

Paper for the Symposium on *Language Contact and the Dynamics of Language: Theory and Implications*, Max Planck Institute (Leipzig), 10-13 May 2007

Theme: *Contact, typology and evolution of languages: a perspective to be explored*

## **Creole Typology and Substrate Typology**

John Holm (University of Coimbra)

### **1 Introduction**

This paper seeks to cast light on two major questions about creole languages. The first is whether creole languages in general share structural similarities that could be said to constitute a typology. Although there has been mounting evidence in recent years that they do not, especially when creoles of the Atlantic area are compared to those of Asia and the Pacific, the question is still open as to what degree particular groups of creoles reflect the typological similarities of their superstrate and substrate languages, thus forming typological groups of their own. Thus the second major question, perennial in the field, must also be dealt with: to what degree do creoles reflect the structure of their superstrate and substrate languages.

This study first compares the typology of six creole languages with Portuguese and/or Spanish superstrates and various African, Asian and Pacific substrates, considering the 97 morphosyntactic features surveyed in Holm and Patrick, eds. (2007). The point is to quantify structural similarity within groupings, particularly by superstrate and substrate.

To deal with the second major question regarding the influence of substrate typology on creole typology, there is next a comparison of the same 97 grammatical features of a particular creole (Guiné-Bissau Creole) with those of its superstrate (Portuguese) and one of its substrate languages (Balanta). The results are tabulated not by areas of syntax but by the patterns of the presence or absence of the features (e.g. present in all or none or only in the creole, etc.) to find what these patterns reveal about the validity of theoretical assumptions about the sources of creole features.

### **2 Creole typology by lexical base**

To begin with the first question above regarding creole typology, Holm (2001) studied a group of creole languages lexically based on Spanish, comparing the 97 morphosyntactic features surveyed in Holm and Patrick, eds. (2007) in Zamboangueno, Papiamentu and Palenquero. The goal was to quantify the structural similarity between creoles to measure the relevance of creoles sharing a superstrate or a substrate. For the latter purpose, the first comparison included a fourth creole which does not share the superstrate, but does share the Austronesian substrate of Zamboangueno: Tok Pisin Pidgin/Creole English. A later study (Holm 2006) expanded the survey to include Portuguese-based creoles with various substrates: Cape Verdean and Guiné-Bissau Creole Portuguese (hereafter CP), Angolar, and Korlai CP.

There are at least two ways of quantifying this kind of typological data in order to compare

the structural similarity of languages. The first, which matches pairs of languages as to presence or absence of features, will be discussed here. The second, which quantifies the general tendency in each of a number of languages towards a particular typology, will be discussed below.

The first quantificational method was to count only exact matches in pairs of languages, e.g. +=+, 0=0 (where + indicates the attested presence of a feature and 0 its attested absence), ignoring possible matches, e.g. R=+ (where R indicates that a feature has been attested but is rare). This yielded the following percentages of shared features:

<u>Spanish-based creoles</u>		<u>Portuguese-based creoles</u>	
Zamboangueno = (Tok Pisin)	49%	Angolar = Korlai	51%
Zamboangueno = Papiamentu	53%	Cape Verdean = Korlai	65%
Palenquero = Papiamentu	62%	Angolar = Cape Verdean	73%
Zamboangueno = Palenquero:	63%		

That is, of the 97 morphosyntactic features surveyed, Zamboangueno shares 49% with Tok Pisin, 53% with Papiamentu, etc. Regarding the Spanish-based creoles, these figures offer some surprises about the importance of both the superstrate and the substrate as a source of creole grammatical features. The lowest percentage of parallel constructions is found between Zamboangueno and Tok Pisin, which share an Austronesian substrate but which have two different superstrates: Spanish and English. A higher percentage of parallel features is found between Zamboangueno and the two other Spanish-based creoles: Papiamentu on the one hand (53%) and Palenquero on the other (63%)--although the Niger-Congo languages that form the substrates of the latter two are typologically quite distinct from Austronesian languages. About the same percentage of parallel constructions is found between Palenquero and Papiamentu (62%)--languages that share both a superstrate and a substrate (even though there are some important typological differences between the West African "Kwa" languages that influenced Papiamentu and the Central African Bantu languages that influenced Palenquero, although both subgroups are part of the larger Niger-Congo family).

