Theories of morphology and theories of creole emergence: the inner connection

Morfologia e emergência das línguas crioulas: a conexão entre as teorias

Alain Kihm
Université Paris-Diderot, CNRS, France
alain.kihm@linguist.jussieu.fr

Abstract: Creole languages have long been considered to be morphologically challenged languages in comparison to their lexifiers. The present paper opposes such a view. It argues that the significant difference between creoles and their lexifiers as far as inflectional morphology is concerned is not a matter of richness, a dimension where inequalities are real but not dramatic, but it has much more to do with the overall predictability of the system, a distinction which only morphological models sharing the Word and Paradigm (WP) theoretical framework are equipped to state in a meaningful way. This is shown by comparing parts of the verbal inflections of European Portuguese and Portuguese-related Guinea-Bissau and Casamance Kriyol for the two dimensions of richness and predictability, using Paradigm Function Morphology (PFM) as a descriptive model. The second claim of the paper is that there exists a connection between models of inflectional morphology and models of creole emergence in terms of conceptual and empirical plausibility. In particular, to the extent that WP models are cognitively more realistic than their competitors, the Second Language Acquisition (SLA) model of creole emergence can be shown to win the prize for psychological and historical plausibility.

Keywords: Inflection, richness/complexity, predictability.
Resumo: As línguas crioulas têm sido consideradas morfologicamente deficientes, ao contrário das suas línguas lexificadoras. O presente artigo opõe-se a esta perspectiva, defendendo que a principal diferença entre as línguas crioulas e as línguas lexificadoras, no que diz respeito à morfologia flexional, não é do domínio da riqueza flexional, uma dimensão em que as desigualdades são reais mas não dramáticas, mas do domínio da previsibilidade geral do sistema, uma característica que apenas modelos morfológicos de tipo *Word and Paradigm* (*wp*) conseguem captar devidamente. Demonstraremos o nosso argumento através de uma comparação parcial da flexão verbal do Português Europeu e dos crioulos portugueses da Guiné-Bissau e de Casamance, no que diz respeito à riqueza flexional e à previsibilidade, recorrendo ao modelo *Paradigm Function Morphology* (*PFM*). Defenderemos ainda que existe uma conexão entre modelos de morfologia flexional e modelos de emergência das línguas crioulas, relativamente à plausibilidade conceptual e empírica. Em particular, na medida em que modelos de tipo *wp* são cognitivamente mais realistas do que modelos concorrentes, uma abordagem da emergência das línguas crioulas segundo o modelo SLA (*Second Language Acquisition*) parece ser historicamente e psicologicamente mais provável.

**Palavras-chave:** Flexão, riqueza/complexidade, previsibilidade.

1 Introduction

Creole languages have long been considered to be morphologically challenged languages in comparison to their lexifiers. Because of what now begins to be recognized as the unfair prejudice that it is (Crowley 2008; Holm 2008), creole studies have tended to focus on syntax, neglecting the crucial contribution comparative morphology could bring to our understanding of creole emergence. The near extinction of morphological processes, especially of the inflectional sort, and their replacement by syntactic devices such as particle-verb sequences instead of inflected verb forms was taken to be a fact, an unavoidable stage in creole formation. And it was thought to be sufficiently explained by the traditional and unquestioned notion according to which morphological phenomena are but complications brought about by the accidents of history. Therefore they have no or little place in languages that are supposed to be closer to the innate blueprint for language (Bickerton 1981) and/or still too young to bear the scars of history (McWhorter 1998).
In the present paper I aim to oppose such a view. I first argue that it is not adequately countered by exhibiting a motley collection of morphological phenomena that creole languages do show. Actually, the significant difference between creoles and their lexifiers as far as inflectional morphology goes is not a matter of richness or the number of distinct forms, a dimension where inequalities are real but not dramatic, but it has much more to do with the overall predictability of the system, a distinction which only morphological models sharing the Word and Paradigm (WP) theoretical framework are equipped to state in a meaningful way (see below). This I will try to demonstrate in sections 2 and 3 by comparing parts of the verbal inflections of European Portuguese (henceforth Portuguese tout court) and of Portuguese-related Guinea-Bissau and Casamance Kriyol for the two dimensions of richness and predictability, using Paradigm Function Morphology (PFM) as a descriptive model (see below)¹.

Then I will show that there exists an inner connection between the theory of inflectional morphology one adopts and the process of creole emergence one finds to be the most probable. Concerning the latter, theories are numerous, but they all include some reference to a process of unguided second language acquisition (SLA) by adults. Thus, the Language Bioprogram Hypothesis (LBH) crucially requires a pidgin or jargon stage that is nothing but very imperfect adult SLA (Bickerton 1981, 1984). Similarly, to the extent that relexification belongs to the mental processes responsible for creole emergence (Lefebvre 1998, 2001; and see discussion in Siegel 2004), it certainly did not affect the natively spoken lexifier, but a partially acquired L2 version of it. McWhorter (2011) explicitly relates the alleged morphological simplicity of creoles to their having been first acquired as L2: ‘The normal state of language is highly complex... Languages significantly low in this kind of complexity... owe this state to second-language acquisition in the past’ (pp. 1-2). Chaudenson (2003) accepts that accounting for the exceptional divergence of French creoles with respect to all dialects of French requires an SLA episode.

Since SLA appears to be such a necessary ingredient of every attempt to account for creole emergence, one comes to wonder why it should occupy a subordinate position in explanatory set-ups. Why could it not be considered the

¹Kriyol’s actual lexifier is sixteenth century Portuguese, also called Middle Portuguese. As far as verbal inflection is concerned, however, Modern Standard European Portuguese does not significantly differ from its ancestor. I therefore feel free to use it as the other term of the comparison.
primary factor to which other factors such as substrate influence or universal
tendencies are subordinate (Muysken 2001; Becker & Veenstra 2003)?

Retaining SLA as the principal agent of creole emergence then implies that
it must first of all be crucially responsible for the two properties of creole
inflectional morphology I wish to demonstrate, namely its richness and its
predictability. Not all notions of inflectional morphology, however, will allow
for these properties to result from an extended SLA process, up to creole
emergence. I think it can be shown that ‘Item and Arrangement’ or ‘Item and
Process’ approaches to inflectional morphology such as Lieber (1992) and Halle
& Marantz (1993) do not allow it because they take inflectional morphemes
to be objects of lexical knowledge like denoting lexemes and therefore to be
involved in the same kinds of operations by speakers. In that sense, they can
be called ‘morphemic’. WP approaches (Blevins 2006) and realizational models
such as PFM (Stump 2001, Bonami & Stump 2013, to appear), in contrast,
make the opposite assumption, namely that inflectional morphs do not exist
as such, but they are abstracted from paradigmatically organized word forms.
Word forms and the paradigms they constitute (or not) are thus the only
input to acquisition, be it L1 or L2 acquisition. Speakers seem to manipulate
morphs when what they actually do is generalize from the paradigms they
already know — as when a speaker of English automatically pluralizes a noun
she has never encountered before by ‘adding s’ to it. This makes WP and the
SLA theory of creole emergence a natural match, as I will try to show. Not
only does the latter, that is the fact that SLA processes may give rise to new
languages, support the WP view of inflectional morphology, but the presence
of inflectional morphology in creole languages in the format predicted by WP
is best explained if the said languages include an SLA episode in their histories.
The first half of the argument will be sustained in section 4, the second half
in section 5 by examining the possible origins of verbal inflections in Kriyol
and another significantly different Portuguese-related creole: Korlai, spoken
in India near Bombay.

In the same way, then, that WP models seem to be cognitively more realistic
than their competitors, the SLA theory of creole emergence wins the prize for
psychological and historical plausibility. This is my conclusion.
2 Comparing Kriyol and Portuguese for morphological richness

According to morphemic models, only forms involving bound morphemes such as English (she) walked or Portuguese cantou ‘s/he/it sang’ or forms proceeding from ablaut such as (she) sang fall within the province of morphology. All the rest, e.g. (she) has sung or tem cantado ‘s/he/it has sung’, as it seems to involve free forms only, is reputed to be syntactically constructed\(^2\). There is indeed a grain of truth in such a view, namely that has sung or tem cantado are verbal tense periphrases made up of an auxiliary (has, tem) and a main verb (sung, cantado), which stand in a syntactic relationship that may differ in different languages. For instance, from the fact that the main verb phrase may be topicalized in English (Sung this song, she has) we infer that the auxiliary and the main verb head two distinct VP’s (Bonami & Webelhuth 2013). Such a construction being ungrammatical in Portuguese (*Cantado esta cantiga, tem), it follows that the syntactic structure of the periphrasis must be that of one flat VP — which does not preclude the auxiliary and the main verb from not being contiguous (cf. She has often sung that song, Tem muitas vezes cantado esta cantiga).

However, it is no less a fact that such periphrastic tense forms fill cells in the paradigms of the respective verbs just as the synthetic forms sung and cantou do. Moreover they are clearly distinct from other verb complexes (or complex predicates) involving modals (She wants to sing, Quer cantar) or serial verb constructions (svc’s) as in the Saramaccan example (1) (Veenstra 1996: 93; and see Ackerman & Stump 2004 for a list of criteria for identifying periphrases; also see Lehmann 1985):

(1) A wáka kó a mi písi.
3SG walk come LOC 1SG yard
‘He walked to my yard.’

\(^2\)Actually, morphology in models such as Lieber (1992) and Halle & Marantz (1993) is formalized as a kind of syntax. This is not important for my argument, however, since this ‘word syntax’ has properties that make it distinct from phrasal syntax.
Both elements in modal constructions and SVC’s have their own lexical identifier (LID). In contrast, HAVE and TER as parts of tense periphrases are divorced from their main verb alter egos meaning ‘to possess’. They then have no proper meaning and share the LID of the main verb whose tense reference they contribute to define. In other words, has sung and tem cantado mean ‘SING-in-the-past’, not ‘HAVE + SING’.

This warrants the conclusion that tense periphrases are creatures of morphology just as much as synthetic forms are. Taking preterite and perfect to be the tense values of cantou and tem cantado respectively, we are then allowed to write the following two Paradigm Functions, that is functions that take as their domain a lexeme and a set of morphosyntactic features associated with it and return the content of a cell in that lexeme’s paradigm (Stump 2001: 32):

\[
\begin{align*}
(2) \text{PF: CANTAR} \land \{\text{TENSE Pret, MOOD Indic, AGR 3SG}\} &= \text{cantou} \\
(3) \text{PF: CANTAR} \land \{\text{TENSE Perf, MOOD Indic, AGR 3SG}\} &= \text{tem cantado}
\end{align*}
\]

In plain talk, (2)-(3) state that cantou ‘s/he/it sang’ is the 3rd person singular of the indicative preterite of the verbal exeme CANTAR, while tem cantado ‘s/he/it has sung’ is the 3rd person singular of the indicative perfect of the same lexeme.

The difference between (2) and (3) is that, whereas the former is complete, the latter requires two statements to the effect that (a) tem is referred for its form (although not for its meaning) to the 3SG indicative present form of the lexeme TER, and (b) cantado is referred to the past participle cell of CANTAR’s paradigm (Bonami & Webelhuth 2013).

Another thing that morphology can take care of instead of syntax are verb complexes involving a verb form and clitics denoting the predicate’s arguments (which may be expressed as full NP’s in the same syntactic environment in case of doubling). Such complexes are typical of Romance languages in which, as shown by Spencer (2005) and Miller & Monachesi (2010), they should be analysed as so many cells in the verb’s paradigm. Portuguese offers especially convincing examples with its mesoclitics as in cantá-la-ei (a cantiga) ‘I’ll sing

---

3 ‘the value of LID is a list of semantic frames that canonically specify the (fine-grained) meaning of a lexeme’ (Sag 2012: 84).

