23 Contact and Portuguese-lexified creoles
Hugo C. Cardoso

1 Introduction

When, in August 1415, King João I of Portugal led an armada across the mouth of the Mediterranean Sea to lay siege to the city of Ceuta, he was in fact launching the first wave of a process of European overseas expansion which would impact the world for centuries to come. Over the course of the next century and a half, the Portuguese would establish themselves around the globe, with outposts in South America, the Atlantic islands, along the coast of East and West Africa, and in Asia and the Pacific, all the way to Japan and Timor. This historical process of political and economic expansion was necessarily accompanied by linguistic diffusion, which accounts for the global nature of Portuguese in the modern world, but also for the development of the Portuguese-lexified creoles (henceforth PLCs), which occupy us here.¹

PLCs have long been prominent in scholarship about creole languages. There are several reasons for this, not least the fact that, because of the chronology just mentioned, PLCs include some of the oldest of the European-lexified creoles which developed out of the colonial expansion of the Early Modern Age. In addition, even though several members of this group have died out (especially in Asia) and some of the others remain among the most under-studied of all creoles, PLCs also include some of the most vital, vibrant, and conspicuous creole languages anywhere.

Besides, the extremely different social, historical, and political trajectories of individual PLCs mean that the group includes very diverse languages, which cover some of the most influential typologies of creole languages proposed in the literature. All three types in Bickerton’s (1988) sociohistorical classification² are represented, not just fort creoles (such as Diu Creole), but also plantation creoles (like Santome), and maroon creoles (like Angolar). Chaudenson’s (1979) opposition between exogenous and endogenous creoles, based on whether or not their formation involved the relocation of the contributing populations, is also covered, with some PLCs (such as Capeverdean Creole) belonging to the first type, and others (including Malacca Creole) to the second. In


² Itself a reinterpretation of Reinecke’s earlier (1937) classification which, in addition to trade jargons (corresponding to pidgins), distinguished plantation creole dialects from settlers’ creole dialects.
addition, because of their wide geographical spread, the formation of the various PLCs involved contact with languages of many distinct families (including Niger-Congo, Indo-European, Dravidian, Austronesian, and Sino-Tibetan), producing very different linguistic encounters and significant typological diversity among them. Such diversity in terms of their formative and developmental settings makes this group of languages particularly adequate to test out theories of creole formation. In that respect, it is noteworthy that, in the late nineteenth century, two of the most important early creolists, Adolfo Coelho and Hugo Schuchardt, relied heavily on data from PLCs to develop their contrasting theories of language contact.

There are, however, other dimensions to the diversity encapsulated by PLCs. One of them relates to the changing role of the lexifier language. In fact, while some PLCs (like those of Guinea-Bissau or Daman) coexist with Portuguese to this day, others (including those of Sri Lanka or Korlai) eventually lost significant contact with it after their formation, making it possible to explore the impact of this variable on their diachronic trajectory, not only in terms of the source and/or development of particular linguistic features, but also of the prevalence of sociolinguistic variation, speakers’ attitudes, official recognition, vitality, and so forth. Another important dimension to these languages’ diversity, which opens up new possibilities of linguistic enquiry, pertains to their socio-political status within their host societies. While some PLCs, such as the creoles of Cape Verde and Guinea-Bissau, have become stable languages with large speech populations and have made or are making significant progress towards standardisation and official recognition, others never progressed past a fragile minority status within their wider contexts and have either died out (as in the case of the creole of Mangalore) or are currently facing linguistic contraction and potential death (as in the cases of Principense or Macau Creole).

Evidently, such a diverse set of languages can be (and has been) approached by linguists from very different perspectives, and contributes many different insights with respect to contact-induced language change. As it would be impossible to survey all of these here, we will focus on select studies and debates which in some way derive from one of the factors which make them stand out among the creole languages that resulted from European overseas expansion: their time depth. The fact that PLCs include, to the best of our knowledge, the first of that group of languages to form not only raises the possibility that they may have influenced other subsequent contact languages, but also enhances the potential for post-formative diachronic dynamics to have left a mark. Before embarking on these discussions, however, it is important to describe in a little more detail the current and past distribution of the PLCs.

2 The Portuguese-lexified creoles

It has already been mentioned that the beginning of the Portuguese overseas expansion is symbolically placed in 1415, but this does not mean that all PLCs have their origin around that period. The progression of navigation and exploration made coastal West Africa the main locus of linguistic contact involving Portuguese throughout the fifteenth century – and, indeed, the formation of the West African PLCs has been located within that period –, while linguistic diffusion to Asia and the Americas is a sixteenth-century phenomenon. The geographical reach of the Portuguese between then and the end of the colonial era was substantial, stretching from Brazil to Japan. PLCs may have formed anywhere along that geographical span – and even beyond, as a result of the migration of previously influenced populations. In fact, those that either persist or have been recorded do stretch across the globe, even if, in some places – including Brazil and Japan, precisely, but also Goa or Angola.
the formation of PLCs is a matter of conjecture, as no conclusive records are to be found.3

At present, PLCs are spoken in Western Africa [Cape Verde, Guinea-Bissau, Senegal, São Tomé e Príncipe, Equatorial Guinea] and Asia [India, Sri Lanka, Malaysia, Singapore, Macau] (see Cardoso, Hagemeijer & Alexandre 2015) by populations of widely variable dimensions, which range from close to a million, in the cases of the creoles of Cape Verde and Guinea-Bissau, to less than 10 speakers, in the case of the one spoken in Kerala (South India). Researchers of these languages tend to group them in geographically-defined clusters which also reflect significant linguistic continuities. These clusters operate at different levels, as demonstrated in Table 1 below: while, since at least Ferraz (1987), a higher-order classificatory split is proposed between the Atlantic and the Asian creoles, lower-order regional clusters such as e.g. Upper Guinea or Southern South Asia can also be recognised.