To continue with the Portuguese-based creoles, the lowest percentage of parallel constructions is found between Angolar and Korlai CP, which share their superstrate but not their substrate (Niger-Congo and Indic, respectively). Surprisingly, Cape Verdean and Korlai CP, with the same respective substrates, have a considerably higher percentage of common features (65%). Less surprisingly, Angolar and Cape Verdean have the highest percentage of such features (73%), both having substrate languages from the Niger-Congo family, albeit from different subgroups (Kwa and Bantu for Angolar, Mande and West Atlantic for Cape Verdean).

The second method of quantifying data in a typological survey is that used by Thomason and Kaufman (1988:315): "Each feature is assigned a weight of one point, and a marginal or less decided feature is given half a point. Plus (+) means present; minus (-) means absent; marginal features are in parentheses; (?) means unknown or unclear." Adjusting this for the symbols above, each plus for a particular language is given one point, each 0 is given 0, each R is given half a point, and each ?--indicating that the presence of the feature in the relevant

creole is unknown in Holm and Patrick, eds. (forthcoming)--remains untabulated. This yields the following results:

Palenquero CS	52.5%
Korlai CP	53%
Zamboangueno CS	57%
Cape Verdean CP	65%
Angolar CP	67%
Papiamentu CS	73%

My original selection of structural features was guided by those discussed in the chapters on syntax in Holm 1988, and this selection was influenced by my own research up to that point, which had focussed largely on the Atlantic creoles, and particularly those of the Caribbean with substrate languages that tended to come from the eastern part of the Guinea coast (i.e. not the West Atlantic and Mande languages that form the substrate of Cape Verdean and Guiné-Bissau CP). This seems to account for Palenquero Creole Spanish (hereafter CS) being at the low end of the scale (although it is surprising that it is marginally lower than Korlai CP and Zamboangueno CS) and for Papiamentu being at the high end.

In the above comparative survey, the lowest percentage of parallel constructions occurs between the two creoles which share the same substrate but which have two different superstrates, while the highest percentage of parallel constructions occurs between two creoles that share both a superstrate and a substrate. All in all, this survey suggests that there are indeed areal groups of creoles that reflect the typological similarities of their superstrate and substrate languages. This provides a start to answering the first major question posed above.

### 3 Superstrate and substrate typologies

To deal with the second major question, the degree to which creoles do or do not reflect the structure of their superstrate and substrate languages, one must begin by systematically comparing a substantial number of representative grammatical features of a particular creole with those of its superstrate and substrate. This has never been done, to my knowledge, because this combination of data (particularly regarding the substrate) is difficult to obtain. However, I am fortunate to be working with Incanha Intumbo, a graduate student at Coimbra University, who is working on a grammatical comparison of his first language, Guiné-Bissau Creole Portuguese (GB CP) with Balanta (BAL, one of the creole's substrate/adstrate languages, and his second native language) and Portuguese (PORT, the creole's superstrate and his language of instruction).

