The Meaning of ‘European’: The Challenge of High-Contact Varieties for Linguistic Taxonomy *

Hugo Cardoso

This article addresses the multiplicity of criteria involved in linguistic labeling, in particular with regard to the establishment of genetic taxonomies, and points out the largely extralinguistic considerations often involved in the resulting classifications and terminology. The matter of genetic classification is particularly complex when dealing with high-contact varieties, as their typological traits are likely to unveil the influence of a number of (often unrelated) ancestral languages. An analysis of the Portuguese-lexified creoles of Asia, in particular the Diu variety of Indo-Portuguese, not only makes it clear that applying the ‘European’ label to them is only weakly supported by typological evidence but can have detrimental consequences with respect to the languages’ social embedding in modern Asian societies as well as their maintenance. All these factors considered, it is suggested that linguists apply taxonomical labels only sparsely and clearly motivate their use, demonstrating sensitivity to the social echoes and possible implications of their terminology.

1. Introduction

From the very early days of scientific linguistic research, efforts to 1) define the boundaries between languages and 2) classify languages in relation to each other have been high on the agenda of linguists. Sir William Jones’ 1786 proposal of the Indo-European linguistic family is

---

* The research leading to this article was funded by Grant SFRH / BD / 18509 / 2004 of the Fundação para a Ciência e a Tecnologia, Portugal. I am indebted to Umberto Ansaldo, Salikoko Mufwene and a reviewer for important suggestions and comments on earlier drafts, as well as their encouragement. Any possible shortcomings, however, are entirely my responsibility.
a major example of this line of research, but some of modern typology still aims at the production or refinement of linguistic classifications. Issues of family relations between languages are in fact one of the most fascinating linguistic issues for the public in general, although it has been noticed time and time again that such issues as the definition of language/dialect boundaries, or the proposal/denial of the genetic relationship between two languages are not devoid of political significance and, as such, are rather sensitive issues.

The notion of 'genetic taxonomy', or 'genetic classification', is defined in this study as the practice of classifying the world’s languages in phylogenetic branches, i.e., with regard to their putative ancestors and the maintenance of ancestral traits. It is necessary to define the terminology this clearly because, as shown below, modern typological research often takes a different approach – classifying languages under a specific 'type' does not necessarily make any phylogenetic predictions, although typological similarity can be used to motivate taxonomic categories (see below). To illustrate the original endeavour of typology, one can recall the 19th-century work of Schlegel and von Humboldt, who proposed to divide the languages of the world according to their basic morphological structures into the categories of 'isolating', 'agglutinative' and 'fusional' languages (with a fourth category, that of 'polysynthetic' languages, sometimes added to the descriptive arsenal). The distinctions in modern typology are much more fine-grained and comprehensive than these early proposals; one might even say that the number and size of typological categories proposed depends on the level of analysis a particular author wants to engage in, which is not independent of theoretical inclinations. Similarly, it has been noticed that, when faced with a wealth of linguistic data and the challenge of organising languages/varieties in genetic families, some linguists tend to favour the atomisation of nodes in their trees ('splitters') while others prefer to define their entries in wider terms ('lumpers'). On a related note, the concept of 'phylum' is now in general use in linguistic typology to denote broad phylogenetic categories. In their work, Heine
and Nurse use the term ’to refer to a language grouping larger, less well defined, and less widely accepted than a ’family’, and which typically contains several families’ (2000: 4). When classifying the languages of Africa at the level of the phylum, the authors identify four categories (viz. Afroasiatic, Nilo-Saharan, Niger-Congo and Khoisan) but, going into more detail, subdivide each of these phyla into several families.

It becomes evident that, even from a purely typological point of view, there is some malleability in the classification of languages. I will suggest that while genetic taxonomy draws on linguistic data, in actuality it often surpasses the linguistic domain. More specifically, this paper will draw attention on the implications of imposing a ‘European’ label onto certain languages actively spoken in Asia; it will be proposed that any such classification must not be taken lightly and, if at all necessary, must combine (diachronic) typological considerations with sociolinguistic factors (e.g. native usage, the languages’ social role(s) or the ethnolinguistic allegiance of the speakers).

2. Factors in linguistic taxonomy

The Portuguese-lexified Creoles of Asia, and Indo-Portuguese in particular, will be invoked to assess the validity of linguistic labelling. As a foreword, though, it is convenient to clarify that linguistic taxonomy operates on a number of different levels and may not be entirely straightforward.