4 A fundamental character of PFM and all realizational models is their being strictly declarative.
it (the song)' (Luís & Spencer 2005). Now, if cantá-la-ei counts as a verb form in CANTAR’s paradigm, there is no reason why cantei-a ‘I sang it’ should not be one as well. And if it makes Portuguese look like a Bantu language, so much the better!

We now are in a position to attempt an informed comparison of verbal inflection in a creole language and its lexifier. As mentioned, Guinea-Bissau and Casamance Kriyol is the creole, Portuguese is the lexifier. Will the comparison vindicate the conventional claim that Kriyol shows much less inflections than does Portuguese. But is this the right question? Is quantity what really matters?

If one insists on brute numbers, it seems true that the paradigms of Kriyol verbs do not include so many distinct forms as do their Portuguese counterparts. Yet, it may be one of those accepted facts that need to be checked again.

As explained in Kihm (1994), Baptista et al. (2007), and Intumbo et al. (2013), the core tma system of the Kriyol verb consists in five constructs, one of which involves the bare form of the verb:

I use Huddleston & Pullum’s (2002) system of tense and aspectuality, easily extended beyond English. With perfective aspectuality, the situation is presented in its totality, as a complete whole… With imperfective aspectuality, the situation is not presented in its totality; it is viewed from within, with focus on some feature of the internal temporal structure or on some subinterval of time within the whole’ (Huddleston & Pullum 2002: 124). Progressive and habitual-iterative are two particular values of imperfectivity. In conjunction

---

5Following these authors, I distinguish aspectuality, a property of predicates, from lexical aspect.

6Other terms one finds in the literature, which I take to be synonymous, are ‘completive’ vs. ‘incompletive’ or ‘accomplished’ vs. ‘unaccomplished’.
with the future, perfective and imperfective mean the same as 'punctual' and 'nonpunctual' respectively.

Bare perfective forms standardly mean past with dynamic verbs like \textit{kanta}, whereas they mean present with stative verbs like e.g. \textit{kuda} 'to think' or \textit{burmedju} '(to be) red'. Progressive is interpreted as temporally present in the default case, but it may also be interpreted as (unspecified) future, more or less as in English (cf. 'I'm leaving tomorrow'). By default (i.e. without explicit indication to the contrary), habitual-iterative also includes utterance time (deictic time in Huddleston & Pullum's terminology).

Following Henri (2010: 268ff.), I analyse progressive \textit{na} and habitual-iterative \textit{ta} not as auxiliaries, but as 'markers', that is syntactic functors selecting lexical heads and analogous to determiners in NP’s (Sag 2012: 86). This analysis rests on the fact that, unlike auxiliaries such as Portuguese \textit{ter} in \textit{tem cantado}, \textit{na} and \textit{ta} have no LID whatsoever: \textit{ta} never occurs outside the construct \{ta V\}; when not a marker, \textit{na} is a preposition meaning 'in' or 'on', that is a distinct item in the present-day language despite probable (but irrelevant) historical connections. Syntactically their position is that of specifier of \textit{vp}:

\[
\text{(4) VP} \quad \text{mkr} \quad V' \\
\quad \quad V \ldots
\]

In morphology they are inflectional clitics, which must be adjacent to the verb hosting them phonologically (e.g. \textit{na kanta} /na=kan'ta/) — unless the predicate negation \textit{ka} precedes them, in which case they are enclitic to it (e.g. \textit{ka na kanta} /'ka=na kan'ta/). Such duality of attachment, being typical of clitics, is what prevents me from regarding \textit{na/ta} as prefixes. Being a morphophonological phenomenon triggered by the presence of negation, on the other hand, it should not deter us from analysing \textit{na/ta} as inflectional elements and accounting for them by means of the following two rules of exponence (see Bonami & Stump 2013 for the PFM formalism I adopt):

\[
\text{(5) RE: } X_{V} & \{\text{vform fin, asp ipfv-prog, tense }< \} \Leftrightarrow na=X \\
\text{(6) RE: } X_{V} & \{\text{vform fin, asp ipfv-hab/iter, tense }< \} \Leftrightarrow ta=X
\]

\textit{PAPIA}, 24(1), ISSN 0103-9415, e-ISSN 2316-2767
Rules (5) and (6) should be read thus: given a verbal lexeme $X_V$ and a feature set including ‘finite’ ($\text{fin}$) as a value for the attribute ‘verb form’ ($\text{vform}$), ‘imperfective-progressive’ or ‘imperfective-habitual/iterative’ for the attribute ‘aspectuality’, and the default-inherited value (notated by empty angled brackets) for the attribute ‘tense’, the verb form realizing this complex is $na=X$ or $ta=X$ ($X$ the stem realizing the lexeme), and the complex exhaustively describes the verb forms. (Rules of exponentce read both ways.)

As for the bare verb form, it is dealt with by way of (7), with aspectuality valued as perfective, and tense defaultly valued as past if dynamicity is part of the verb’s lexical identity, as present otherwise:

\[
(7) \text{ re: } X_V \& \{\text{vform fin, asp pfv, tense < >} \} \Leftrightarrow X
\]

These three rules constitute a block — call it Block I ($\text{tma}$) — and they give us three forms. The Kriyol progressive and habitual forms may be equated with the Portuguese present (as it has both values), indicative ($\text{canto}$ ‘I sing’, $\text{cantas}$ ‘you sing’, and so forth) or subjunctive ($\text{cante}$ ‘that I sing’, $\text{cantes}$ ‘that you sing’, and so forth). Likewise, the Kriyol perfective corresponds to the Portuguese preterite and perfect, indicative ($\text{cantei}$ ‘I sang’, $\text{cantaste}$ ‘you sang’, and so forth, $\text{tenho cantado}$ ‘I have sung’, $\text{tens cantado}$ ‘you have sung’, and so forth) or subjunctive ($\text{cantasse}$ ‘that I sang’, $\text{cantasses}$ ‘that you sang’, and so forth, $\text{tenha cantado}$ ‘that I have sung’, $\text{tenhas cantado}$ ‘that you have sung’, and so forth). Hence thirty-six Portuguese forms ($6 \times 6$), i.e. a ratio of 12:1\(^7\).

If subject clitic pronouns are taken into account, however, the ratio falls dramatically to 2:1 since Kriyol now turns up with eighteen forms as shown in Table 2.

Just as $na/ta$ cannot be separated from the verb, $N$, $bu$ etc. must be adjacent to $na/ta$ or to the bare verb and they cannot be realized in any other environment (see Kihm 1994). It follows that Kriyol subject clitics are aptly analysed as inflectional clitics, as much part of the verb form as are endings in Portuguese conjugated verbs.

\(^7\)Increased to 16 to 1 if the 12 indicative and subjunctive present forms of periphrastic $\text{estar a cantar}$ ‘to be singing’ are added to the total. They should not be, however, since the Portuguese progressive, unlike the Kriyol progressive and unlike the Portuguese simple present, cannot refer to future events, so both forms do not really match.
Tab. 2: Kriyol paradigm for the three core forms.

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N (na/ta) kanta</td>
<td>no (na/ta) kanta</td>
</tr>
<tr>
<td>2</td>
<td>bu (na/ta) kanta</td>
<td>bo (na/ta) kanta</td>
</tr>
<tr>
<td>3</td>
<td>i (na/ta) kanta</td>
<td>e (na/ta) kanta</td>
</tr>
</tbody>
</table>

The numerical discrepancy between Portuguese and Kriyol is therefore far from spectacular and it results entirely from the absence of a subjunctive mood in the latter, not from a loss in TMA and agreement features. In fact, the change from Portuguese cantou to Kriyol i kanta to express ‘s/he/it sang’ would be quite similar to the change from Classical Latin cantabit to Portuguese cantará ‘s/he/it will sing’ through Late Latin cantare habet (Lehmann 1985), were it not for the likely fact that cantou (and the other inflected forms) were at no time part of the input for the transitional system that finally led to Kriyol, which probably only included uninfl ected all-purpose forms such as mi canta(r) (see Kihm & Rougé 2013 and below). This is of course a reminder that creole emergence, although it also involves grammaticalization processes like does ‘ordinary’ language change, should not be confused with the latter.

Moreover, as pointed out by Lehmann (1985), cantará is just as much a synthetic form as its forebear cantabit (a pungent argument against any finalist view of language change). In both cases, subtracting the fusional exponents -ará and -abit (i.e. the theme vowel /a/ plus tense-cum-agreement /rá/ and /bit/) leaves a form /kant/ that can be called a ‘root’ but has no grammatical actuality. Kriyol kanta, in contrast, is a grammatical form, namely the Imperative ‘sing!’. The person-number subject proclitics N, bu, etc. (which I prefer not to call ‘pronouns’) and the markers na/ta are what Lehmann (1985) calls ‘agglutinative affixes’ (and Spencer 2005 ‘inflecting clitics’), meaning that, as already pointed out, i na kanta may be considered as synthetic as cantará, differing from it only in the morphological identity of the exponents. Like markers, Kriyol-style person-number clitics have no Lid, they are the exponent of morphosyntactic (agreement) features associated with a given lexeme. This is summarized in Table 3, where auxiliaries, markers (including clitics) and affixes are compared for the features Lid and Fusion. Being pure exponents (no Lid, no reference) and easily segmented (no fusion), markers thus look like ‘ideal’ functional morphs.

---

8This is not generally true. For instance, cantou must have been part of the system out of which Korlai emerged (see below).
Tab. 3: Markers as ideal functional morphs.

<table>
<thead>
<tr>
<th>Auxiliary</th>
<th>Marker</th>
<th>Affix</th>
</tr>
</thead>
<tbody>
<tr>
<td>lid</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Fusion</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 3 therefore implies a second rule block for person-number (agreement) values, whose first rule is given in (8) where \(X_{VI}\) is a verb form resulting from functional application of Block I:

(8) Block II, RE 1: \(X_{VI} \& \{\text{AGR 1SG}\} \leftrightarrow N=X\)

And the same for the remaining agreement values. The total paradigm function for the core system of Kriyol verbs is then as in (9):

(9) \(pf: \{II \{I \{X_{V}\}\}\}\)

whereby Kriyol verb forms result from the successive functional application to a verb stem of Rule Blocks I and II.

All eighteen \((3 \times 6)\) forms in Table 2 can be given an anterior meaning by means of the postposed marker \(ba\) (Casamance \(ban\)) thus multiplying their number by two (e.g. \(N na kanta-ba\) ‘I was singing’). Comparison with Portuguese is no longer so straightforward, however, for morphosyntactic and semantic reasons.

The first issue is whether \(ba(n)\) is really a TMA marker analogous to \(na/ta\) since it may be separated from the verb by a complement. For some speakers, the complement may only be a clitic object pronoun as in (10), and it must appear between the verb and \(ba(n)\) (*\(N kanta ba l\)).

(10) \(N 1\text{sg.s} kanta-l -ba.\)

\(1\text{sg.s} \text{ sing-3sg.o} \text{ ant} \)

‘I had sung it.’

I assume these speakers analyse such forms as involving mesoclitics (tmesis), analogous to Portuguese \(cantá-lo-ei\) ‘I will sing it’, so \(ba(n)\) in their grammar may be considered a suffix or an enclitic (this is actually rather immaterial).