Table 23.1. Proposed geographical groupings of the current PLCs (identified by several recurrent glossonyms); designations in italics correspond to the classification proposed by Schuchardt (1889a) for the Asian PLCs, based on the taxonomy of their substrate/adstrate languages.

<table>
<thead>
<tr>
<th>GEOGRAPHICAL CLUSTERS</th>
<th>CURRENTLY-SPOKEN LANGUAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic</td>
<td></td>
</tr>
<tr>
<td>Upper Guinea</td>
<td>Capeverdean Creole/Kabuverdianu [Cape Verde]</td>
</tr>
<tr>
<td></td>
<td>Guinea-Bissau Creole/Kriyol [Guinea-Bissau]</td>
</tr>
<tr>
<td></td>
<td>Casamance Creole/Kriyol [Senegal]</td>
</tr>
<tr>
<td>Gulf of Guinea</td>
<td>São Tomé Creole/Santome/Forro, Príncipe Creole/Principe Creole/Principe Creole/Lung'Ie, Angolar/Lunga Ngola [São Tomé e Príncipe]</td>
</tr>
<tr>
<td></td>
<td>Annobón Creole/Annobonese/Fa d’Ambu [Equatorial Guinea]</td>
</tr>
<tr>
<td>Asia</td>
<td></td>
</tr>
<tr>
<td>South Asia [Indo-Portuguese]</td>
<td>Diu Creole, Daman Creole, Korlai Creole [India]</td>
</tr>
<tr>
<td></td>
<td>Cannanore Creole [India]</td>
</tr>
<tr>
<td></td>
<td>Sri Lanka Portuguese [Sri Lanka]</td>
</tr>
<tr>
<td>Southeast + East Asia [Malayo-Portuguese]</td>
<td>Malacca Creole/Papia Kristang [Malaysia + Singapore]</td>
</tr>
<tr>
<td>East Asia [Sino-Portuguese]</td>
<td>Macau Creole/Makista/Patuá [Macau, China]</td>
</tr>
</tbody>
</table>

3 For an overview of the issue with respect to Brazil, see Lucchesi, Baxter & Ribeiro (2009), especially ch. 1. For the Goan case, see Dalgado (1921) and Cardoso (2016: 82-84).
The last column of Table 1 is entitled "Currently-spoken languages" because, as a matter of fact, the current distribution of the Portuguese-based creoles mentioned there represents only a portion of these languages' past implantation (see Smith 1995). It is important to bear in mind not only that several of them were abandoned over time (in some cases, potentially, without any records), but also that, according to some hypotheses which will be explored in this chapter, some transformed or were absorbed into other, rather different contact languages, leaving behind linguistic traces which are sometimes evident and largely uncontroversial, and sometimes not.

Not unlike creole languages in general, there are no substantial records for most of the Portuguese-based creoles before the late nineteenth century. The one of Sri Lanka (then called Ceylon) is a notable exception, since, in the first half of the century, several publications in or about this language were produced, including grammars, dictionaries, religious texts, and translations. Nonetheless, because several early creolists developed a scholarly interest in the PLCs around the turn of the nineteenth century, we actually have records of a few of the extinct varieties, such as that of Batavia (modern Jakarta) and nearby Tugu, in Indonesia (Schuchardt 1891), or that of Nagapattinam, in Southeastern India (Dalgado 1917). In this respect, Hugo Schuchardt stands out for the intensity of his

---

4 For a survey of sources about the PLCs in general, see Cardoso, Hagemeijer & Alexandre (2015). For Asia, see also Tomás (1992).
scholarly production and the encompassing nature of his descriptive studies (see Schuchardt 1882a/b, 1883a/b, 1888a/b/c, 1889b/c). Adolfo Coelho (1880-86) also cast a global look at the PLCs to support his universalist theory of creole formation, as did Leite de Vasconcelos (1901) in his global dialectological survey of Portuguese. Other philologists of the time specialised in particular regions, such as – to name but a few – Joaquim Botelho da Costa and Custódio José Duarte (1886), and António de Paula Brito (1887), who documented the creole of Cape Verde, or Sebastião Rodolfo Dalgado (1900, 1902-3, 1906), who focused on those of South Asia.

3 Quest for the origins

Creole languages posit a number of questions, some of which stem from a certain recurrence of particular features among them, regardless of their formative settings. One of the proposed solutions to account for perceived similarities between creoles (in general or within particular clusters) has been to posit a common origin, and the inaugural role of Portuguese in the European language expansion of the Modern Age has made it a logical component of several such theories. The strongest formulations of a monogenetic theory proposed that all pidgin and creole languages resulting from European overseas expansion ultimately derived from a Portuguese-lexified pidgin (henceforth PLP) formed in West Africa in the fifteenth century – perhaps itself a transformation of the preceding Mediterranean Lingua Franca, or Sabir –, then spread around the globe through the networks of trade (including the slave trade) and adapted to new lexical bases (English, French, etc.) through relexification (Whinnom 1956; Taylor 1961; Thompson 1961). The presence of certain Portuguese-derived lexemes in non-Portuguese-lexified creoles – some recurrent, such as *pikinini/pikni/pequin/*... *child* (from Ptg. *pequen(in)o *'small') or *savi/savvy* 'to know' (from Ptg. *sabe* 'know(s)'), others circumscribed to particular (groups of) languages, such as *palaver* 'dispute' (from Ptg. *palavra* 'word') in West Africa or *briga* 'to fight' (from Ptg. *briga(r)* '(to) fight') in Guyanese (see Baker & Huber 2001 for further examples) – were seen as remnants from this process.