We have written a joint paper (Holm and Intumbo 2007) comparing the data on GB CP in the chapter on *Kabuverdianu or Cape Verdean, and Kriyol, or Guiné-Bissau Creole Portuguese* by Baptista, Mello and Suzuki in Holm and Patrick, eds. (2007) with corresponding structures in Balanta and Portuguese. It should be noted that while Baptista, Mello and Suzuki (2007) (hereafter BMS) is a superb study by three very competent linguists, one of whom is a native speaker of Cape Verdean (which is closely related to GB CP), none of the authors is a native speaker of the Guinean creole and all of their data on this language was taken from the published literature. This literature did not happen to

cover all 97 structures surveyed in Holm and Patrick, eds. (2007) but some of the missing structures were part of Intumbo's native speaker competence. To ensure the accuracy of the Portuguese data, all sentences in this language were monitored by Liliana Inverno, a linguist who is a native speaker of the European standard. Unfortunately the documentation of Balanta (despite its 367,000 speakers) appears to consist of a single 21-page article (Wilson 1961) but this has not deterred Intumbo, who has published three comparative studies involving Balanta (2005, 2006a, 2006b).

#### 4 Discussion of the data on GB CP, Balanta and Portuguese

Constraints on space prevent the discussion here of examples of all 97 structures in the three languages or the lack thereof. Instead, the tables that come at the end of the sections on various areas of syntax in BMS showing the presence or absence of particular structures have been expanded to include Balanta and Portuguese as well as GB CP. The structures discussed in BMS have sequential numbers (1.1, 1.2 etc.) that correspond to those of parallel structures in the creole languages whose grammars are outlined in the other chapters of Holm and Patrick, eds. (2007). The reader is referred to that discussion for any clarification needed regarding the nature of the features, and to Holm and Intumbo (2007) regarding their occurrence in the three languages being compared here.

The tables below are organized not by grammatical features but by the patterns of the presence or absence of the features (e.g. present in all or none or only in the creole, etc.) to reveal structural similarities and dissimilarities among the three languages for the light this casts on the validity of the theoretical assumptions that have guided research on creole languages. Some time ago, I characterized my position as that of a "moderate substratist" (Holm 1988: 11), defining this as an understanding that the creoles' "common syntactic features reflect the influence of both superstrate and substrate languages, as well as universals of adult second-language acquisition, creole-internal innovations, and often the convergence of all or some of these" (Holm 1988:144).

We will begin by examining the grammatical features that were found to be missing in all three languages: creole, substrate and superstrate.

<b>Table 1: Feature totally absent</b>		<b>GB CP</b>	<b>BAL</b>	<b>PORT</b>
2.5	Anterior (or past) with locative predicates	0	0	0
4.2	Progressive marker for habitual	0	0	0
6.4	Anterior plus irrealis = future perfect	0	0	0
8.3	'For' as a (quasi-) modal	0	0	0
14.7	Serial verbs constructions with 4+ verbs	0	0	0
15.4	Plural marker (= 'they')	0	0	0
15.7	Demonstrative plus definite or plural	0	0	0
15.8	Relative clause followed by definite or plural	0	0	0
16.1	Nouns: juxtaposition [possessor + possessed]	0	0	0
16.3	Nouns: poss. adj. [possessor HIS possessed]	0	0	0

Why are these features missing from the relevant languages in Table 1? Why was their occurrence expected or at least considered possible? As noted above, I wrote the two volumes of *Pidgins and Creoles* in the 1980s when most creolists were focusing on the

Atlantic creoles of the Caribbean and West Africa. The most relevant group of substrate languages for these creoles, given known historical patterns of the slave trade to the Americas, were believed to be those on Africa's Guinea Coast, particularly the group then known as the Kwa languages, spoken from Liberia in the west to Nigeria in the east, including languages such as Akan, Ewe, Yoruba, Igbo, Kru, Bini, and Ijo (Pulleyblank 1987:966). I had studied Yoruba at the City University of New York and found many features in its grammar that had parallels in Caribbean varieties of Creole English and Creole French, and I focussed on these parallels in my 1988-89 book. These became the salient features in what I then considered to be the likely Atlantic creole typology. However, a number of these features (e.g. 15.4, 15.7 and 15.8 in Table 1 above) are not relevant to the Mande and West Atlantic languages that make up the substrate of Guiné Bissau Creole Portuguese, so it is not surprising that none of the three languages being compared contain these features.