Many speakers, however, find it acceptable to insert an object NP or even an adjunct between the verb and \(ba(n)\)^9:

^9I am grateful to Jean-Louis Rougé for these examples.
(11) N tene bu foto tudu ba.
    1SG have 2SG.POSS picture all ANT
    ‘I had all your pictures.’

(12) I tarbadja na un banku ba.
    3SG.S work in a bank ANT
    ‘He had worked in a bank.’

Nonverbal predicates are modifiable by ba(n) as well:

(13) Pedru i marinheru ba.
    P. 3SG.S sailor ANT
    ‘Pedru was a sailor.’

(14) Si i ka sin ba...
    if 3SG.S NEG so ANT
    ‘If it wasn’t so…’

And in complex predicates involving several verbs, ba(n) may appear at the end of the series even though it actually scopes the first verb in the series, as in (15) where ba directly modifies txiga ‘manage’ (although it can also be viewed as modifying the whole complex event ‘manage to see’):

(15) E gaja li i ten un kau ku N txiga di
    this girl here 3SG.S have a place that 1SG.S arrive of
    odja-l-ba nel.
    see-3SG.O-ANT in.3SG
    ‘This girl, there’s a place where I had managed to see her.’

It would be tempting to assume that ba(n) is not an inflectional exponent for these speakers, but maybe a time adverb meaning something like ‘before’. Note, however, that these same speakers also produce, seemingly in free variation, V-ba(n) XP constructions: e.g. N tene-ba bu foto tudu instead of (11) and I tarbadja-ba na un banku instead of (12). The size of the intrusive XP is limited: if an object NP, a bare noun or, at most, a noun with one or two modifiers such as bu foto tudu ‘all your pictures’; if an adjunct, a minimal PP such as na banku ‘in a bank’.

Instead of and meaning the same as (13) and (14), one frequently hears Pedru i yera-ba marinheru and Si i ka yera-ba sin, where yera (< Portuguese era ‘was’) is the copula’s allomorph in anterior contexts. The alternation
is sociolinguistically grounded: (13) and (14) are reputed to be kriyol fundu ‘deep Kriyol’, so using them is partly a shibboleth of older and/or more or less illiterate people, but it must also have stylistic overtones still in need of investigation insofar as younger people with a school knowledge of Portuguese do use them as well.

This complex situation actually reflects the devious history of the marker. Peck (1988) and Rougé & Kihm (2008) argue for a double origin, from kaba ‘to finish’ (< Portuguese acabar) and from -va, the Portuguese 1st/3rd person singular indicative imperfect ending for 1st group verbs (cf. cantava ‘I – s/he/it sang’). The former must have been present as soon as the pidgin period (see below), and its grammaticalized use in a kind of serial verb construction would account for examples (11)-(14). The latter then entered stage during the vernacularization period, which started a phonological and functional competition with already available (k)ba10. In Cape-Verdean the competition ended with ba being fully reanalysed as a tense suffix that cannot be separated from the verb (Baptista et al. 2007). Kriyol ba(n), in contrast, having conflated both its sources, retained the conflicting selectional and positional properties it inherited from them, possibly as a consequence of continued contact with Atlantic languages (Manjaku and Balanta in particular) in which ‘finish’ sounds ba or ban and is used very much in the same way as Kriyol ba(n) is in (11)-(14) (Rougé & Kihm 2008).

Assume then that Kriyol ba(n) is governed by two alternating grammars (under poorly understood sociolinguistic conditions). In one of them, as already mentioned, it is a suffix with two special properties: it forces tmesis of object clitic pronouns (e.g. N kanta-l-ba ‘I had sung it’, N txoma-u-ba ‘I had called you’, and so forth); it can attach to the last verb in a complex predicate as shown by (15) — and note this is also a possibility in Sotavento Cape-Verdean: cf. N kreba papia ku bo vs. N kre papiaba ku bo, both meaning ‘I wanted to talk to you’ (Jürgen Lang p.c.).

Not taking object pronouns into account for simplicity’s sake, we can therefore safely add the eighteen anterior forms to the stock of Kriyol inflected forms, thus reaching a total of thirty-six forms. The question now is what to compare these anterior forms with in Portuguese. With bare anterior forms all depends on the verb’s semantics. If the verb is dynamic as in N

---

10The change from kaba to ba probably involves an intermediary stage */kba/ rather than wholesale deletion of the first syllable: cf. Cape-Verdean ma ‘that’ < */kma/ < coma. Note that nongrammaticalized kaba meaning ‘to be finished’ was kept entire.
kanta-ba, the closest Portuguese equivalent is the pluperfect, simple (cantara) or periphrastic (tinha cantado), all meaning ‘I had sung’. Anterior stative forms such as N kuda-ba kuma... ‘I thought that...’, on the other hand, correspond to Portuguese imperfects (Achava que... ‘I thought that...’). Now these same imperfects match rather closely with Kriyol anterior habitual forms: compare N ta kanta-ba and cantava both translatable as ‘I used to sing’. Finally, Kriyol anterior progressive forms (e.g. N na kanta-ba ‘I was singing’) are best rendered into Portuguese as a simple imperfect again or, more explicitly, via the periphrastic progressive estava a cantar ‘I was singing’.

Three Portuguese tenses thus offer themselves for comparison: simple imperfect, imperfect progressive, and pluperfect. Since both paradigms of the latter are fully equivalent in meaning, we shall retain only one, say the simple one as it may be the oldest. As in the present, each tense occurs in two moods, indicative and subjunctive. This gives us six paradigms with six forms each, hence thirty-six forms to add to the already known thirty-six, that is seventy-two forms.

Remarkably the 2:1 ratio is preserved, confirming that the main cause of the numerical difference is the absence of a distinct subjunctive mood in Kriyol.

As mentioned in Table 1, Kriyol verb inflection also comprises two periphrastic futures in addition to the progressive, which often receives future interpretation: a perfective (punctual) future (PFV.FUT) (na) bin V exemplified in (16) and an imperfective (nonpunctual, continuative) future (IPFV.FUT) (na) ba ta V exemplified in (17) (also see Kihm 1994: 108-117). In nonmatrix clauses, na is usually dropped as shown in (17)\(^\text{11}\):

(16) **Es na bin ntxumbanta-N.**
this.one PROG PFV.FUT fail.exam.CAUS-1SG.O
‘This one will make me fail my exam.’

(17) **Gosi N na ba ta yanda ku kuidadu**
now 1SG.S PROG IPFV.FUT HAB walk with caution
anti di no bin kontra.
between of 1PL.S PFV.FUT meet

‘From now on I’ll be going around cautiously before we meet [so we don’t meet].’

\(^{11}\)Both examples are from Manuel Júlio’s comic books: (16) from *Ntori Palan, falsu pursor di filosofia* (‘N.P. fake philosophy professor’; (17) from *Dotor Po sen diploma: falsu mediku di veterinaria* (‘Dr. Po without a diploma: fake vet’).
As an autonomous verb, *bin* means ‘to come’. *Ba* is a reduced form of *bay* ‘to go’. In fast speech, *na bin* may be realized as */nin/*, *na ba ta* as */nɑːtɑː/*. If we count (as we may) subordinate *bin V* and *ba ta V* forms as distinct from matrix *na bin V* and *na ba ta V* forms, and discounting object clitics for the moment, twenty-four (4 × 6) forms are thus added, for a total of sixty (36 + 24) forms.

Portuguese is also rich in future forms, but they do not match easily with their Kriyol counterparts. In particular they never seem to contrast in terms of perfectivity (punctualness), but rather in terms of the speaker’s commitment or belief concerning the actual occurrence of the future event (Cunha & Cintra 1987: 457-459). However that may be, it seems fair to only retain two Portuguese forms for comparison: the simple future (*cantarei* ‘I’ll sing’) and the *ir-V* periphrasis (*vou cantar* ‘I’m going to sing’), since the *haver-de-V* periphrasis (*hei-de cantar* ‘I am to sing’) is a modal rather than a tense form. These two tenses do not really have subjunctive counterparts (the so-called subjunctive future is only used in conditional clauses, whereas the periphrastic future is apparently never used in the subjunctive). Hence twelve forms to add to seventy-two for a total of eighty-four.

The ratio has decreased further, from 2:1 to 1.40:1. Now we must be clear that such a nice proportion will not maintain itself as we continue to enter more cells into the respective paradigms. The total paradigm of a Portuguese 1st group verb such as *cantar* numbers 4110 cells, including all finite and nonfinite forms and all possible collocations with object enclitics. The total paradigm of a Kriyol verb, also taking object enclitics into account and including negated forms, since the predicate negation *ka* may reasonably be analysed as an inflectional proclitic, numbers 1320 forms. The overall numerical ratio from Portuguese to Kriyol therefore settles at roughly 3:1.

Two comments must be made about these bare figures, however. First, three times more implies an entirely different order of magnitude than the traditionally alleged much versus nothing or next to nothing. Kriyol is not a morphologically impaired language, it is just not exuberant in this department. And it certainly does not sit on the lowest rungs of the morphological richness ladder. I am confident this conclusion can be extended to all creole languages and to expanded pidgins as well (*‘pidgincreoles’* in Bakker’s 2008 terms).

Secondly, and to repeat, such a numerical decrease does not in any way result from a loss in the amount of primary distinctions in terms of

---

\[12\]It is unclear whether collocations with object proclitics should be counted as paradigms cells.
tense, aspectuality, person-number features of the actants, and even voice, since Kriyol avails itself of two voices, a passive and a causative. What got lost in the transition process from Portuguese to Kriyol, as we saw, is the easily dispensable mood contrast indicative vs. subjunctive, as well as the conditional as in *cantaria* ‘I would sing’, *teria cantado* ‘I would have sung’. Here again, however, one must be accurate. The functional load of the indicative-subjunctive contrast did not disappear as it was carried over to the bare form, one of whose properties is to be the only admissible form in certain types of embedded clauses — precisely these types that require the subjunctive in Portuguese. What Kriyol does not show are just the inflections specifically distinguishing one mood from the other. Likewise, conditionality is taken care of by combinations of aspectuality and anterior tense (Kihm 1994: 94-95).

Table 4 gives a summary of Kriyol’s eight ‘tenses’, abstracting away from agreement and voice.

Number of distinct forms or paradigm extension is therefore not a relevant criterion to discriminate Kriyol (or, I assume, any creole) from the lexifier. Of course, it takes a particular theory of morphology, one that does not restrict the domain to fusional affixation and ablaut/umlaut, to reach this conclusion. In that sense, the conclusion is as always dependent on the theory — actually a good thing as it is becoming increasingly clear that realizational approaches fare better than their ‘morphemic’ competitors in terms of descriptive adequacy and cognitive plausibility.

What is then the relevant criterion? In the following section I try to demonstrate that predictability is the dimension that really differentiates Kriyol from Portuguese inflection.
3 Comparing Kriyol and Portuguese for predictability

Predictability can be defined as the proportion from the number of forms that have to be listed — i.e. stored in the lexicon and assumedly memorized as such — to the number of forms that can be inferred and possibly constructed from the listed forms (Blevins 2006). If few listed forms allow one to guess many or all the other forms, predictability is strong or — to put it in information theory terms — entropy is low because the system is tightly organized. If, on the contrary, many forms have to be listed and few are inferrable, this makes for a relatively opaque system with high entropy. Richness and predictability are therefore two independent dimensions.