It is no longer widely held that a single origin is necessary to account for similarities among pidgins and creoles. One important challenge to the *Lingua Franca* and *West African Portuguese Pidgin* hypotheses comes from non-European-lexified pidgins and creoles (see e.g. Buchstaller, Holmberg & Almoaily 2014), which could hardly have been part of the same process of diffusion and relexification, even though only some authors ever considered the scope of a strong form of monogenesis to extend beyond the European-lexified creoles. However, the precedence of Portuguese in many former colonial settings and various instances of population displacement are historically attested facts; therefore, more circumscribed processes of linguistic transfer and speciation involving this language continue to be identified and discussed in the literature. In this section, we will explore some such cases. In 3.1, we survey diachronic proposals applicable within sub-groups of PLCs, while in 3.2 we focus on those that imply the influence of a PLC or PLP on contact languages currently not classified (strictly) as Portuguese-lexified.

3.1 Diffusion and speciation

The earliest PLCs – those of West Africa – developed during the fifteenth century, a period in which written records largely fail to document linguistic variation explicitly. As such, the actual process of language contact and creole formation occurred in something of a historiographical mist, which motivates many questions and different theories.

---

6 For overviews of monogenetic theories, see Holm (1988:44-52), and den Besten, Muysken & Smith (1995).
However, there is one type of textual evidence from the period which is relatively plentiful, and whose relevance for the formation of PLCs has been the matter of debate. With the confluence in Portugal (especially Lisbon and the southern regions) of Africans from several parts of the continent, in the fifteenth century, a particular form of L2 Portuguese came to be associated with them. This sociolect, known then as *Língua de Preto* ‘Language of the Black’ or *fala Guiné* ‘Guinea speech’, was often portrayed (mostly for comical effect) in Portuguese plays and literary works (see Lipski 2005; Pereira 2006: 69-73). Though certainly prejudiced and stereotypical, this highly variable corpus of attestations contains sentences such as the following, taken from Gil Vicente’s play *O Clérigo da Beira* [The clergyman of Beira], from the 1520s (transcribed in Kihm & Rougé 2013: 267):

(1)  

A mi bae furtá em tanto / camisa que sa na muro.  
1SG go steal meanwhile shirt REL COP on wall  
‘Meanwhile, I’m going to steal the shirt that’s on the wall.’

Such passages are a testament to the social visibility of African communities in fifteenth- and sixteenth-century Portugal, but they also interest linguists to the extent that they appear to reflect certain forms and constructions of the modern West African PLCs. With respect to the sentence in (1), for instance, pronominal *a mi* is reminiscent of Portuguese *a mim* ‘to me’ but, while in Portuguese this preposition-pronoun combination can only occur in object position, here it appears as a subject; more significantly, however, it seems to reflect the same morphological and syntactic reinterpretation which produced the first person pronouns of the current West African PLCs (see Lipski 1991), either *ami* or *(a)mi* (also, for other persons and numbers, *abó/(a)bo* ‘2SG’; *anós/(a)nos* ‘1PL’; etc.; see Kihm & Rougé 2013: 260). Likewise, the copula *sa*, probably a derivation of Ptg. *estar* ‘to be’ or *sam/são* ‘[they] are’, is also identified in some modern West African PLCs, namely among those of the Gulf of Guinea (Lipski 2005: 53-54; Kihm & Rougé 2013: 267-8). For some authors (e.g. Naro 1978; Kihm & Rougé 2013) – although this is not without controversy – the contact varieties developed among Africans in Portugal and represented as *Língua de Preto* were then carried to West Africa on board the Portuguese ships and formed the seedlings of the local PLCs.

Regardless of whether or not all West African PLCs had a common ancestry in the form of *Língua de Preto*, the split between Upper Guinea and Gulf of Guinea represented in Table 1 is linguistically as well as socio-historically motivated. While the languages which make up both clusters do share a number of linguistic traits – e.g. the presence of distinct strong and weak pronouns or negative concord –, there are also many features which separate both regions but are recurrent within each – e.g. discontinuous negation in the Gulf of Guinea Creoles (Hagemeijer & Alexandre 2012) or a postverbal anterior morpheme (-)ba (though with somewhat different morphological and syntactic constraints) in the Upper Guinea Creoles (see Holm 1988: 152). This has been interpreted as evidence of two distinct formative histories, leading to attempts at reconstructing the diachrony of each regional cluster.

In the case of Upper Guinea, the internal consistency is relatively uncontroversial (see e.g. d’Andrade & Kihm 2000; Baptista, Mello & Suzuki 2007). The bone of contention has been that of the ultimate origin of the proto-language in the ancestry of these creoles – if indeed they do share a common ancestry. One theory posits that the original contact language was established on the continent, around the Portuguese outposts of Guinea, and was then carried over to Cape Verde; conversely, another theory posits an origin on the Capeverdean island of Santiago and later diffusion both to the other islands of the archipelago and the West African coast.7

---
7 See Jacobs (2010) for an overview.
The directionality of spread is somewhat less obscure in the Gulf of Guinea. On the islands of the Gulf of Guinea, two PLCs are spoken in São Tomé (Santome and Angolar), one in Príncipe (Lung'Ie), and one in Annobón (Fa d’Ambu), and, while their mutual intelligibility is restricted, their relatedness has been determined on linguistic grounds since at least the work of Ferraz (1979). More recently, comparative studies of the four languages (Hagemeijer 2011; 2015) have not only strengthened the case for their genetic connection, but also suggested that the proto-language received a substrate contribution constituted mostly of Niger-Delta languages, with Bantu influence kicking in later on (a fact supported also by a predominance of the Nigerian genotype among the population of São Tomé and Príncipe, see Hagemeijer & Rocha 2017). In the case of the Gulf of Guinea creoles, linguistic, sociodemographic, genetic and historical evidence converge to define the scenario of a proto-creole forming on São Tomé towards the end of the fifteenth century, and then diffusing to the islands of Príncipe and Annobón in the sixteenth century. Angolar, on the other hand, appears to have broken away from the proto-creole due to the constitution of a community of runaway slaves in Southeastern São Tomé.