<b>Table 2: Miscellaneous</b>	<b>GB CP</b>	<b>BAL PORT</b>	
4.3 Marker for habitual only	0	+	+
5.1 Completive only (before/after verb)	?	?	0
9.6 Relative clauses (zero relative pronoun)	0	+	0
10.2 Discontinuous double negation	0	0	+
15.11 Gender agreement	0	0	+

These features could be considered random linguistic static. Feature 10.2, for example, is found in some Bantu languages but apparently made its way from São Tomé Creole Portuguese into Brazilian Vernacular Portuguese and then into European Portuguese. Feature 15.11 is the kind of feature that is highly unlikely to occur in a basilectal creole or most African substrate languages. With only five members, this group is not especially important statistically.

<b>Table 3: Feature found in the creole only</b>	<b>GB CP</b>	<b>BAL PORT</b>	
6.2 Anterior plus irrealis = conditional	+	0	0
8.2 'For' as an infinitive marker	+	0	0
8.4 'For' introducing a tensed clause	+	0	0
8.6 Distinct subordinator after verb of speaking	+	0	0
14.1 Directional with 'go'	+	0	0
14.4 Serial 'say' meaning 'that'	+	0	0
14.6 Serial verb constructions with three verbs	+	0	0
15.3 Definite article (from superstrate deictic)	+	0	0
19.2 Zero prep. after motion verb + place	+	0	0

This is the first group of any real relevance to theory. Since these features are found in the creole only, we might expect them to be the result of creole innovation, e.g. the extension of existing systematicity by analogy. In actual fact, most of them are found in African languages that could have been part of the creole's substrate, although they do not happen to occur in Balanta. On the other hand, feature 15.3 seems to be a universal.

<b>Table 4: Feature found in the creole and its superstrate</b>		<b>GB CP</b>	<b>BAL</b>	<b>PORT</b>
4.1	Zero marker for habitual	+	0	+
8.5	Subordinator from superstrate ‘that’	+	0	+
9.5	Relative clauses (rel. pronoun = obj. of a prep.)	+	0	+
10.3	Negative concord	+	0	+
11.1	Passive construction	+	0	+
12.7	Comparison as in superstrate	+	0	+
13.2	Locative copula (with expression of place)	+	0	+
13.6	Existential (‘have’ = ‘there is’)	+	0	+
14.2	Directional with ‘come’	+	0	+
15.9	Prenominal adjective	+	0	+
16.4	Possessive adjectives: prenominal	+	0	+

First, we must remember that the presence of a feature in two languages is not in itself proof that one acquired it from the other. On the other hand, it is also fair to observe that a completely random selection of three features—what could be expected from an honest slot machine producing combinations of peaches, bananas, and apples—would produce more or less equal numbers of the various possible combinations. Instead, we find very unequal numbers here (see below). Moreover, many of the features in table four are those a creole could logically be expected to acquire from an Indo-European superstrate language.

<b>Table 5: Feature found in the creole and its substrate</b>		<b>GB CP</b>	<b>BAL</b>	<b>PORT</b>
1.1	Statives with non-past reference	+	+	0
1.2	Statives with past reference	+	+	0
1.3	Non-statives with past reference	+	+	0
1.4	Non-statives with non-past reference	+	+	0
2.4	Anterior (or past) with adjectival verbs	+	+	0
3.3	Anterior plus progressive	+	+	0
3.4	Progressive with adj. verb = inchoative	+	+	0
5.2	Completive plus adjective	+	+	0
5.3	Anterior (or other marker) plus completive	+	+	0
6.3	Anterior plus irrealis = future in the past	+	+	0
7.1	Irrealis plus progressive	+	+	0
7.2	Anterior plus irrealis plus progressive	+	+	0
8.7	Zero subordinator	+	+	0
12.1	Preverbal markers before adjectives	+	+	0
12.2	Preverbal markers before nouns	+	+	0
12.3	Preverbal markers before locatives	+	+	0
12.4	Predicate clefting: adjectives or adjectival verbs	+	+	0
12.5	Predicate clefting: other verbs	+	+	0
12.6	Comparison with ‘pass’	+	+	0
13.3	Zero copula with adjective	+	+	0
14.5	Serial ‘pass’ meaning ‘more than’	+	+	0
15.1	Bare nouns (generic, definite)	+	+	0
15.5	Personal noun plus plural marker	+	+	0
16.5	Possessive pronouns: distinct	+	+	0
16.6	Poss. pronouns as emphatic poss. adjectives	+	+	0