Portuguese verb inflection is a classroom example of a system that is both rich and opaque. Opacity is primarily due to the existence of three inflection classes or conjugations characterized by three stem vowels: /a/ for the 1st conjugation, /e/ for the 2nd conjugation, /i/ for the 3rd conjugation. The shape of many exponents for given TMA values depends on which conjugation a verb belongs to: cf. 1st conjugation cantei ‘I sung’ vs. 3rd conjugation parti ‘I left’. Only knowing the forms which show the stem vowel allows one to predict these exponents. For instance, 1sg indicative present parto ‘I leave’ does not reveal the verb’s membership in the 3rd conjugation as 1st conjugation canto ‘I sing’ and 2nd conjugation temo ‘I fear’ exhibit the same ending. On the other hand, 2sg partes ‘you leave’ and 3sg parte ‘s/he/it leaves’ do exclude 1st conjugation (cf. cantas ‘you sing’, canta ‘s/he/it sings’), but they leave open the possibility of 2nd conjugation (cf. temes ‘you fear’, tene ‘s/he/it fears’). Only 1pl partimos ‘we leave’ (in addition to the infinitive partir) is diagnostic. In other words, to be able to conjugate a Portuguese verb one need to know its so-called ‘principal parts’, the number and identity of which may vary depending on inflection class or even individual items — brute memorization always remaining a possibility of course.

Nonpredictability also comes from the need to list several stems for every verb. Even so-called ‘regular’ verbs such as cantar do not escape it: 1sg indicative present canto ‘I sing’ implies /kant/ as a stem, if the TMA exponent is /o/ ; 2sg cantas ‘you sing’ implies /kanta/ as a stem since the TMA exponent has to be /-s/. In the simple past, /kant/ appears as a stem in 1sg and 3sg cantei ‘I sang’ and cantou ‘s/he/it sang’, /kanta/ in the other persons: cantaste

---

13Being deducible does not entail not being listed as well, especially if one is dealing with frequent forms (Baayen et al. 2003). The argument is not affected by this.
‘you sang’, cantámos ‘we sang’, cantaram ‘they sang’. Future cantarei, cantarás etc. is built on a stem /kan’tar/ identical to the infinitive.

Things are even more complex with irregular verbs of which Portuguese is replete. For instance, ferir ‘to wound’ shows /fir/ as a stem in 1SG indicative present (firo ‘I wound’) and in every person of the subjunctive present (fira ‘that you wound’, firas ‘that you wound’, etc.), but /fer/ elsewhere: cf. feres ‘you wound’, feriu ‘s/he/it wounded’, etc. Trazer ‘to bring’ has four stems: /traz/, /trag/ (e.g. trago ‘I bring’), /tros/ (e.g. trouxe ‘I brought’) and /trar/ (e.g. trarei ‘I’ll bring’). Very irregular verbs show fully suppletive alternative stems: cf. vou ‘I go’ vs. irei ‘I will go’, sou ‘I am’ vs. era ‘I was’.

Forms involving object clitics give rise to idiosyncratic morphophonological phenomena. In cantá-lo(s)/la(s)-ei ‘I’ll sing it/them’, for instance, the stem /kan’tar/ loses its final /r/ and the direct object clitic o(s)/a(s) (cf. canto-o(s)/a(s) ‘I sing it/them’) appears as lo(s)/la(s). Only the 3rd person direct object clitic triggers such stem changes: cf. cantar-lhe-ei ‘I’ll sing to him/her’, cantar-te-ei ‘I’ll sing to you’. The homophonous definite determiner o(s)/a(s) ‘the’ in the same phonological environment does not vary in this way: cf. Quero cantar a cantiga ‘I want to sing the song’, not *... cantá-la... 14.

Total predictability is the rule for Kriyol verb inflection, in contrast. Verbs present one stem only, which shows up unmodified in the perfective bare form: cf. kanta ‘to sing’ (N kanta ‘I sung’), kume ‘to eat’ (N kume ‘I ate’), sibi ‘to go up’ (N sibi ‘I went up’). Unlike in Portuguese and other Portuguese-related Creoles such as Korlai (Clements 1996, 2007 ; Luís 2008a), the final vowels /a/, /e/, /i/ (and /u/ in a few verbs of non-Portuguese origin) do not signal different inflection classes, but they are part of the root, itself nondistinct from the stem. It is therefore enough to know this root-stem-verb form as well as the TMA and voice exponents to predict all cells of the global paradigm.

Clitic attachment raises no difficulties. There is one small hitch with the few verbs that end in /j/, e.g. pui ‘to put’ (< Portuguese põe ‘s/he/it puts). When followed by a nonsyllabic clitic, the stem loses its final /j/: cf. I pu-N ‘S/he/it put me’, N pu-l ‘I put her/him/it’. This can be considered a phonological adjustment, however: standard orthography notwithstanding, there are good reasons to analyse Kriyol /j/ as a consonant, here in coda position (Kihm 1994). Keeping it in front of N and l would therefore yield complex codas /jn/ and /jl/ which Kriyol phonotactics do not tolerate. Causative formation might be another source of relative opacity. Comparing kambanta ‘to make cross, to ferry’ (kamba ‘to cross’) with intrindí ‘to fill up’ (intri ‘to be full’), one sees that the causative exponent is a suffix consisting in a prenasalized voiced or unvoiced coronal /nd/ or /nt/ and a vowel. The vowel copies the stem’s last vowel, whereas the coronal shows the opposite value for voicing than the onset of the stem’s final syllable. This can be formalized as a template as in (18):

14This is precisely what happens in Portuguese’s sister language Galician, however.
Theories of morphology and theories of creole emergence. . .

Morfologia e emergência das línguas crioulas. . .

61

(18) Template for Kriyol causative verbs

\[ <\ldots C_{\text{avoice}}V_i\text{-NT}_{\text{avoice}}V_i> \]

The question now is whether such a template or any kind of rules that would generate it can reasonably be assumed to be part of the speakers’ active competence. Or are causative forms learned and memorized as such, with template (18) no more than a linguist’s artefact? It is an empirical question since the answer largely depends on whether causative formation is really fully productive modulo semantic adequacy as argued by Biagui (2010). If Biagui is right, that is if new causatives are being coined that conform to template (18), then causativization, despite its relative complexity, does not depart from the system’s full predictability since there are no exceptions to template (18).

Apart from the absence of stem alternations and more or less opaque morphophonological phenomena, Kriyol predictability is to be explained essentially by the overwhelmingly analytic (periphrastic) or semi-analytic (inflecting clitics) character of the system as against the no less overwhelmingly synthetic (affixal and fusional) character of Portuguese verb inflections. Here, I believe, is the difference that matters: not so much that Kriyol verbs show less distinct forms than Portuguese verbs do — that is a fact, but not a really meaningful fact — but that most distinct forms it does show are (semi-)analytically rather than synthetically distinguished.

In Table 3, for instance, we see that only two forms out of eight (25%) are or may be considered synthetic: the perfective bare form (i) since it conveys perfectivity by contrasting with non-bare forms, not via some ‘zero’ marker; and the anterior form (iv) in the grammar where \textit{ba} is a suffix. In Portuguese, 100% of the corresponding forms are synthetic provided we do not take into account the progressive periphrasis (always replaceable by the simple present), the perfect (not equivalent to the Kriyol perfective), and the periphrastic futures (since the simple future is semantically compatible with both Kriyol periphrastic futures). Such figures should probably not be taken too seriously, but they certainly are indicative of a tendencial process.

That (semi-)analytic forms involving elisis or periphrasis are more predictable (less opaque, less entropic) than synthetic forms need little elaboration. What must be explained, however, is why such a transition from the synthetic to the (semi-)analytic inflection mode for a substantial part of the system — but rarely or perhaps never for the whole of it — should be a necessary ingredient of creole emergence, not only in the case of Kriyol with respect to Portuguese, but generally, for all creoles in comparison to their lexifiers. (It is indeed a historically explicable but linguistically accidental fact...
that all creole lexifiers, not excluding English or Bantu, show predominantly synthetic morphologies.)

I will now try and provide the required explanation by looking at the conceptual and empirical connection between the WP approach to morphology illustrated in the present section and the emergence of creole forms in a process, a necessary and, I believe, essential component of which is unguided SLA by adults, whatever other factors may come into play. In the next section, I show that assigning such a central role to SLA indeed supports the WP view of inflectional morphology. Then I argue that, given such a view, the way inflectional morphology lives on in creole languages follows straightforwardly from the SLA episode.

4 SLA and WP

Paradigms are one of the two basic elements of WP as we saw. Words or, more accurately, word-forms are the other. Whatever elements can be isolated below word level (so-called ‘morphemes’) only exist at first in the capacity of abstractions from the relations between word-forms within the same or related paradigms. Doing the abstraction job may be the linguist, or it may be the language learner and user who notices the formal regularities and differences associated with semantic regularities and differences.

Given this, the tasks facing children acquiring their native language and adults learning a second language without guidance are very much the same: to memorize word-forms and their denotations and to detect which forms are variants of the same lexeme differently specified and which forms represent distinct lexemes — to find out, for instance, that foot and feet constitute one lexeme, whereas boot and beet entertain no such relation; or that Portuguese livros ‘books’, gatos ‘cats’, mesas ‘tables’, and so on and so forth share the property of being plural count nouns, and that their plurality happens to be associated with the final /s/ they have in common, which is not heard when the reference is to single exemplars of the denoted entities (um livro, um gato, uma mesa). Confronted with such facts, the cognitive faculties or ‘instinct to learn’ of learners of Portuguese will allow, may force them to subconsciously infer an informal rule such as ‘if it’s plural, it ends in /s/, and if it ends in /s/, it’s plural’ and apply it to any new items they encounter knowing they denote sortable entities.

Of course the example chosen is elementary. Even in a simple language like Portuguese there are many snags, for instance that the final /s/ of lápis ‘pencil’ does not mean plural (as there is no *lápi); or that the plural of cão ‘dog’ is cães, not *cãos, and so forth. In languages with more complex inflections (say
Russian), learners will have to take into account more information (including morphomic properties such as declension class) than just number to achieve mastery of plural marking. In still other languages, Kriyol is a good example, plural marking is easy, but the semantic and pragmatic conditions determining whether a noun must be pluralized or not go much beyond mere cardinality of the referent (Kihm 2007). Yet, the basic gestaltist process of extracting variation from a background of uniformity remains the same.

Memorizing words and their relations is not all there is to learning a language, naturally. Given the aims I am pursuing, however, it is what I want to focus on. Moreover, it is a very important part of the learning process. After all, words and word collocations are what unguided learners hear. It is their primary foothold.

Research in first language (L1) acquisition shows that assembling paradigms — which may or may not conform to the end product of adult grammar — is an essential part of what children do when acquiring the morphology of their native language, and they are especially good at it (Pinker 1984; Courtney 2000; Yang 2002; Krajewski et al. 2011). I quote below the conclusion of a study of the acquisition of French verbal morphology (more precisely the passé simple) by young native francophone children (Labelle & Morris, p. 28):

Two crucial points emerge from the present study. Firstly, phonological factors play an important role in the participants’ production: the phonological form of the stem, the syllabic structure of the stem and the choice of stem given the vocalic nature of the ending all explain some aspects of the children’s creative analogies. Secondly, learners take into account a variety of paradigm cells in order to compute the PS [passé simple]: the present, participle, and infinitive are used to compute the ending, and the imparfait is used to compute the stem. Therefore learners are sensitive not only to phonological factors, but also to paradigmatic structure.

There are two crucial expressions in this quote. One is ‘creative analogies’. Children seem to be very apt at them, adults very much less so (at least as far as language learning is concerned). The other is ‘paradigmatic structure’, to which children are ‘sensitive’ and across which they pursue their creative analogies.