The PLCs of Asia present new challenges when it comes to reconstructing potential patterns of intra-continental diffusion, but even their relationship with the creoles of the Atlantic is an important diachronic question. Since, historically, Portuguese overseas exploration proceeded southwards and eastwards around the Cape of Good Hope (as did the Asian-bound vessels that sailed the Cape route), one could expect that whatever phenomena of language contact occurred on West Africa would be carried over to East Africa and Asia. However, an influential comparative study by Ferraz (1987) argued that the Asian PLCs constitute a separate cluster (cf. Table 1) with a number of linguistic commonalities that set them apart from the West African PLCs, suggesting a fundamental genetic break between the two spheres (but see Clements 2000 and Holm 2009 for less radical views). Some of those “Luso-Asianisms” include the ones given in Table 23.2. It is important to notice that these features are not necessarily universal among (modern) Asian PLCs, but they do cut across geographical clusters, which may be indicative of a common origin.8 Clements (2000) admits some carry-over from West Africa but does posit an “Asian pidgin” which could be responsible for diffusing the linguistic features that crop up in various Asian PLCs. He further specifies that this Asian pidgin is likely to have been shaped in South Asia, and, more specifically, around the earliest Portuguese settlements in Asia (chiefly, Cochin), on the Malabar Coast of Southwestern India (see also Clements 2009). The hypothesis that the initial language contact that occurred on the Malabar Coast (primarily involving Portuguese and Malayalam) determined the common features of Asian PLCs has not been fully worked out, but Baxter & Bastos (2013), in a study of the prototypical Asian-Portuguese possessive construction (see Table 2, nr. 1), find that the conditions of the initial encounter are consistent with the development of this particular structure.

Table 23.2. Some widespread features among Asian PLCs identified by Ferraz (1987)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a possessive construction with the structure: Possessor + su/sū/s(u)a + Possessed</td>
</tr>
<tr>
<td>2</td>
<td>syncrletic dative-accusative case-marking, expressed by an adposition derived either from Portuguese para ‘for’ or por ‘by’, or from Portuguese com ‘with’</td>
</tr>
<tr>
<td>3</td>
<td>preverbal Tense-Aspect markers derived from Portuguese já ‘already’ (Past/Perfective), está/estar ‘is/to be’ (Non-punctual aspect) and logo ‘immediately, later’ (Future/Irrealis);</td>
</tr>
<tr>
<td>4</td>
<td>syncrletic existential/possessive/copular verbs</td>
</tr>
<tr>
<td>5</td>
<td>a special future negator derived from Portuguese não há-de ‘shall not’</td>
</tr>
<tr>
<td>6</td>
<td>pluralising reduplication</td>
</tr>
</tbody>
</table>

8 For an overview of debates surrounding the (dis)unity of the Asian PLCs, see Ansaldo & Cardoso (2009).
certain lexical items, such as *ada/ade/adi* ‘duck’

Multimorphemic words involving a derivation of Ptg. *laia* ‘fashion’ (e.g. question word *kilai* ‘how’)

Several other instances of linguistic diffusion may be identified within the Asian PLCs, but two are especially interesting and will be described here: the Malacca-Macau axis, and the establishment of a PLC on the island of Java.

The first of these helps to explain the conflation of “Southeast + East Asia” as a second-order cluster in Table 1, and concerns the ultimately Malaccan origin of the creole of Macau. At least since the work of Batalha (1965-66), it has been a truism that the creoles of Malacca and Macau share a particular genetic bond (see also Holm 1989: 296-298; Baxter 2009a) regardless of the geographical distance between them and their different linguistic environments (Malay-dominated in the case of Malacca, Cantonese in Macau). Since then, comparative studies have provided further evidence that certain features of Macau Creole are best explained by transfer from Malay than from Cantonese, including its causative-facilitative serial verb construction (Baxter 2009b), and the structure of comparative constructions (Cardoso 2012). The following examples illustrate the reasoning in the case of the comparative constructions:

(2)\(^9\)

<table>
<thead>
<tr>
<th>Example</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. <em>Eles mais pior que japonês.</em></td>
<td>Macau Creole</td>
</tr>
<tr>
<td>3PL more bad CMPR Japanese</td>
<td>‘They [were] worse than the Japanese.’</td>
</tr>
<tr>
<td>b. <em>E ela mais velha de que eu.</em></td>
<td>and 3SG.F more old CMPR 1SG</td>
</tr>
<tr>
<td>‘And she [was] older than me.’</td>
<td></td>
</tr>
<tr>
<td>c. <em>Iloutro ung pouco mais alto grau de nós.</em></td>
<td>3PL one little more high degree CMPR 1PL</td>
</tr>
<tr>
<td>‘They [were] a little higher-class than us.’</td>
<td></td>
</tr>
<tr>
<td>d. <em>elí más altu di Pio</em></td>
<td>Malacca Creole</td>
</tr>
<tr>
<td>3SG more tall CMPR Pio</td>
<td>‘She is taller than Pio.’</td>
</tr>
<tr>
<td>e. <em>Dia lebih tinggi dari saya.</em></td>
<td>Malay (Bahasa Indonesia))</td>
</tr>
<tr>
<td>3SG more tall CMPR 1SG</td>
<td>‘He is taller than me’</td>
</tr>
<tr>
<td>f. <em>Gāmyaht yiht gwo kàhmyaht.</em></td>
<td>Cantonese</td>
</tr>
<tr>
<td>today hot CMPR yesterday</td>
<td>‘Today is hotter than yesterday.’</td>
</tr>
</tbody>
</table>

Of the three slightly different comparative constructions in Macau Creole, which essentially differ in the form of the comparative marker, (2a/b) are very close to Portuguese (which uses the comparative marker *que* or *do que*), a language that has always played (and continues to play) an important role among the Macanese community; (2c), on the other hand, is practically equivalent to the construction in Malacca Creole (2d). Both of these have a very close model in the Malay construction in (2e), with respect to the order of the terms but also the presence of an index (*lebih* ‘more’, equivalent to *más/mais*) and the semantics of the comparative marker, which, in all these cases (*dari* in Malay, *di* in Malacca Creole, *de* in Macau Creole) also marks ablative. By contrast, even though we might expect Cantonese (or Hokkien, prominent in the earliest stages of colonial Macau, see Ansaldo & Matthews 2004) to have contributed to the structure of comparative constructions in Macau Creole, its comparative construction (2f) contains no index, and the comparative marker *gwo* is not an ablative marker, but rather a verbal