18.2	'And' joining sentence parts: distinct	+	+	0
19.1	General locative preposition	+	+	0
20.1	Word order: questions SVO	+	+	0
20.2	Sentence final -o	+	+	0

All of the remarks regarding Table 4 are, *mutatis mutandis*, applicable to Table 5.

<b>Table 6: Convergence</b>		<b>GB</b>	<b>CP</b>	<b>BAL</b>	<b>PORT</b>
2.1	Statives with past reference	+	+	+	
2.2	Non-statives with (past-before-) past reference	+	+	+	
2.3	Anterior = counterfactual	+	+	+	
3.1	Indicating progressive	+	+	+	
3.2	Indicating future	+	+	+	
4.4	Anterior plus habitual	+	+	+	
6.1	Future (= progressive marker)	+	+	+	
7.3	Other auxiliary-like elements	+	+	+	
8.1	Zero infinitive marker	+	+	+	
9.1	Subordinate clauses (non-embedded)	+	+	+	
9.2	Subordinate clauses (embedded)	+	+	+	
9.3	Relative clauses (relative pronoun = subject)	+	+	+	
9.4	Relative clauses (relative pronoun = direct object)	+	+	+	
10.1	Single negation (verbal)	+	+	+	
11.2	Passive equivalent	+	+	+	
13.1	Equative copula (with NP)	+	+	+	
13.4	Highlighter with question words	+	+	+	
13.5	Highlighter with other structures	+	+	+	
15.2	Indefinite article	+	+	+	
15.6	Demonstrative	+	+	+	
15.10	Postnominal adjective	+	+	+	
16.2	Nouns: preposition [possessed (of) possessor]	+	+	+	
17.1	Personal pronouns: first person singular	+	+	+	
17.2	Personal pronouns: second person singular	+	+	+	
17.3	Personal pronouns: third person singular	+	+	+	
17.4	Personal pronouns: first person plural	+	+	+	
17.5	Personal pronouns: second person plural	+	+	+	
17.6	Personal pronouns: third person plural	+	+	+	
17.7	Reflexive pronoun: distinct form	+	+	+	
17.8	Interrogative pronouns: some bimorphemic	+	+	+	
17.9	Relative pronouns	+	+	+	
18.1	'And' joining sentences	+	+	+	

We find what logic would predict: a feature is most likely to occur in a creole when it is found in both parent languages because it is a common denominator that can be built on.

To summarize:

**Table 7: Percentage of features in each category**

Table 1: Feature totally absent	11.2%
Table 2: Miscellaneous	5.1
Table 3: Feature found in the creole only	9.2
Table 4: Feature found in the creole and its superstrate	11.2
Table 5: Feature found in the creole and its substrate	29.6
Table 6: Convergence	<u>32.7</u>
	99%