What if paradigmatic structure was an unessential byproduct as argued by proponents of morphemic approaches such as Lieber (1992) or Halle & Marantz (1993)? According to these authors, sound pieces such as /s/ are
lexical items just like /kʰæt/ (elements of List I for Marantz 1997), so children would face the task of acquiring the former (and its association to the meaning ‘plural’) in the same way as they acquire the latter (and its link to a furry, miaowing and scratching signifié). The obvious rub is that children do hear /kʰæt/, but they never hear /s/; they hear /kʰæts/ and understand it means the same as /kʰæt/, except there are more of them.

I do not think any proponents of the morphemic approach ever pushed their own argument to the absurd extreme of maintaining that children acquire morphology in such a ‘distributed’ fashion. They know that, of course, children have to somehow tease out morphemes from full word-forms. Having granted this, however, they have (or ought) to acknowledge, firstly, that paradigms are not to be pushed out of the picture, and secondly that, should their model be endowed with some measure of psychological reality, morphology would be almost unlearnable — unless children were born with an innate knowledge of all humanly possible functional categories, for which they would then search the primary data (more or less Bickerton’s hypothesis about first generation children in creole-emergence situations, except that they do not find the categories in the data). Not only is such an assumption very unlikely given what we now know of first language acquisition (e.g. Clark 1993, 2002, 2008; Tomasello 2009), but even then the search could only be conducted within some paradigmatic space. There remains little doubt, therefore, that the paradigmatic organization of the lexicon and the correlated human propensity (which may well be called an ‘instinct’), especially keen in children, to look for regularities among variations are of paramount importance in first language acquisition, at least as far as morphological phenomena are concerned.¹⁵

Compared to children, adults come to the language learning task with two handicaps. First, but for a few exceptional individuals, their general capacity for ‘creative analogy’ is but a shadow of what it was during childhood. (Not that adults are not capable of analogy, but it is most often not of the creative sort, as it generates confusion rather than novelty.) Secondly, and as a correlate of this shortcoming, adults seem to be much less sensitive than children to paradigmatic structure. In the absence of explicit tuition of the ‘L-as-a-foreign-language’ type, even such simple paradigmatic relations as that of cat vs. cats may remain unperceived — for a long time and without much serious inconvenience since, after all, plural marking is so often futile because

it is redundant (cf. ‘many cats’), or implicit in the context, or immaterial (cf. ‘I love cats’).

This is where nongrammatical factors come into play. Unhandicapped children come into the world with an ‘instinct to learn’ (Fitch 2011) and a biological urge to fully integrate the human group that surrounds them and whose protection they depend on for their survival. Perfect acquisition of the group’s idiom is one aspect of this integration, so essential that it would succeed even in the absence of a specific faculty of language.

In contrast to this idyllic (for a time) state of affairs, adults trying to learn a second language without formal tuition confront the task at an age where their instinct to learn has waned and, more often than not, in a not exactly benevolent mood given the most frequent causes (emigration, deportation) that led them to such a situation. (I leave aside the situation where plurilingualism, often already beginning in childhood, is part of the normal state of affairs, as it is not directly relevant to the present discussion.) The will to integrate, hence the incentive to learn the new language as well as possible in order to emulate the native speakers may therefore be weak, especially in such situations as deportation into slavery, where it would be doomed to failure for the majority in any case. Add to this that the native speakers themselves usually care little whether the newcomers speak correctly or not, when they do not raise effective (although not necessarily conscious) barriers to prevent them from doing so, for instance by never correcting mistakes or by pointedly addressing them in a foreigner talk variety.

Combining these socio-cultural hindrances with the linguistic-cognitive problems explains why unguided SLA by adults, especially when a collective phenomenon following a migration or deportation wave, generally leads not to the final mastery of the ‘target language’ (at least in the first generation), but to the formation of more or less stable approximative varieties or ‘interlanguages’ (Selinker 1972; also see Plag 2008), called ‘Basic Varieties’ (BV’s) when they achieve a significant degree of stability.

The classic study of the grammatical properties of BV’s is Klein & Perdue (1997). The driving factor behind these properties, which distinguishes interlanguages from the acquisition stages in children, is that adults, having already mastered at least one language, implicitly know what must absolutely be expressed for everyday communication to be effective (and, if possible, satisfactory). In the situations where BV’s emerge, where discussing labour conditions or how to procur food is on the agenda, not Kant’s theory of knowledge, such a minimum will suffice. And experience shows this minimum to be really minimal as far as morphosyntactic features are concerned: nearly all notions that morphosyntactic exponents mark in words, such as number,
case, gender, tense, aspectuality, and so forth, can be discarded, left implicit or entrusted to independent items like ‘a lot’ or ‘yesterday’, without harm done to practical communication (This is to the point that one seriously wonders why they came about in the first place, why language became complex. But that is another discussion, for which Carstairs-McCarthy 1999 is a recommended reading.).

Owing to this parsimony of effective communication, BV speakers feel no pressing motive to try and overcome their learning limitations, in particular the trouble they, like all adults, have with regrouping apparently unlike words into paradigms. As a result, lexical items are overwhelmingly invariable in BV’s: /kʰæt/ says CAT, no matter whether the referent is unique, multiple or generic; /drɪŋk/ means DRINK, whatever the time location of the event may be; and so forth. It is therefore a fact that, because of paradigm blindness and everything that concurs in not curing it, inflectional morphology (and productive derivation) strongly tends to fall away in BV’s.

But notice that the account goes through only if paradigm recognition is the only avenue to feature retrieval as assumed by the WP approach. Indeed, if functional morphemes such as plural /s/ were part of the native mental lexicon of the unguided learners together with denoting items such as cat, it would be harder to grasp why these learners are not more successful than they actually are in fishing out such morphemes among the lexical items they memorize. If they feel no difficulty in realizing that English cat is the equivalent of their own native gato (assuming their L1 to be Portuguese), why don’t they just as easily take in the same equivalence between English /s/ and Portuguese /s/ if both are lexical items like cat and gato? Moreover, partial homophony should help in the occurrence. Yet, SLA studies show absence of inflection or erratic inflection to be a general trait of early interlanguages and BV’s irrespective of the learners’ first languages.

Of course, it does not prove that morphemic approaches are wrong. Perhaps functional morphemes are lexical items. But then they have to be special lexical items, since unguided L2 learners cannot readily access them, whereas they have no such trouble with denoting lexical items like cat or passport. WP approaches account for this discrepancy without extra assumption.

Now BV’s, although relatively stable, are still interlanguages, therefore destined to perish. They end in either one of two ways. One, clearly the most frequent, is that their users finally master a good approximation of the target.
language and/or (if they settle) that their children become native speakers of the same. The other way out is creole emergence, a rare event.

Here I wish to be quite clear, as this point has given rise to heated debates (e.g. DeGraff 2002), unnecessarily I feel. Creole emergence occurs rarely not because creoles are exceptional qua languages; not even because the factors that trigger the emergence are in any way exotic. It occurs rarely merely because the conjunction of these triggering factors, none of which is sufficient by itself, is a rare event. Although creole emergence per se is not the subject matter of this article, it is not out of place at this point to insist that all we know about the diachrony of attested creole languages shows that it took a once-in-an-era concourse of historical circumstances — in a nutshell what we eurocentrically see as the ‘Great Discoveries’ and their aftermath — for such conditions to be met that entirely new languages rapidly emerged from contact situations that should ordinarily have evolved towards the above-mentioned run-of-the-mill outcomes.

Creole emergence is characterized by a considerable expansion of expressive resources in comparison with the BV stage. According to the SLA account, the primary agents of this expansion are BV speaking adults as, owing to the special circumstances that give rise to creoles, they find themselves in changed sociocultural conditions that require much more expressive power than a BV can provide. An essential area in need of enrichment, then, is morphology, that is the formal means for fast and concise expression of the various specifications that make utterances less ambiguous, to the benefit of both the speaker, who is then able to formulate her thoughts more precisely, and the hearer who is less exposed to misunderstandings (see Siegel 2004).

In order to avoid the latter in the present discussion, I need to point out that of course the function of morphology does not stop at making language a more effective tool for communication. It does that to an extent: it is certainly worth some inflectional exertion if it helps the hearer quickly to understand that the house is burning and there’s still time to do something about it (but fast!), which house burn does not convey with as much clarity and urgency; case marking of arguments helps be clear about who does what to whom, thereby freeing word order for pragmatic functions such as topicalization or focussing; and so forth. On the other hand, there is no doubt that many

---

16 If the contact was important in number and duration — like, say, after the Norman invasion of England — then of course the target may be modified as a result. Although related to some extent, that is a distinct issue from creole emergence, however (see Thomason & Kaufman 1988).
morphological processes serve no obvious communicative or expressive function, especially when the system is a complex one: *oxes would have drawn ploughs as well. Then, however, another, maybe more important function of (complex) morphology takes over, namely the ethological function: complex morphology makes learning the language by any other means than natural acquisition a difficult task; therefore, it contributes to the group’s identity and seclusion with respect to neighbouring groups, effectively although not to the point that all exchanges are barred. Song diversity within the same bird species plays a similar role — but for the fact that it precludes mating, which language diversity does not do, or much less.

It would be interesting to assess to what level morphological enrichment rose in different creoles. Did it exhaust itself as soon as sufficient means of expression were made available to the newly formed language community? Or did it go beyond, to the point that morphological processes exceeded their practical function and began to fulfil an ethological one? I am sure the answer to the second question is ‘yes’ for such creoles as Kriyol, Korlai or Saramaccan. I am uncertain whether any creole calls for a ‘no’ answer.

The issue now is how the enrichment proceeded. Here we are working with a fact and a robust assumption, both already mentioned. The fact is that the emerging morphology is largely (but never entirely) of the (semi-) analytic type; the assumption, that adults are primarily responsible for the enrichment. From a WP perspective, the relation is self-evident: the predominant (semi-) analyticity of creole morphology is a consequence of the adults’ endemic paradigm insensitivity. Actually, paradigm insensitivity set in at two junctures. Because of it (in conjunction with other factors), the BV first came up with almost no morphological equipment; because of it again, synthetic resources could not be retrieved (or to a limited extent) from the lexifier when the enrichment process was set in motion (see Roberts & Bresnan 2008 for what inflections were then kept).

Two circumstances then proved crucial. Firstly, as we saw, all attested lexifiers happen to use periphrastic expressions next to synthetic ones, mainly in the verbal paradigms (e.g. English I will sing, Portuguese hei-de cantar, etc.), but also, to a lesser extent, in the nominal paradigms (e.g. English bigger vs. more intelligent). Such expressions were apparently large enough and stuck out sufficiently for paradigm-challenged adults to notice them and put them to good use. Secondly, grammaticalization is an irresistible process that will fire as soon as the conditions for it are met. Thanks to it, new analytic paradigms were created to fill up the morphological (near) void of the BV and apparently substitute the synthetic paradigms of the lexifiers — apparently only, since there actually was no direct replacement.

_PAPIA_, 24(1), ISSN 0103-9415, e-ISSN 2316-2767
What the foregoing developments show, therefore, is that the SLA theory of creole emergence, that is the possibility under special circumstances for new languages to emerge out of a process of unguided ‘imperfect’ SLA by adults, supports the WP approach to morphology according to which word-forms constitute the primary lexical reality.

In order to round up the argument, I need now to show that the regular (re)emergence of at least a few and often many morphological phenomena in the emerging creoles thanks to the paradigmatic (re)organization of word-forms and various grammaticalization processes is better accounted for by a theory that crucially includes an SLA episode than by a competitor that does not. This I will attempt to achieve by first examining in some detail the origins of the morphological apparatus of Kriyol as sketched above and of Korlai as described in Clements (1996, 2007) and Luís (2008a). Both languages formed independently in widely different environments (West Africa and India) from a partially similar source, namely a BV version of Middle Portuguese. This is what makes it such a fascinating and highly revealing endeavour to assess what they have in common (not a little) and where they diverge (on many points).