The application of labels to languages always needs to be carefully motivated given that, in reality, the classification of languages into any sort of groupings can be done according to a variety of criteria. The consequence of this observation is that there is no unified interpretation for labels such as ’Nilo-Saharan’, ’Amerindian’, ’isolate’ or ’Asian’. The following is an enunciation of possible classificatory criteria, at once the instruments of classification and its connotations. The list is not intended to be exhaustive; in fact, I would posit that there can be as many clusters and rankings of criteria as there are authors dealing with
language taxonomy. The bottom line is that language labelling cannot be proposed or employed without a clear statement of the facts observed.

2.1. Typological factors

Linguistic typology deals with the classification of languages into types as defined through cross-linguistic comparison of linguistic structures, across genetic families. In modern typological research, a type is defined as any of the strategies observed in the world’s languages to perform a comparable function; as an illustrative example, let us recall that languages vary as to the structures employed to mark possession, with recourse to adpositions, pronouns, affixes and so forth; after compiling the attested structures, it is therefore possible to classify languages according to the strategy(ies) they make use of. As a caveat, some authors (v. Croft 2003: 42) caution that any language may resort to more than one of these.

It is tacitly assumed by most researchers that, ultimately, linguistic taxonomy must reflect typological evidence. In other words, languages classified together must share linguistic traits, or it must be possible to reconstruct the processes of drift that made them diverge. This is true of both typological classification and genetic classification, with the crucial difference that the latter requires these traits to have been inherited from a common ancestor whereas ancestry is not relevant for typological research. Given the amount of variation observed between human languages, any genetic taxonomy may at best hope that the proposed language families reflect an overlap of a significant number of types attested for its members. Whereas in some cases (e.g., the Ibero-Romance varieties) the typological match is relatively straightforward, in other cases it is rather volatile.

Typological correspondence can be sought and defined in several domains and sub-domains of linguistic analysis. In other words, two languages may be classified together or kept apart based on lexical,
morphosyntactic or phonological (dis)similarities. Languages and language varieties resulting from extensive contact between typologically unrelated parent languages (such as creoles are usually said to be, but also pidgins and mixed languages) present a particularly interesting challenge for genetic classification, as one may find that different domains owe more to one or another of the ancestral languages, or even that more than one of the parent languages have left a roughly equal imprint on a specific domain.

Saramaccan, a creole language of Suriname, is an interesting case in point. If its lexicon alone were considered, one would have to classify it alongside its Indo-European ancestral languages (English and Portuguese); but even then the classification within the Romance or the Germanic subfamilies would be disturbed by the fact that English is dominant only in certain sections of the lexicon (adjectives and function words) with Portuguese providing most of the lexemes in the verbal domain. The figures in Table 1, adapted from Smith and Cardoso (2004)\(^1\), reveal the skewed nature of the Saramaccan lexicon.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|}
\hline
 & English & \% & Portuguese & \% & Total \\
\hline
Verbs & 121 & 44\% & 154 & 56\% & 275 \\
Nouns & 179 & 50\% & 176 & 50\% & 355 \\
Adjectives & 51 & 65\% & 27 & 35\% & 78 \\
\hline
\end{tabular}
\caption{Relative contribution of English and Portuguese to Saramaccan lexicon}
\end{table}

The data in Table 2 show the dominance of English-derived morphemes among Saramaccan’s function words. This fact has been interpreted as an indicator that English was the earliest European lexifier, and it makes the superlative contribution of Portuguese to the verbal and nominal domains all the more striking.

\(^1\) Notice that, in this count, only English- and Portuguese-derived lexemes are contemplated, despite the fact that a small part of Saramaccan’s lexicon is derived from West African, Dutch and Amerindian etyma; the percentages reached are therefore only comparative and do not refer to an overall picture of the language’s lexicon.
Table 2: Etymological origin of Saramaccan function words

<table>
<thead>
<tr>
<th>English</th>
<th>Portuguese</th>
<th>Dutch</th>
<th>African</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.8%</td>
<td>16%</td>
<td>14.5%</td>
<td>6.8%</td>
</tr>
</tbody>
</table>

(from Smith 1987)