**5 Conclusions**

What support do these figures give to the various hypotheses that have been offered to explain the sources of creole language structures? The least support would be for a non-theory: that the relationship between the creoles and their purported sources is random. However, unambiguous support for the influence of universals as envisioned by Bickerton's Language Bioprogram Hypothesis (1981) is difficult to discern if any is indeed present. Support for the influence of universals of adult second language acquisition, particularly the encoding of grammatical information in free rather than bound morphemes, could be more clearly adduced if the same phenomenon were not characteristic of so many substrate languages. Evidence of possible creole-internal innovation seems limited to Table 3, which makes up less than 10% of the features in this survey. However, this is not much less than the evidence of possible influence from the superstrate only (Table 4). It is the evidence of possible (could we say 'likely'?) influence from the substrate (Table 5) that characterizes nearly 30% of the features and makes up the largest single category except for convergence (Table 6), which slightly exceeds it. But if we consider the last three categories together—superstrate and substrate influence combined with convergence—we have a clear answer to our perennially unanswerable question: they account for the overwhelming majority of the features: 73.5%.

We need to ask whether it is relevant to these results that Balanta is both a substrate and an adstrate language of GB CP with many speakers bilingual in the creole. In fact we need to question the possible bias in the very grammatical categories chosen for the survey, which assumed an Atlantic creole prototype based on "Kwa" rather than West Atlantic languages. However, there is no way to begin tackling these questions without comparable research on other creole languages.

So it is perhaps still premature to claim that any factor is the most crucial one in determining the structure of a creole language. These are initial findings and will need to be reexamined in the light of future research. However, the present study does suggest a promising agenda.

## REFERENCES

- Baptista, Marlyse, Heliana R. de Mello, and Miki Suzuki. 2007. Kabuverdianu or Cape Verdean, and Kriyol, or Guiné-Bissau Creole Portuguese. In Holm and Patrick, eds. 2007.
- Holm, John. 1988-89. *Pidgins and Creoles*. 2 vol. Cambridge: Cambridge University Press.
- . 2001. Chabacano versus related creoles: (Socio-)linguistic affinities and differences. *Estudios de Sociolingüística: Linguas, sociedades e culturas*. 2:2, 69- 93.
- . 2006. Portuguese- and Spanish-based Creoles and Typologies. *Papia* 16, pp.53-61.
- Holm, John and Incanha Intumbo. 2007. Quantifying superstrate and substrate influence. Paper to be presented to the meeting of the *Associação: Crioulos de Base Lexical Portuguesa e Espanhola*. University of Amsterdam, June 18-20.
- Holm, John and Peter L. Patrick, eds. 2007. *Comparative Creole Syntax: Parallel Outlines of 18 Creole Grammars*. London: Battlebridge.
- Intumbo, Incanha. 2005. Balanta, Guiné-Bissau Creole Portuguese and Portuguese: a comparison of the noun phrase. Paper presented at the Conference on Creole Language Structure between Substrates and Superstrates, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- . 2006a. Papiamentu, Guiné-Bissau Creole Portuguese, and its substrate, Balanta: a comparison of the noun phrase. *Revista Internacional de Linguística Iberoamericana* IV:1, pp. 107-115.
- . 2006b. Crioulo guineense, balanta e português: estudo comparativo do sintagma verbal. Paper presented at the meeting of the *Associação: Crioulos de Base Lexical Portuguesa e Espanhola*, Coimbra.
- Pulleyblank, Douglas. 1987. Niger-Kordofanian Languages. In Bernard Comrie, ed. *The World's Major Languages*. New York: Oxford University Press, pp. 959-1014.
- Thomason, Sarah Grey, and Terrence Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: U.of California Press.
- Wilson, W. A. A. 1961. Outline of the Balanta language. *African Language Studies* 2: 139-160. University of London: School of Oriental and African Studies.

University of Coimbra. Creole typology and substrate typology. 13.45 Lunch. 14h.45 15.15.Â constructed by "typology"™. Chair: Martin Haspelmath. Theme Results of language contact seen from a broad comparative perspective; linguistic areas at a local, continental and a global scale; different "diffusibility" or "areal stability" of different linguistic features; the role of a language's structural type in facilitating or rejecting structural influence; the relation between grammatical and lexical borrowing; cross-linguistic comparability of language contact as a basis for typological.