5 SLA and the development of Kriyol and Korlai verbal inflections

Although we have no direct attestation of the BV on which Kriyol and Korlai were founded, we may assume it was to some extent similar to the BV called Língua de Preto (‘Black Speech’, henceforth LDP) which formed in sixteenth century Portugal among the many African slaves who had been transported there, and which we know well enough through the literature of the time (Teyssier 1959; Tinhorão 1988). The connection from Portugal to West Africa is rather easy to establish (Kihm & Rougé 2013). For India, we may assume that LDP (or something like it) reached there maybe because Portuguese ships had African slaves-sailors on board and certainly because it provided a ready foreigner talk variety to try and make contact with new aliens (see below).

LDP shows complete absence of inflection: nouns never pluralize, pronouns are not marked for case, verbs appear under a single form usually proceeding from the Portuguese infinitive minus final /r/.

---

17 The presence of African slaves aboard Portuguese ships sailing to West Africa in the second half of the fifteenth century is an attested fact. For ships sailing to India, however, it is only a speculation.
Kriyol and Korlai inherited some of these characters. Nouns and noun phrases in Kriyol are not marked for number unless there is a specific reason for doing so (suffixing /s/), which reason depends from the noun’s denotation as well as from the syntactic and discursive context (Kihm 2007). In Korlai, nouns and noun phrases are simply not marked for number at all: *ye om /DEM* man/ can be understood as ‘this man’ or ‘these men’ (Clements 2007: 167). Marathi, the neighbouring Indo-Aryan language Korlai has most contacts with, does distinguish number, at least in determiners (Clements 1996: 101), so the Korlai indistinction cannot be ascribed to contact.

Turning to verbal morphology, we see that Kriyol bare forms such as *kanta* ‘to sing’, *bibi* ‘to drink’, *kume* ‘to eat’ clearly proceed from *r*-less Portuguese infinitives (*cantar*, *beber*, *comer*) as do Korlai base forms *kata* ‘to sing’, *bebe* ‘to drink’, *irgi* ‘to get up’. Interestingly, final */r*/ per se raises no problem for Kriyol or Korlai: cf. Kriyol *katxur* ‘dog’ < Portuguese *cachorro*, *kalur* ‘heat’ < Portuguese *calor*, *dur* ‘pain’ < Portuguese *dor*, etc.; Korlai *ter* ‘earth’ < Portuguese *terra*, *kalor* ‘heat’ < Portuguese *calor*, *gadzor* ‘carrot’ < Marathi *gadzor*, etc. It is only final *r* as an infinitive marker that underwent regular apocope, which, given the LDP data, makes a strong case for regarding it as a feature retained from the original BV.

Another thing Kriyol and Korlai retained from the BV stage is lack of synthetic verbal agreement, that is affixal inflection also varying according to the person-number features of the subject as in Portuguese. This is of course not surprising since the BV, as we saw, lacked every kind of inflection. From this stage, however, two evolutive paths are opened and attested. Kriyol followed one as it grammaticalized the autonomous (*np*-like) subject pronouns of LDP which became subject agreement inflecting proclitics as argued above and as summarized in Table 5 which gives both LDP and Kriyol resulting forms.

Tab. 5: From LDP subject pronouns to Kriyol subject agreement inflecting proclitics.

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>mi</em> &gt; <em>N=</em></td>
<td><em>nos</em> &gt; <em>no=</em></td>
</tr>
<tr>
<td>2</td>
<td><em>vos</em> &gt; <em>bu=</em></td>
<td><em>vos</em> &gt; <em>bo=</em></td>
</tr>
</tbody>
</table>
| 3 | *ele* > *i=* | *eles* > *e=*

The same path of grammaticalization was followed by LDP object pronouns, as the latter had kept nothing of the Portuguese object enclitics, which Kriyol recreated using its own resources. This is shown in Table 6.
Tab. 6: From LDP object pronouns to Kriyol object agreement inflecting enclitics.

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>mi</em> &gt; <em>N</em></td>
<td><em>nos</em> &gt; <em>no</em></td>
</tr>
<tr>
<td>2</td>
<td><em>vos</em> &gt; <em>u</em></td>
<td><em>vos</em> &gt; <em>bo</em></td>
</tr>
<tr>
<td>3</td>
<td><em>ele</em> &gt; <em>l</em></td>
<td><em>eles</em> &gt; <em>elis</em></td>
</tr>
</tbody>
</table>

Korlai followed the other path keeping LDP’s autonomous subject and object pronouns, so Korlai verbs effectively do not inflect at all for agreement. I say LDP rather than Portuguese since, as far as form is concerned, the paradigm of Korlai subject pronouns aligns on Table 5 except for 1SG *yo* that seems to come directly from Portuguese *eu* ‘I’ (also encountered in LDP texts albeit more rarely) and 2PL *udzo*, perhaps from *vos outros* ‘you others’ (cf. Kristang *bolotu*, Mauritian *zot* < French *vous autres*)\(^{18}\). See Table 7. 3PL *elo* instead of expected *eli(s)* may result from analogical levelling since all pronouns but 3SG *el* end in /o/ or /ɔ/ or it may be the reduction of an old combination with *outro* similar to Kristang *olotu* ‘they’ (Baxter 1988: 53). Korlai pronoun paradigm also includes a 2nd person polite form of address use, presumably from Portuguese *você* (a sixteenth century reduction of *Vossa Mercê* ‘Your Grace’ — Clements 1996: 227). The Kriyol equivalents *nhu* and *nha* addressing a man and a woman respectively come from Portuguese *senhor* ‘sir’ and *senhora* ‘madam’, and they function like autonomous pronouns.

Tab. 7: From a Portuguese BV to Korlai subject pronouns.

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>eu</em> &gt; <em>yo</em></td>
<td><em>nos</em> &gt; <em>no</em></td>
</tr>
<tr>
<td>2</td>
<td><em>vos</em> &gt; <em>wo</em></td>
<td><em>vos (outros)</em> &gt; <em>udzo</em></td>
</tr>
<tr>
<td>3</td>
<td><em>ele</em> &gt; <em>el</em></td>
<td><em>eles</em> &gt; <em>elo</em></td>
</tr>
</tbody>
</table>

Moreover, Korlai belongs to a language type one may call ‘radical null subject’, as it allows for the identity of the subject to be left entirely implicit

\(^{18}\)Comparing with Mauritian is not an idle move, I believe. Several clues point toward an Indian Ocean sprachbund including the French and Portuguese creoles of the area, maybe even Melanesian English creoles. I hope to be able to demonstrate it in later work.
whenever the speaker feels it to be sufficiently clear from context or irrelevant (Clements 1996: 32). (Recall that Portuguese is a null subject or ‘pro-drop’ language only in the sense that its rich person-number inflection makes subject pronouns redundant, hence pragmatically loaded when used.) Radical null-subjecthood is an areal property shared by Marathi (Clements 1996: 106). Outside the Indian subcontinent Papiá Kristang also shares it (Baxter 1988: 166). On the other hand, it is also observed in BV’s (although not in the LDP texts, which does not mean it was not present in the actual speech). Areal influence and retention of a pidgin feature may therefore have conspired to promote radical null-subjecthood in Korlai and Papiá Kristang — perhaps in Zamboangueño as well judging by this example from Lipski & Santoro (2007: 388): ... mio nana talli na cueva no puede camina my mother be.LOC in cave NEG can walk ‘... my mother is in the cave (and she) cannot walk’.

One should be very cautious with sweeping generalizations in this domain. It seems safe, nevertheless, to claim that no language with inflecting clitics such as Kriyol has could share Korlai’s typology. And this is because inflecting clitics can no more be cut off from finite verb-forms than can inflectional affixes.

Both Kriyol and Korlai obligatorily mark case on pronouns. In the former, the phonological distinction between subject and object clitics in 2/3SG and 3PL has to be traced back to differential evolutions from the LDP uninflected forms vos, ele, and eles, since there clearly is no direct link between Portuguese cantei-o/a ‘I sang it’ and Kriyol N kanta-l ‘id.’, but only through reconstructed LDP *mi cantá ele. Korlai, in contrast, evolved an original system whereby pronominal object case is marked by a prefix derived from the Portuguese and LDP preposition para ‘for’ and whose abstract form may be given as/p(V)(r)-/, as shown in Table 8.

Tab. 8: From a Portuguese BV to Korlai object pronouns.

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(para) mi &gt; par(m)i</td>
<td>(para) nós &gt; pọw</td>
</tr>
<tr>
<td>2</td>
<td>(para) vós &gt; pọw (para) você &gt; puse</td>
<td>(para) vós (outros) &gt; pudzo</td>
</tr>
<tr>
<td>3</td>
<td>(para) ele &gt; pel</td>
<td>(para) eles &gt; pelo</td>
</tr>
</tbody>
</table>

Pace Clements (1996: 108), Marathi influence seems unlikely here. Object case marking in Marathi is by way of a suffix as in tu ‘thou’ vs. tula ‘thee’. Since there was nothing in the lexical input for Korlai that could match this suffix, Marathi pronominal inflection would probably have passed unnoticed in the language contacts that contributed to shaping Korlai. On the other hand, it is difficult to ignore the fact that object pronouns (and NP’s with
animate denotation) are marked with *per* in Mediterranean Lingua Franca (Schuchardt 1909). Mediterranean Lingua Franca, does not seem to have had any impact on the Portuguese pidgin that gave rise to Kriyol and West African Portuguese creoles generally, despite Naro’s (1978) arguments to the contrary. There is indeed no reason why it should have if one accepts Kihm & Rougé’s (2013) demonstration that these creoles directly proceeded from LDP, a Portuguese *bv* that formed entirely within the African slave community in sixteenth century Portugal.

Members of this community and consequently LDP obviously played a more marginal role in India, on the other hand. To quote from Clements (1996: 10-11), ‘The first Portuguese to arrive in Chaul [the Korlai area, in 1505] were largely soldiers, also of socially lower classes … These men knew their dialect of Portuguese, some version of Portuguese foreigner talk, and perhaps some of the lingua franca spoken around the Mediterranean in medieval times as well’ (also see Clements 1992). Under ‘some version of Portuguese foreigner talk’ I suggest we understand LDP which Portuguese living at the beginning of the sixteenth century could easily pick up in the streets of Lisbon (Kihm & Rougé 2013) and imitate to address the natives they soon had intimate contacts with. That this imitation became ‘contaminated’ by the more or less second-hand knowledge they probably had of Mediterranean Lingua Franca as well stands in the order of things.

It is noteworthy that all West African Portuguese Creoles and Atlantic Creoles more generally constrast subject and object pronouns as Kriyol does, through word order only for some items (cf. Kriyol 1sg *N*), through word order and suppletive forms for other items (cf. Kriyol 3sg *i* vs. *l*). None does it by inflecting the pronominal stem — a device for which Romance and Germanic languages offer no model, let it be said in passing. Nor do they mark a subset of object NP’s by means of a fully grammaticalized preposition such as Korlai and Papiá Kristang *ku <* Portuguese *com* ‘with’ (Clements 1996: 160; Baxter 1988: 153). Viewing things in terms of contact and substrate, it may be significant that morphological case marking of nouns and noun phrases is largely absent in the Niger-Congo phylum, to the best of my knowledge, whereas it is well entrenched in Indo-Aryan and Dravidian languages (see Clements 1996: 158 for Marathi, Smith 1979 for Sri Lanka Portuguese). However that may be, such a macro-typological difference between Western and Eastern Portuguese Creoles confirms that, at the very least, the pidginized variety that formed the original input was significantly (although not entirely) different in both areas.