---

\(^9\) All sentences in example (2) given in Cardoso (2012: 101; 102; 105; 107).
element meaning ‘to cross/pass’. It is linguistic correspondences such as this, coupled with sociohistorical evidence of early population movements from Malacca to Macau (Batalha 1974; Baxter 2009a), that motivate the interpretation of Macau Creole as an offshoot of Malacca Creole, later transformed by speciation and the unavoidable linguistic influence of Sinitic languages (see e.g. Ansaldo & Matthews 2004, for an exploration of the Sinitic source of certain aspects of Macau Creole reduplication).

The other instance of linguistic transfusion involving Asian PLCs to be discussed here concerns the implantation of one such creole in the city of Batavia (modern-day Jakarta) and nearby Tugu, on the island of Java. Considering that this particular territory never had an official Portuguese presence, the use of a PLC until at least the late nineteenth century may come as a surprise. Batavia and Tugu Creole, now extinct, is known to us through a few written sources – most importantly, Schuchardt (1891) – which have been collated and analysed in Maurer (2011). As it turns out, the establishment of this language on Java was the product of multiple migrations converging onto Batavia – earlier the centre of the Verenigde Oost-Indische Compagnie (Dutch East Indies Company) – from across Asia, especially locations which the Dutch had conquered from the Portuguese around the mid-seventeenth century. The phenomenon is clearly put in a 1708 memorial by two protestant priests in charge of the “Portuguese Church” of Batavia:

“The Portuguese language is used daily and privately by the slaves of the families who come from Ceylon and the [Coromandel; HC] Coast; by all the owners of the slaves and their children in their daily interaction with the slaves and the indigenous Christians; by the families and people who come from Siam, Malacca, Bengal, the Coromandel Coast, the island of Ceylon, the Malabar coast, Surat and even Persia; and even the heathens who live in this city and trade with the Christians or their slaves learn to speak Portuguese.” (quoted in Lopes 1936: 67-68; transl. HC)

The multiple provenance of these PLC-speaking populations appears to have created, among this community, a setting easily conducive to dialect levelling. It is true that, as noted by Baxter (1996: 310), Batavia and Tugu Creole data look remarkably like Malacca Creole, which is unsurprising considering the geographical proximity and similarities in terms of the linguistic ecologies of colonial Malacca and Batavia. However, when we dig deeper, we also find certain forms, lexical as well as functional, which establish links with South Asian PLCs (see Cardoso 2019), although this connection has not been systematically explored.

3.2 A foundational role?

One of the most interesting effects of the inaugural role of Portuguese in the European global linguistic expansion of the Modern Age is that, for various reasons and to different degrees, Portuguese often came to impact the linguistic repertoire of the colonisers who followed and, on many occasions, occupied places where this language had already created roots. Even discarding strong formulations of a monogenetic nature, several theories have been proposed that involve Portuguese or a Portuguese-lexified contact variety in the formation or development of a particular language which does not strictly fall under the PLC category. Here we will mention only a few of them.¹¹

¹⁰ See also Maurer (2013) for linguistic samples and grammatical analysis. One further source, not included in this lot, consists of a manuscript recently identified in the library of Lisbon’s Museu Nacional de Arqueologia (see Castro et al. 2019).

¹¹ Further cases which will not be treated here involve Palenquero, for which a connection with the Gulf of Guinea PLCs has been proposed (cf. Schwegler 1999); Hawaiian Pidgin, on which the potential impact of Portuguese is explained by the presence of migrants from Portugal, from the
The various English-lexified creoles of Suriname are a case in point. While there is a discernible Portuguese lexical element in all of them, including Sranan (Smith 1987) and Ndyuka (Huttar 1989), the Western Maroon Creoles, which consist of Saramaccan and Matawai, stand out in this respect. Consider the following excerpt of a story in Saramaccan (transcribed in Bakker, Smith & Veenstra 1994: 165-166); the elements which can be ascribed a Portuguese etymon have been underlined:

(3)  
\begin{verbatim}
Hɛn totómbotí táa we a o-du lúku tu
Then woodpecker said well 3SG IRR-do look too

'Gaamá, mi a-gó nákí lúku.' Hen déé otowán táki táa:
"Chief, I am going to try to hit it." Then the others said:

Kú ún-búka, i lánga bákahédi ku di gaán taku fi-i de?
with which-beak, 2SG long back.head with DEF.SG big ugliness of-2SG there
"With what beak, you long back-of-the-head, with your great ugliness?"
\end{verbatim}

A study of the Saramaccan lexicon by Smith & Cardoso (2004) which pitched Portuguese-derived lexemes against English-derived lexemes (i.e., ignoring the other, minority lexical sources) found a nearly equal distribution in the domain of nouns (176 from Ptg., 179 from Eng.), a smaller but still highly significant proportion of Portuguese-derived adjectives (27 from Ptg., 51 from Eng.), and a majority of Portuguese-derived verbs (154 from Ptg., 121 from Eng.; see also Perl 2000). Such an even division of labour makes it possible to consider Saramaccan a mixed creole, although the distribution of the Portuguese- and English-derived lexicon is not entirely equal: according to Smith (1987), while the Portuguese element is very solid among the basic vocabulary of Saramaccan, accounting for nearly 35% (but just under 4% in Sranan), it is much less well-represented in the repertoire of function words, at just 16% (and 1.5% in Sranan). Smith’s interpretation is that this typifies a scenario in which, in the case of Saramaccan, an originally English-lexified creole integrated the Portuguese element at a later stage, but the truth is that several different explanations have been proposed for the remarkable lexical impact of Portuguese in Suriname. Some of these run contrary to Smith’s scenario. Voorhoeve (1973), for instance, argues for a Portuguese-lexified origin for Saramaccan (and Sranan), and later relexification towards English.  

