Philip J. 2007. The port of Geba: At the crossroads of Afro-Atlantic trade and culture. *Mande Studies* 9. 21–50.

Heine, Bernd. 2003. Grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 575–701. Oxford: Blackwell.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Ichinose, Atsushi. 1993. Evolução da expressão equacional no kiriol da Guiné-Bissau [Evolution of the equational expression in Guinea-Bissau Kiriol]. *Papia* 2(2). 23–31.
Jacobs, Bart. 2010. Upper Guinea creole: Evidence in favor of a Santiago birth. *Journal of Pidgin and Creole Languages* 25(2). 289–343.
Jacobs, Bart & Nicolas Quint. 2016. Relevance of classical Portuguese features in four Atlantic creoles. Armin Schwegler, John McWhorter & Liane Stroebel (eds.), *The Iberian Challenge: Creoles beyond the plantation setting*, 67–83. Madrid/Frankfurt am Main: Iberoamericana/Klaus Vervuert.
Karlik, Jan. 1972. *A Manjaku grammar*. London: SOAS dissertation.
Katz, Aya. 1996. *Cyclical grammaticalization and the cognitive link between pronoun and copula*. Houston, TX: Rice University dissertation.
Kihm, Alain. 1994. *Kriyol syntax: The Portuguese-based creole language of Guinea-Bissau*. Amsterdam: Benjamins.
Kihm, Alain. 2007. The two faces of creole grammar and their implications for the origin of complex language. In Regine Eckardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, selection, development: Probing the evolutionary model of language change*, 253–305. Berlin: Mouton De Gruyter.
Kihm, Alain. 2011. Substrate influence in Kriyol: Guinea-Bissau and Casamance Portuguese-related Creole. In Claire Lefebvre (ed.), *Creoles, their substrates, and language typology,* 81–103. Amsterdam: John Benjamins.
Kihm, Alain & Jean-Louis Rougé. 2013. Língua de Preto, the Basic Variety at the root of West African Portuguese creoles: A contribution to the theory of pidgin/creole formation as second-language acquisition. *Journal of Pidgin and Creole Languages* 28(2). 203–298.
Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13. 301–347.
Lang, Jürgen. 2009. *Les langues des autres dans la créolisation: théorie et exemplification par le créole d'empreinte wolof à l'île Santiago du Cap Vert.* Tübingen: Gunter Narr.
Li, Charles N. & Sandra A. Thompson. 1977. A mechanism for the development of copula morphemes. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 419–444. Austin, TX: University of Texas Press.
McWhorter, John. 1997. Lost in transmission: A case for the independent emergence of the copula in Atlantic creoles. In Arthur K. Spears & Donald Winford (eds.), *The structure and status of pidgins and creoles*, 241–264. (Creole Language Library 19). Amsterdam: John Benjamins.
Melissaropoulou, Dimitra. This volume. Variation and change in comparative structures as a conspiracy of language internal and language external factors: evidence from Asia Minor Greek. In Barbara Hans-Bianchi, Chiara Truppi & Barbara Vogt (eds.), *Speakers and structures in language contact: pluralistic approaches to change and variation.* (Language Contact and Bilingualism 31). Berlin/Boston: Mouton De Gruyter.
Michaelis, Susanne M. & Martin Haspelmath. 2020. Grammaticalization in creole languages: Accelerated functionalization and semantic imitation. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization scenarios from Africa, the Americas, and the Pacific*, vol. 2, 1109–1127. Berlin: De Gruyter.
Moreira, Karina T. 2020. Documentation et description grammaticale et lexicale du créole afro-portugais de l'île de Fogo (République du Cap-Vert, Afrique de l'Ouest). Paris: INALCO dissertation.
Mufwene, Salikoko S. 1996. The Founder Principle in creole genesis. *Diachronica* 13(1). 83–134.
Mufwene, Salikoko S. 2001. *The ecology of language evolution.* Cambridge, MA: Cambridge University Press.