More than anything else, this difference is probably what explains the rather wide discrepancy between the TMA systems of Kriyol and Korlai. Although not the most spectacular, the main difference lies in the distinct fates of
the BV-pidgin all purpose bare form of verbs. In Kriyol, as we saw, this form evolved into a default form interpreted as perfective in matrix clauses, perfectivity being then interpreted as referring to past events with dynamic verbs, but to present states with stative verbs. In Korlai, in contrast, only five verbs can be used bare as the main verbs of main clauses with a perfective (present-referring) meaning: the two psych-verbs κετε ‘to want’ and sαβ(e) ‘to know’, the copula τε, and the two modals pο ‘can’ and mασhi ‘must’ (Clements 2007: 154). All other verbs must be inflected in some way to be given their proper TMA value.

This property of Korlai verbs correlates with the fact that Korlai has two affixally inflected nonperiphrastic tenses to refer to past events: a preterite involving a /o/ or /w/ suffix depending on conjugation class: cf. yo kato ‘I sung’, yo bebew ‘I drank’, yo irgiw ‘I got up’; and a perfect involving a /d/ suffix: cf. yo katad ‘I have sung’. The former obviously proceeds from the Portuguese 3SG simple past (pretérito perfeito) form generalized to all person-number values (cf. cantou ‘s/he/it sang’, bebeu ‘s/he/it drank’, ergueuse ‘s/he/it got up’). Interestingly, the same generalization of the same form independently occurred in the Tonga Portuguese of São Tomé (Rougé 1992). The perfect results from the reduction of the Portuguese periphrasis ter-past participle (e.g. tem cantado ‘s/he/it has sung’).

We therefore have to assume that both these inflected forms were part of the BV-pidgin foundation of Korlai at the time of the first contacts with native speakers of Portuguese. That of Kriyol, in contrast, must have contained only uninflected forms, being thereby closer to LDP as we know it from the texts and farther removed from Portuguese. When Kriyol creolized to achieve the system exposed in Table 3, bare forms remained next to the new tenses. Because they were now in paradigmatic contrast with these new forms, however, they became reinterpreted as inflected by default, that is by the very absence of overt inflection. Their perfective meaning then resulted from the conjunction of two factors: (a) it was not expressed by the new tenses; (b) being completed (i.e. seen as a whole) may be considered the default interpretation for events or non-continuants (NC) defined in terms of Zemach’s (1970/1979: 74-75) first ontology of entities bound in space and time. In Korlai, on the other hand, the presence from the start of the two inflected forms referring to completed events left no room for the bare verb as a finite tense form.

19 In the 4th conjugation class of etymologically non-Portuguese verbs ending in /u/, the suffix /w/ does not show up, probably for phonological reasons: cf. yo tepu ‘I heated up’ (Luís 2008b).
Another Portuguese construction that must have been there from the beginning in India is the progressive periphrasis consisting in *estar* ‘to be’ and the gerund, e.g. *está cantando* ‘s/he/it is singing’, as it is the only possible origin of the Korlai now synthetic progressive -n form of e.g. *yo katan* ‘I am singing’. This is supported by ancient documents showing forms like *tá comprá* or *tá comprand* ‘be buying’ in the Norteiro dialects of the Bombay area at the turn of the twentieth century (Leite de Vasconcellos 1901/1987: 139; Dalgado 1906). Clements (1996: 118) mentions that ‘around 50 years ago’ the progressive form in Korlai included the copula as in *ê katan* ‘be singing’.

No trace of this construction can be found in Kriyol or the other West African Portuguese Creoles, in contrast. As we saw in section 2, the Kriyol progressive involves a marker *na* homophonous with the locative preposition (*N na kanta* ‘I am singing’). That homophony is the synchronic reflex of ancient identity is supported by the fact that, very much as in Korlai, *na V* appears to result from the reduction of a more complex periphrasis *sta na V* involving the locative copula *sta* (< Portuguese *estar*) and in which *na* may well be analysed as a preposition. This construction is still in use in the more conservative Casamance variety. Building on this, we can now relate (*sta) na V to the other Portuguese progressive periphrasis *estar a V* (*é está a cantar ‘s/he/it is singing’

In Cape Verden (Sotavento) *sta (ta) V* corresponds to Kriyol (*sta) na V* (Baptista et al. 2007: 56). The Cape Verden form is represented in Santome as *sa ka V* (often reduced to *sk a V*) and in Angolar as *θa ka V*, where *sa ~ θa = sta and ka = ta*. A propos the Angolar form Lorenzini (2007: 4) attributes a ‘gerund-like property’ to *ka V*, thereby relating it — and the Santome and Cape Verden equivalents by transitivity — to the Portuguese *estar Vndo* periphrasis. An infinitive-like property would be just as plausible, however. In fact, *ta ~ ka* happens to be the habitual marker as well, and in that capacity it also proceeds from Portuguese *estar* — or more probably *está ‘s/he/it is’ commonly pronounced /ta/ (see below). Given the complex and irretrievable lexical crossbreedings that accompany BV formation, then, *sta ta V* may just as well represent *estar a V* as *estar Vndo*.

What we have to assume in any event — as it could not have been otherwise — is that both periphrases were competing for a time at the juncture of Middle and Early Modern Portuguese. If some version of *estar Vndo* is primeval in Korlai, being present as soon as 1502, it indeed corresponds with a stage in the history of Portuguese where that periphrasis was dominant, perhaps still free from competition. The emergence of *na V* in Kriyol at least and its equivalents in other West African Portuguese Creoles possibly would then
have occurred later, at a time when *estar a V\_inf* was overcoming *estar V\_ndo* in the popular speech of Portugal.

As mentioned above, the only possible origin for the *ta ~ ka* marker of West African Portuguese Creoles seems to be *está*, which relates *ta/ka V* to *estar a V\_inf* or *estar V\_ndo* as well. What of the mismatch between the progressive meaning of the Portuguese periphrasis and the mostly habitual meaning of *ta/ka V*, however?

An interesting piece of evidence comes from the Bozal Spanish data cited in Lipski (1987: 432), which show *ta V* actually to be compatible with all aspectuality readings that include the present interval irrespective of the progressive vs. habitual contrast: cf. *yo tá murí* ‘I am dying’ vs. *mí tá sabé que tú no tá queré a la negra Yeye* ‘I know that you don’t like black Yeye’.

This suggests the construction arose in the BV itself, probably as a collocation *(es)tá V*\_inf, not directly related to the Portuguese periphrasis *está a V\_inf* (possibly not available yet at the time), but used primarily to distinguish imperfective (non completed) from perfective (completed) states of affairs\(^{20}\). Depending on the successor creoles, the habitual-progressive merger was continued or, as in Guinea-Bissau Kriyol, there emerged a special progressive form now probably proceeding from *(es)tá a V\_inf*.

Does this story apply to the Korlai construct consisting in the habitual marker *ta* followed by the base form (*yo to kata* ‘I sing’)? It is true that *ta* sounds like *ta*. But consider the past counterpart *ti* of *ta* (*yo ti kata* ‘I used to sing’). Its etymology appears rather secure: it has to come from Portuguese *tinha* ‘had’, as confirmed by the older form *tinh* (Dalgado 1900, 1906). Also note that following negation *to* gets replaced by *tε* (*yo nu tε kata* ‘I do not sing’ — Clements 2007: 162), homophonous with the copula *tε* ‘be’, itself a reflex of Portuguese *tem* ‘has’, just like the past marker *ti* is homophonous with the past form of the copula *ti*. Actually, it seems safe to assume that synchronic homophony results from past identity and that present-day *ta*, clearly a product of destressing, has *tε < tem* as its main source, with possibly *ta < está* as a secondary source (Clements 1996: 111).

The question now is how one goes from ‘have’ to ‘be’. The obvious link is the impersonal use of *tem* meaning ‘there is/are’. This is often described as Brazilian usage, whereas the standard European Portuguese form should be *há*, the 3\(^{rd}\) person singular indicative present of *haver* ‘to have’. This is not

\(^{20}\)To be honest I must acknowledge that the LDP corpus does not contain one single instance of this construction.

*PAPIA*, 24(1), ISSN 0103-9415, e-ISSN 2316-2767
accurate, however, as *tem ‘there is/are’ is in constant use in the colloquial speech of Portugal and has probably been so for quite a long time as part of a general process leading to complete replacement of *haver by *ter.

Let us assume then that Korlai to proceed from Portuguese tem meaning ‘there is/are’. Habitual constructions such as el to kata ‘s/he/it sings’ therefore originate in a BV expression we can reconstruct as *el tem canta(r), not yet specialized for a particular aspectuality value as suggested by the Bozal Spanish evidence. A comparison then immediately comes to mind, namely with the yâna V or ya V construction of so-called français tiraillou, that is the French pidgin in use among French colonial troops at the beginning of the previous century (Delafosse 1904: 265; Anonymous 1916; Valdman 1978). This yâna or ya, which was used both as a copula and apparently to signify a predicate’s present relevance (cf. mwa ya di ‘I’m saying’), is nothing but colloquial French y a /there has/ or y en a /there of.it has/ ‘there is (of it)’. Would it be too far-fetched, in addition, to suspect a further connection with the progressive marker de of several English-related Creoles (e.g. Jamaican, Krio), whose formal derivation from ‘there’ is not in dispute, although the semantic path remains uncharted?

If these comparisons make sense, they might incite us to adopt a cautious version of the monogenetic theory that goes no farther than claiming that the pidginization of English and French in West Africa conformed to a large extent to the model provided by the Portuguese pidgin that was there first (Dillard 1979; Hull 1979). The presence in this pidgin or BV of an element originally meaning ‘there is/are’ that could be used as a copula and to actualize predicates seems therefore well supported. Apparently this element survived in both these functions in India only. In the Western hemisphere, TEM (te in Angolar, ten in Guinea-Bissau Kriyol, tin in Papiamentu, etc.) just means ‘there is/are’. As a copula and/or a TMA marker it was supplanted by ta < esta(r), with which it had been competing probably from the very beginning — not excluding possible contaminations, as only the vowel discriminates the two items. Even in the East, Papiá Kristang kept tem (teng) as a copula (Baxter 1988: 181), but not as a TMA marker, in which function it was replaced by ta.

Such reconstructions are of necessity speculative. Nevertheless, the fact that Korlai to V as well probably as its past counterpart ti V proceed not from equivalent Portuguese periphrases, but from constructions that arose during the BV stage seems solidly established. A similar case is that of the future form lo V, in which lo is a reflex of Portuguese logo ‘soon’. The use of logo as a tense-indexing adverbial is attested as soon as LDP: cf. logo a mi bae trazee/soon 1sg go fetch/‘I’ll go fetch it right away’ in Gil Vicente’s Fraga de Amor (‘The Forge of Love’ — Kihm & Rougé 2013). Korlai lo, like Papiá Kristang and Papiamentu lo, represents the final (so far) grammaticalization of the item, as evidenced by its phonetic attrition and closing up to the verb. (Tok Pisin bainbai from English by and by and reduced to preverbal bai or bai shows an intriguingly similar evolution — see Sankoff 1991; Faracas 2007.)
6 Bringing it all together

To summarize, the TMA systems of Kriyol and Korlai appear to result from three processes: (a) the regrammaticalization of Portuguese simple forms or periphrases taken over in and through the BV stage; (b) the grammaticalization of constructs devised during the BV stage; (c) the paradigmaticization of all such elements.