There has been some disagreement on the origin of the Portuguese (or Iberian, in some theories) lexical element in Suriname. Several possible routes have been proposed, including the diffusion of the Mediterranean Lingua Franca or a West African pidgin/creole, the impact of previously-influenced African languages, or of a French-lexified baraguoin circulating in the Caribbean from the sixteenth century onwards which had, however, absorbed several Ibero-Romance words (see Arends 2017: 39-43). However, according to some authors, the most important historical development in this respect was the influx of (Judeo-)Portuguese-speaking Sephardic Jews from the short-lived Dutch colony in Northeastern Brazil (1630-1654) in the mid-seventeenth century. When the Portuguese retook Pernambuco, the resident Sephardic Jews moved to Suriname (either directly, or by way of Guyana or Europe) and set up several plantations on the Upper Suriname River. Some authors believe that, in the process, they also brought over significant numbers of slaves from their earlier Brazilian plantations (e.g. Goodman 1987), but this is controversial.12 Be it as it may, by the eighteenth-century, several historical
sources mention a language known as Djutongo, which appears to have been a PLC used in these plantations and is thought to be the precursor of the West Maroon Creoles (see Arends 2017: 182-188).

This particular mid-seventeenth-century migration of Portuguese-speaking Jews from Brazil has also been seen as relevant to understand parallel linguistic developments in Aruba, Bonaire and Curacao. The creole spoken on these islands, Papiamentu, was traditionally classified as a Spanish-lexified creole, but, here too, the Portuguese lexical element is evident. Whereas several lexemes can only be derived from Spanish, many may be either from Spanish or Portuguese, while others still can only be derived from Portuguese (e.g. ainida ‘still’, bai ‘to go’, bon ‘good’, forsa ‘strength’, nasemenu ‘birth’, tresè ‘to carry, wear’; see Grant 2008; Lipski 2008). First of all, as with Saramaccan, this has motivated the question of whether Papiamentu originated as a Portuguese- or Spanish-lexified language, which has proven a bone of contention among scholars and speakers alike. In addition, the origin of such a significant Portuguese element has also attracted considerable debate. A number of explanatory theories have been proposed, including: a) the transfer of a West African PLC or PLP; b) the influence of the Judeo-Portuguese spoken by the Sephardic Jews relocating from Northeastern Brazil or from Europe; and c) even the influence of an Asian-Portuguese creole (see Grant 2008). Recent comparative work has favoured an Upper Guinea connection, on the basis that some essential characteristics of Papiamentu have very concrete correlates among the PLCs of Cape Verde and Guinea-Bissau. Jacobs (2009) explores a few of these, including the fact that both Papiamentu and the Upper Guinea PLCs make a distinction between simple and emphatic forms within their pronominal system – related to the pronominal forms (a)mi, (a)ba, (a)nas discussed above in connection with example (1) – but also correspondences concerning the distribution of mono- or bimorphemic question words, the form and semantics of prepositions, conjunctions, and so forth. The debate is still ongoing, but the recognition of the pervasiveness of the Portuguese element has now made it common to classify Papiamentu as an “Ibero-Romance-lexified” or “Spanish/Portuguese-lexified” creole.

Over in Asia, similar proposals have also been advanced. One of the earliest – and perhaps the most contentious one – has to do with a putative Portuguese origin for Chabacano, the cluster of Spanish-lexified creoles of the Philippines. The recognition of certain similarities between Chabacano and the Asian PLCs (especially those of Southeast Asia and Macau, see e.g. Batalha 1960) motivated the hypothesis – formulated particularly strongly by Whinnom (1956; 1965) –, that the origin of Chabacano lay in the resettlement of a PLC-speaking community from the Moluccan island of Ternate to the Philippines, around 1660, followed by relexification towards Spanish. The applicability of this scenario to the history of all varieties of Chabacano has been questioned, although the possibility remains that it may be relevant for the formation of at least one: Ternateño (see Fernández 2012: 226).

Less controversial than the proposed Portuguese past of Chabacano is the role that Portuguese – or rather, a PLC or PLP – played in the development of Chinese Pidgin English. For a long time, the Portuguese in Macau had a privileged role in the Canton trade, and, as such, a contact variety of Portuguese definitely circulated in the Pearl River Delta. It was only in the eighteenth century that the British became especially prominent in the region’s trade (van Dyke 2005). This transition in terms of economic influence led to the gradual replacement of a Portuguese-based contact language with an English-based one, which came to be known as Chinese Pidgin English. Interestingly, in a study of records of Chinese Pidgin English, Hall (1944) notices that the earliest ones were characterised by a large number of Portuguese-derived words (such as e.g. joss for ‘God’, a term preserved in the widespread English term joss-stick), which, as time went by, were progressively replaced with English-derived terms. In fact, one of the earliest known attestations of

---

13 See Lipski (2008) and Jacobs (2009) for an overview.
Chinese Pidgin English (Noble 1762: 240), collected in the 1740s during a trip to Canton, actually consists of a sentence that is mostly composed of Portuguese-derived terms:

(5)  Carei grandi hola, pickenini hola?
    want old whore young whore
    ‘Do you want an old whore or a young whore?’