Mufwene, Salikoko. 2006. Grammaticalization is part of the development of creoles. *Papia* 16. 5–31.
Plag, Ingo. 2002. On the role of grammaticalization in creolization. In Glenn G. Gilbert (ed.), *Pidgin and creole linguistics in the twenty-first century*, 229–246. (Studies in Ethnolinguistics). New York, NY: Peter Lang.
Quint, Nicolas. 2000. *Le cap-verdien: origines et devenir d'une langue métisse.* Paris: L'Harmattan.
Quint, Nicolas. 2012. Les influences du portugais contemporain sur le système verbal du capverdien santiagais. In Claudine Chamoreau & Laurence Goury (eds.), *Changement linguistique et langues en contact: Approches plurielles du domaine prédicatif*, 155–178. Paris: CNRS Éditions.
Quint, Nicolas. 2015. Capeverdean words derived from Portuguese non-infinitive verbal forms: a descriptive and comparative study. *Papia* 25(2). 189–215.
Quint, Nicolas & Karina Moreira Tavares. 2019. The common African lexical core of the Upper Guinea Creoles and its historical significance. [Special issue]. *Journal of Ibero-Romance Creoles* 9(1). 113–144. http://www.acblpe.com/revista/volume-9-2019/the-common-african-lexical-core-of-the-upper-guinea-creoles.
Rabanus, Stefan. This volume. Two modes of contact-induced change in minority languages: phonology and syntax vs. inflectional morphology. In Barbara Hans-Bianchi, Chiara Truppi & Barbara Vogt (eds.), *Speakers and structures in language contact: pluralistic approaches to change and variation.* (Language Contact and Bilingualism 31). Berlin/Boston: Mouton De Gruyter.
Rougé, Jean-Louis. 1986. Uma hypótese sobre a formação do crioulo da Guiné-Bissau e da Casamansa [A hypothesis about the formation of the creole of Guinea-Bissau and Casamance]. *Soronda* 2. 28–49.
Rougé, Jean-Louis. 1999. Apontamentos sobre o léxico de origem africana dos crioulos da Guiné e de CaboVerde (Santiago) [Notes on the African-derived léxicon of Guinea-Bissau and Cape Verde (Santiago)]. In Klaus Zimmermann (ed.), *Lenguas criollas de base lexical española y portuguesa*, 49–67. Frankfurt am Main: Vervuert.
Rougé, Jean-Louis. 2013. Créole de Casamance: Émergence de nouvelles variétés. [Special issue]. *Travaux de Cercle Linguistique d'Aix en Provence* [CLAIX] 24. 201–213.
Rougé, Jean-Louis. 2019. Eléments pour l'histoire du kriol de Guinée Bissau et de Casamance: de la variation à la variation. *Faits de langues* 49(1). 47–68.
Salas Barrena, Nicté-Ha I. 2006. Another look at the problem of the copula in the Santiago variety of Cape Verdean Creole. *Papia* 16. 32–52.
Schuchardt, Hugo. 1887. Review of Joaquim Vieira Botelho da Costa and Custódio Duarte (1886), O creôlo de Cabo Verde: Breves estudos sobre o creôlo das ilhas de Cabo Verde, offerecidos ao dr. Hugo Schuchardt [The creole of Cape Verde: Brief studies on the creole of Cape Verde islands, offered to Dr. Hugo Schuchardt]. *Literaturblatt für Germanische und Romanische Philologie* 8. 132–141.
Schuchardt, Hugo. 1888a. Beiträge zur Kenntnis des kreolischen Romanisch II: Zum Negerportugiesischen Senegambiens. *Zeitschrift für romanische Philologie* 12. 301–312.
Schuchardt, Hugo. 1888b. Beiträge zur Kenntnis des kreolischen Romanisch III: Zum Negerportugiesischen der Kapverden, *Zeitschrift für romanische Philologie* 12. 312–322.
Stassen, Leon. 2013. Nominal and locational predication. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/119. Accessed on: February 28, 2019.
Swolkien, Dominika. 2014. *The Capeverdean Creole of Sao Vicente: Its genesis and structure*. Coimbra: University of Coimbra dissertation.
Torrence, William H. 2005. *On the distribution of complementizers in Wolof*. Los Angeles, CA: University of California dissertation.

Truppi, Chiara. 2019. Copulas in contact: Kriyol, Upper Guinea Creoles, and their substrate. [Special issue]. *Journal of Ibero-Romance Creoles* 9(1). 84–112. http://www.acblpe.com/revista/volume-9-2019/copulas-in-contact-kriyol-upper-guinea-creoles.

Truppi, Chiara. 2021a. Predicational encoding in Upper Guinea linguistic area. [Special issue]. *Revista do GEL* 18(3). 191–218. https://revistas.gel.org.br/rg/article/view/3167

Truppi, Chiara. 2021b. The copula cycle in Kriol. [Special issue]. *Journal of Historical Syntax* 5 (12). 1–27. https://ojs.ub.uni-konstanz.de/hs/index.php/hs/article/view/42

Truppi, Chiara & Tjerk Hagemeijer. 2018. Grammaticalization of tense in the Upper Guinea Creoles. Paper presented at the 20th Diachronic Generative Syntax Conference (DiGS), York, June 18–21, 2018.

Veiga, Manuel. 2000. *Le créole du Cap-Vert: étude grammaticale, descriptive et contrastive*. Paris: Karthala.

Wilson, William A.A. 1962. *The crioulo of Guiné*. Johannesburg: Witwatersrand University Press.

Wilson, William A.A. 2007. *Guinea languages of the Atlantic group: Description and internal classification*. Frankfurt am Main: Peter Lang.