Process (a) accounts for the Kriyol perfective (bare) and progressive (na=V) forms, as well as for the Korlai preterite (Vw), progressive (Vn), and perfect (Vd) forms. The Kriyol and Korlai habituals (ta=V, to=V, ti=V), the Kriyol anterior (V=ba), and the Korlai future (la=V) are due to process (b). Finally, paradigmaticization explains why the Guinea-Bissau Kriyol bare form evolved from the all-purpose form it was in the BV to a perfective form contrasting with two imperfective forms. The same processes are responsible for the Guinea-Bissau Kriyol agreement paradigms involving proclitics for subject agreement and enclitics for object agreement. Korlai on the contrary remained in the same no-agreement condition as the BV.

Grammaticalization as a creole-building process is often objected to on the ground that it ordinarily takes a lot of time for its effects to be felt. A possible rejoinder is that creole emergence precisely takes place in ordinary circumstances and that the kind of grammaticalization we must be talking about is ‘contact-induced grammaticalization’, a possibly more rapid process as language contact is known to be an accelerator of change (Heine & Kuteva 2003; also see Holm 2004). Yet, it is not clear how the evolutions we have just reviewed are ‘contact-induced’, that is, under common interpretation, due to interference from the L2 learners’ first languages or substrate.

Korlai’s verbal system seems to owe little to Marathi (Clements 1996: 171-172). The LDP speaking slaves taken to Portugal between 1450 and 1550 came from the whole extent of the West African coast, so their L1’s were just too many and too diverse that a search for precise interferences can be seriously conducted. LDP texts nevertheless reveal a number of features that may be attributed to a very broadly delineated substrate: in morphosyntax, nonrecognition of Portuguese gender distinctions, expected from persons whose L1’s either ignore such distinctions altogether or do not base them on a pseudo-sexual contrast; in phonology, a preference for CV syllabification and rhotacism of Portuguese intervocalic /d/ as in maruro for maduro ‘ripe’ and toro for todo ‘all’, the latter pointing towards the Atlantic languages of Upper Guinea, that is present-day Guinea-Bissau and Casamance (Kihm & Rougé 2013).

Concerning Kriyol specifically, to the extent that it fully creolized in situ (Rougé 1986), its main substrates appear to be Mandinka (Niger-Congo,
Mande), on the one hand, Balanta (Niger-Congo, Atlantic), on the other hand, and it seems to be a fact that the verbal system of Kriyol presents several significant parallels with that of Balanta (Intumbo 2007; Kihm 2012) This difference between Kriyol and Korlai is probably to be related to the different stages of elaboration of the interlanguages out of which they grew. More Portuguese forms entered Korlai than did Kriyol, thereby leaving less room for transfer from L1.

Moreover, the distinct evolutions of Kriyol and Korlai show that ‘transfer’ must be used in the sense of transfer in L2 use rather than in L2 acquisition: ‘Transfer in L2 use ... is the result of learners (or former learners) falling back on their L1 knowledge when communicating in the L2. This occurs when their knowledge of the L2 is inadequate to express what they want to say’ (Siegel 2004: 350). Having only a Portuguese pidgin similar to LDP at its disposal, the emerging community amidst which Kriyol crystallized was precisely in this situation. Proto-Korlai speakers, in contrast, availed themselves of a richer medium and must therefore have felt less need for transfer.

Considering both ends of the processes, the lexifier and the creole, one sees quite various relations between corresponding forms in the former and the latter. Sometimes a fully affixal form of the lexifier corresponds to a form involving inflecting clitics in the creole: cf. Portuguese canta ‘s/he/it sings’ vs. Kriyol i ta kanta or Korlai (el) ta kata. In other cases it is a periphrasis in the lexifier that corresponds to the same type of form: cf. Portuguese está a cantar ‘s/he/it is singing’ vs. Kriyol i na kanta or Cape Verdean i sta kanta. Or the lexifier’s periphrasis corresponds to a creole affixal form: cf. Portuguese está cantando ‘s/he/it is singing’ vs. Korlai el katan, or Portuguese é cantado/a ‘s/he/it is sung’ vs. Kriyol i kantadu. Or both the lexifier and the creole forms are affixal: cf. Portuguese cantou ‘s/he/it sang’ vs. Korlai el kato or Kriyol i kanta (inflected by default).

In every case, the lexifier and the creole forms are morphological objects definable by the same kinds of rules. Loss of morphology is therefore not a disease creole languages suffer from — which diagnostic, as we saw, requires a definition of morphology such that words and paradigms, not morphemes, are the building blocks. But in every case also the creole form is patently not a direct continuation of its source in the lexifier, but there is always a gap, an interval during which things occur that would not have occurred otherwise. This is especially clear with correspondences such as canta vs. i ta kanta in which, as argued above, the ta V construct that would emerge as the creole form must first have appeared during the interval. But it is just as true when the correspondence seems more direct, as with cantou vs. (el) kato, since the latter does not include the agreement features of the former.
The gap or interval alluded to is of course the SLA episode. The crucial fact now for creole emergence is that this episode untypically does not end with near perfect learning of the second language, so that all innovations issuing from the SLA process itself such as the ta V construct are erased. Nor, again untypically, does it end with a scattering of imperfectly acquired individual versions of the L2, which may preserve the innovations but do not elaborate on them. What it ends up with is a new, fully functional grammar where the contributions of the SLA process are preserved (or at least some of them) and elaborated (on elaboration, see Siegel 2004). Such an outcome necessarily correlates with special sociolinguistic conditions given a view of the function of grammar such as that advocated in Labov’s (1990) unsurpassed study into the development of tense in creole languages, of which I quote the following:

One might say that a developing grammar serves the need of stylistic variation. But it would be more accurate to say that grammar is style [...] On the whole, grammar is not a tool of logical analysis: grammar is busy with emphasis, focus, down-shifting and upgrading; it is a way of organizing information and of taking alternative points of view. (Labov 1990: 45, emphasis in the original)

Moreover it is consonant with Baker’s (1996) well supported position that, in creolization contexts, the ‘full’ L2 ceases at some point to be the target of acquisition (if it ever was). What the incipient creole speakers are trying to acquire and use is an already efficient interlanguage, a Basic Variety. And where they differ from usual L2 learners such as the migrant workers studied in the 1980’s in France, Germany, the Netherlands, etc. (see Klein & Perdue 1997 for references; also see Kotsinas 2001; Lowie 1998) is in not stopping at the BV or developing it in the direction of the L2.

All this is well-known and part of the creolist vulgate. What may not be so generally recognized, however, is that, for all the formal and semantic changes that took place during the interval, correspondences such as those just pointed out imply that the interlanguage that provided the input to the creole was sufficiently close to the full learning end that the original meanings of the forms retained were always compatible with the meanings finally assumed in the creole. For instance, the fact that Korlai el katan and Kriyol i na kanta proceed more or less directly from Portuguese está cantando and está a cantar, respectively, implies that the meaning of these periphrases was understood and preserved when their forms changed. This sounds like a truism, but it shows that theories such as the Language Bioprogram Hypothesis which assume
minimal knowledge of a lexifier reduced to a ‘jargon’ and assign no crucial role to adult SLA cannot be right.

It also shows that we must be as accurate as possible in our assessment of the degree of advancement of the SLA process at the point where creole emergence starts to become the actual outcome. It should not be too high or it would result in full learning or partial restructuring (Holm 2004). Nor should it be too low in any event. If the L1 mental lexicons of adult L2 learners included morphemes along with lexemes, minimal exposure would indeed presumably be enough for them to recognize that, say, final /s/ means ‘plural’ in Portuguese nouns — except when it doesn’t! Things do not happen this way, however. As we know, even after years of practice with the L2 adult learners may not yet master its simplest paradigms and go on using interlanguages where nouns are unmarked for number.

From this it follows that the pre-Kriyol stabilized interlanguage or BV or pidgin must have reached a stage where Portuguese number paradigms had been formally identified — compare Portuguese gato(s) ‘cat(s)’ and Kriyol gatu(s) ‘id.’ — but the semantic principle that governs plural marking — roughly ‘pluralize if there is more than one’ — had not been retained and was being replaced by another, not merely arithmetic one (Kihm 2007). We cannot be sure whether some TMA contrasts were already present or not, although the ta V form seems likely. In the pre-Korlai BV, on the other hand, number paradigms apparently went unrecognized, but a tense paradigm contrasting the Portuguese 3SG simple past form with at least the base form and probably more was in place.

What this comparison shows as well is that one cannot expect consistency across BV’s. They must be sufficiently elaborated to allow for the leap into creole emergence, but the elaboration pattern itself varies from case to case, not showing general implications such that, if stage x (e.g. plural marking) has been reached, then stage y (e.g. tense marking) must also be reached. This is one reason why creole languages, even from the same lexifier, differ as they do.

The crucial step then is vernacularization, to use Manessy’s (1995) fitting term. It is indeed essential that, for a variety of reasons, a particular group begins using the BV at the right stage of elaboration — not too low, not too high — as their in-group medium. (The full L2 may still be around or it may not. What is important is that the members of the group do not try to know more of it than they already do.) It is at this point that the insufficiencies of the medium will be keenly resented and have to be compensated for. Further learning of the L2 no longer being an actual and/or psychological possibility, two paths only remain open. One is spontaneous
elaboration through paradigmatization, as when the Kriyol bare form acquires perfective meaning from being enmeshed in a network including a habitual and a progressive form. The other is transfer in use from not yet forgotten L1’s or already known L2’s (not to be neglected in multilingual environments such as West Africa).

Of course, both paths are bound to cross. To give one last example, the Guinea-Bissau Kriyol causative forms described above result from the generalization and paradigmatization of a few Portuguese verb phrases such as a ferventar a água ‘to boil the water’ which indeed convey a causative meaning with respect to the related simple verb: cf. a água está a ferver ‘the water is boiling’, except there are too few such cases in Portuguese to assume a causative derivation. Guinea-Bissau Kriyol pairs such as firbi vs. firbinti, in contrast, instantiate a productive causative formation, as we saw. This would probably not have happened, however, were it not for the fact that the substrate-adstrate languages Mandinka and Balanta both have similar formations involving suffixes that accidentally sound very much like the Guinea-Bissau Kriyol suffix (Kihm 1994: 246ff.).

Who made the rapprochement and the generalization and when, we do not know. What we do know is that we do not want to assume they had any awareness of the suffixes as such. All the knowledge they needed was of the few Portuguese items and the Mandinka and/or Balanta paradigms, plus the cognitive ability to relate the two and draw the (subconscious) conclusion.

Once in the process of vernacularization, the incipient creole will at some point be transmitted to children who acquire it as an L1 (not necessarily their only L1). There is no evidence that this event constitutes a significant step in the development of the language. The children of course continue to elaborate it, probably accelerating the process thanks to their acute capacity for ‘creative analogy’, but they do so by using the same means as did their parents.

Abbreviations

1SG 1st person singular, 1SG.S 1st person singular subject, 2SG.POSS 2nd person singular possessive, 3SG 3rd person singular, 3SG.O 3rd person singular object, ANT anterior, ASP aspectuality, CAUS causative, FIN finite, FUT future, HAB habitual, LOC locative, MKR marker, PFV perfective, PROG progressive, VFORM verb form.
References


PAPIA, 24(1), ISSN 0103-9415, e-ISSN 2316-2767
Theories of morphology and theories of creole emergence...

Morfologia e emergência das línguas crioulas...