In this sentence, only hola has an English source; all the other terms are Portuguese-derived: carei from Ptg. querer ‘to want’, grandi from Ptg. grande ‘big’, and pickenini from Ptg. pequenino ‘small’. Examples like this, not atypical in the 1700s, call into question the classification of the language as an English-based pidgin. One hypothesis which has received some attention is that earlier contact varieties of Portuguese did not simply provide loanwords to Chinese Pidgin English, but were in fact its ancestors. According to this possibility, Chinese Pidgin English would have been formed, to some extent, through a process of relexification of a previous PLC or PLP. Matthews & Li (2012), for instance, interpret certain functional features of Chinese Pidgin English (such as the use of for as a complementiser or have as a locative copula) as retentions from a Portuguese-based ancestor, since similar constructions are attested elsewhere among the Asian PLCs.\footnote{\textsuperscript{14}}

\section*{4 Caught in the middle}

One consequence of the PLCs’ relative antiquity is that the opportunity presents itself, more than perhaps in the case of younger languages, for socio-political and linguistic conditions to change over time and for various types of language change to operate. In fact, as already mentioned, among the PLCs there is a diversity of situations: some have had constant and substantial contact with the lexifier but not with the major substrate(s) (e.g. Cape Verdean); some have remained in touch with their substrate language(s) but not the lexifier (e.g. Sri Lanka Creole); yet others have maintained contact with both (e.g. Guinea-Bissau Creole); and others lost touch with both at some point (e.g. Fa d’Ambu). In addition, significant changes in these languages’ linguistic and socio-political contexts (including colonial takeovers and migrations) occurred at different times in history, which makes it possible to observe the long-term consequences of variables such as the presence or absence of the lexifier, substrate languages, or adstrate languages, but also of the position of the creole and its speakers in their wider communities.

Wherever PLCs have remained in close contact with Portuguese, a number of different trajectories can be identified. In places such as Cape Verde and Guinea-Bissau, the creoles have gained wide currency but, for most of their history, have been subjugated to Portuguese in the formal and official domains. In fact, Portuguese remains the official language of these two countries, although the creoles are increasing their public visibility and taking steps towards an enhanced political status. Nonetheless, this particular coexistence has led to the development of considerable linguistic stratification, which, in the case of Guinea-Bissau, has been interpreted as a creole continuum brought about by decreolisation (Couto 2009). In other places, however, Portuguese has exerted a more pernicious pressure. In São Tomé and Príncipe, for instance, the creoles have been losing ground very quickly, to the point that their demise is now a possibility (Hagemeijer 2016). In Diu, it has been observed that, decades after Portuguese lost any official status in the territory, it still exerts some normative pressure on the creole, posing certain communicative constraints and defining language attitudes among the speech community.

\footnote{\textsuperscript{14} Notice however that, as mentioned in section 3, pickenini is part of a set of Portuguese-derived words which occur in many non-Portuguese-lexified pidgins and creoles; therefore, the use of this term in (5) could be the result of diffusion through (a variety of) English.}

\footnote{\textsuperscript{15} For a summary of the issue, see also Li & Matthews (2016).}
(Cardoso 2007; 2016). As for Macau, the fact that the creole is now all but gone has also been attributed to a gradual process of shift – or rather, according to Baxter (2009), assimilation of the creole into a Macanese norm of Portuguese –, motivated by the institutionalisation of Portuguese during the late colonial period.

In some former Portuguese outposts that went through drastic political changes, normally accompanied by the demotion or removal of Portuguese, the local PLCs have disappeared. Among those are the creoles of Cochin, Mangalore, Nagapattinam or Bengal (India), of Burma (i.e., Myanmar), and of Flores (Indonesia), to name but a few. In other places, however, these languages have stood their ground, but the removal of the lexifier from the ecology had profound effects on their linguistic development. The Asian PLCs provide a very interesting laboratory to observe the effects of such an event, since those which have lost contact with Portuguese did so at different points in time, and also because, as endogenous creoles (for the most part), they did remain in close contact with their former substrates. This situation has resulted in considerable convergence towards the adstrate(s) – or even, according to certain interpretations, metatypy. The creole of Sri Lanka is one of the most iconic examples of this process, as demonstrated by Smith (1979). The island was taken over by the Dutch in the mid-seventeenth century, after one and half centuries of Portuguese presence, and, while this did not mean the complete eradication of Portuguese from the territory, it did bring about a very radical break. For the ensuing centuries, then, the PLC of Sri Lanka developed with little to no contact with the lexifier but in close proximity with the substrate/adstrate languages, Tamil and Sinhala, which the speakers themselves mastered. As a result, currently, the language – which is preserved mostly in Tamil-majority areas – shows evident signs of grammatical convergence towards Tamil. The following is an example given by Smith (1979: 201), in both the creole (Sri Lanka Portuguese, SLP) and Batticaloa Tamil (BT), to illustrate the degree of isomorphism that obtains between the two:

(6) SLP ew eli-pa  diñe:ru ja:- dá:.  
    BT na:n avan-ukku calli -ya kúTu -tt -an.  

1SG 3SG.M-DAT money -ACC PST- give -PST -CONC  
‘I gave him the money.’

Despite some differences, there are also important similarities, not only in terms of word order but also of case-marking, which become even more significant once we compare the sentences in (6) with the equivalent in Portuguese (7):

(7) Eu dei-lhe o dinheiro.  

1SG give.1SG.PST-3SG.OBL DEF.M money  
‘I gave him the money.’

Sri Lanka Portuguese is by no means the only Asian PLC to exhibit the structural influence of its substrate(s) or adstrate(s) (settings such as those that produce endogenous creoles call into question the very distinction between the two concepts), but the degree of convergence is perhaps only matched, among this group of languages, by the creole of the Malabar, which is understandable since the uprooting of Portuguese occurred in much the same period and circumstances as in Sri Lanka. In fact, recent comparative studies of the Asian PLCs have drawn attention to the long-term impact of the removal of the lexifier from a creole’s ecology. Looking at word order features and comparative constructions, respectively, both Smith (2012) and Cardoso (2012) attempted to rank the various Asian PLCs (and also Chabacano, in the case of Smith’s study) according to their similarity with either the lexifier or the relevant substrate/adstrate. The results are given in Table 3:

Table 23.3. Ranking of Asian creoles according to their similarities with the lexifier or the substrate/adstrate; adapted from Cardoso (2012) and Smith (2012)
There are certain regularities in the two rankings which correlate neatly with the chronology of the break (if any) with the lexifier. At the end of the scale that indicates a closer similarity to the lexifier, we consistently find the Asian PLCs which remain in contact with Portuguese or only began to lose some connection in the twentieth century (Macau, Daman, and Diu); at the other end, signalling a strong resemblance with the relevant substrate(s)/adstrate(s), are the ones for which contact with Portuguese was substantially severed in the mid-seventeenth century (Sri Lanka, Malacca, Batavia/Tugu; also the Malabar, in the case of Cardoso's study); the creole of Korlai, spoken in a region from which Portuguese was dislodged in the eighteenth century, hovers in the middle. This exercise provides a clear visualisation of what Smith (2012: 141) calls the 'tug-of-war' the lexifier and the substrate(s)/adstrate(s) enter into as potential developmental models for the creole and sheds new light on the sociolinguistic condition of many creoles as languages caught in the middle of competing diachronic pulls.

4 Conclusion

Earlier in this chapter, we made the point that PLCs were especially important for the early creolists, who developed their approaches to pidgins and creoles in the late nineteenth century and the beginning of the twentieth century. What is interesting is that, despite the fact that creole studies did develop a discernible focus on English- and French-lexified creoles (and especially those of the West Atlantic) throughout the twentieth century, PLCs have more than stood their ground within the field. In this chapter, we have discussed a number of prominent issues in scholarship about PLCs which help explain their relevance for creole studies – and for language contact as a whole –, and also illustrate the diversity of debates that require a reference to them. Due to space constraints, the range of possible issues to cover was restricted here to those that in some way relate to the circumstance that the PLCs are among the oldest European-lexified creoles – which, in turn, derives from the fact that the Portuguese were the initiators of European overseas expansion in the Modern Age.

Within this narrow domain, issues of historical diffusion and speciation become paramount, and, as we have seen, much research has gone into disentangling genetic connections between particular subgroups of PLCs but also, perhaps more unexpectedly, between PLCs/PLPs and a number of English-lexified and Spanish-lexified contact languages (an exercise that could be extended to include Dutch-lexified and French-lexified ones as well). In addition, issues of diachronic change are also amplified by the relative ancientness of the PLCs among the creoles that resulted from European
expansion. In this respect, we have highlighted the contribution of these languages to an understanding of how post-formative stages of creoles can be impacted by the competing influences of the lexifier and of the adstrates, and to what extent changes in the status of one or the others can determine their evolution.

However, evidently, these do not exhaust the contributions that PLCs have made and stand to make to linguistics. Untouched in this chapter are, for instance, the many ways in which synchronic studies of PLCs have benefited typology and formal linguistics, or how PLCs have also become prominent in sociolinguistics (e.g. language death and preservation, language policy, variation, and diaspora), language acquisition, or the theory and practice of language documentation.

The field of PLC studies is, in fact, quite dynamic. To illustrate the history of creole studies as a discipline, Sousa, Mücke & Krämer (2019: Fig. 1) charted the number of studies focusing on PLCs throughout the years, revealing a significant amount in the 1880s, and then a surge after the 1950s, with a peak in the 1980s. However, their survey ends in 1992. As a matter of fact, in the period since then, some important developments have taken place. One of the most significant of those, I would argue, concerns the documentation and description of some PLCs which were insufficiently known (or entirely unknown) until then – including, among others, those of Annobon, Diu, Daman, Korlai, and the Malabar –, but also of particular varieties of better-known creoles – such as some of the regional variants of Capeverdean Creole. The enhanced range and availability of PLC data, aided by the constitution of freely-accessible corpora, open up fresh analytical possibilities. In addition, the inclusion of many PLCs or varieties of PLCs in comparative databases such as Comparative Creole Syntax (Holm & Patrick 2007) or The Atlas of Pidgin and Creole Language Structures (Michaelis et al. 2013)16 is also highly relevant. If, to the growing availability of PLC data, we add the consolidation of a dedicated scientific society (the Association of Portuguese- and Spanish-Lexified Creoles) and the establishment of focused publications (such as the Journal of Ibero-Romance Creoles), these languages look set to retain their scholarly relevance.

ABBREVIATIONS

1 = first person (pronoun), 2 = second person (pronoun), 3 = third person (pronoun), CMPR = comparative marker, CONC = concord, COP = copula, DAT = dative (case), DEF = definite (determiner), F = feminine (gender), IRR = irrealis, M = masculine (gender), OBL = oblique (case), PL = plural, PLC = Portuguese-lexified creole, PLP = Portuguese-lexified pidgin, PST = past (tense), REL = relativiser, SG = singular

REFERENCES


16 In Holm & Patrick (2007), 4 (or 5, if you include Papiamentu) of the 18 featured languages are PLCs: Angolar and the Creoles of Cape Verde, Guinea-Bissau, and Korlai. In Michaelis et al. (2013), out of 76 languages represented, 14 (or 15, once again depending on the inclusion of Papiamentu) are PLCs: 3 varieties of Capeverdean Creole (Santiago, Brava, and São Vicente), Angolar, and the Creoles of Guinea-Bissau, Casamance, Sáo Tomé, Principe, Annobón, Diu, Korlai, Sri Lanka, and Malacca, and even the extinct creole of Batavia.


Batalha, Graciete N. 1974. *Língua de Macau: O que foi e o que é* [The language of Macau: What it was and what it is]. Macau: Imprensa Nacional de Macau.


Cardoso, Hugo C. 2016. O português em contacto na Ásia e no Pacífico [Portuguese in contact in Asia and the Pacific]. In Ana Maria Martins & Ernestina Carreira (eds),


Couto, Hildo Honório do 1996. *Introdução ao Estudo das Línguas Crioulas e Pidgins* [Introduction to the study of creole and pidgin languages]. Brasília: Editora UNB.