The matter of genetic classification is further complicated as soon as one moves on from the lexicon into the area of morphosyntax. Much has been written on the antecedents of Saramaccan morphosyntax, a result no doubt of the realisation that Saramaccan behaves, with respect to this linguistic domain, in a markedly non-Indo-European way. Bickerton (1984) regarded Saramaccan as the clearest example of how the process of creolisation draws heavily on universally unmarked structures (his Bioprogram Hypothesis, reminiscent of the Chomskyan concept of Universal Grammar) – thus shunning the structural input of both its West African and European ancestral languages. McWhorter (1997) has also singled this language out as one of the ’deepest’ creoles, a label that in his model places Saramaccan particularly close to the prototype of a proposed typological class of ’Creole languages’. Other researchers (e.g., Kouwenberg and LaCharité 2004; Winford and Migege 2007) prefer to emphasise the presence in Saramaccan of linguistic traits from the African languages involved in its formation. Apart from the obvious variety of conflicting hypotheses, the element to be retained here is that a genetic classification based on morphosyntactic considerations alone would yield a much fuzzier result than one focusing solely on the language’s lexicon, and possibly even contradictory. In turn, an analysis of Saramaccan’s phonological system, with features such as tone and implosives (v. Smith and Haabo 2007), establishes important connections with the West African languages involved in its formation. For similar considerations concerning Indo-Portuguese, see section 3.1. below.

Bearing this in mind, several authors have set out to identify the most stable areas of grammar, the ones less pervious to the effects of
The Meaning of ‘European’

diachronic change as well as contact-based change, and therefore more reliable for genetic classification purposes. Thomason and Kaufman (1988: 5-8) survey these attempts; for Sapir (1921), the ‘”deeper” kernel’ corresponded to morphology, in particular inflectional morphology, while Weinreich (1958) considered a language’s basic vocabulary to be the most reliable indicator of its phylogenetic affiliation. However, as Thomason and Kaufman realise when looking at high-contact varieties, ‘neither the inflectional morphology nor the basic vocabulary is sufficiently stable internally and sufficiently impervious to restructuring or replacement through foreign interferences to justify giving it status as the sole criterion for genetic relationship’ (1988: 6). In their view, in case of high-contact, it may become impossible to insist on the genetic link with any of the ancestor languages on linguistic grounds alone. Whether or not one adheres to this position, the fact remains that, when dealing with those languages traditionally called creoles, one may be inclined to classify them into different taxonomic categories (West African, European, Germanic, Romance, ‘Creole’) depending on the level of linguistic analysis in question. In addition, even within a particular domain of analysis evidence can be rather fuzzy – any attempt to classify Saramaccan as an Indo-European language, West African or typologically ‘Creole’ would necessarily require scraping out the considerable typological correspondence with members of the neglected group(s).

2.2. Demographic factors

The definition of language families and genetic relationships is often envisaged to mirror demographic continuity, with ancestral languages being spoken by ancestral populations. Conversely, where a language family has been linguistically defined without direct evidence of a shared population history, such history is often assumed as a result – notice for instance the anthropological and archaeological studies aiming at identifying the Indo-European people and tracing their mo-
vements across the land. However, this link between linguistic tradition and population continuity was recognised from very early on to be problematic. In his 1922 research in search of ‘the home of the Indo-Europeans’, Bender makes it clear that:

[1]linguistic relationship is not in itself sufficient proof of racial relationship. The conquered may adopt the language of the conquerors, or the conquerors may adopt the language of the conquered, or there may be peaceful mingling in irregular proportions of race and language. (...) When we speak of the Indo-Europeans we mean merely the people, whoever they were, that spoke Indo-European, and we imply nothing whatever as to race or racial characteristics. (1922: 10)

In contexts involving the transplantation of a language from one region into another, claims of demographic continuity between the perceived area of origin and the perceived target area are often articulated to legitimise the status of the language in society. In more concrete terms, one finds that if, for instance, a certain community in Asia speaks a (non-metropolitan) variety of what is perceived as a European language, this fact may be invoked as evidence for linear descendant. It will be argued in section 3.1. that, in the case of the present-day Indo-Portuguese-speaking communities, this hypothesis is only partially borne out.

The myth of demographic continuity has often brought to light racial considerations within the genetic debate, even if the very concept of ‘race’ is highly problematic. Mufwene (2008) reasons that, while biology and physical anthropology increasingly stress the social foundations of racial categories, linguistics has not yet questioned the biological reality of the concept and echoes it in claims amounting to the exceptional nature of Creoles. The field of creole studies is ripe with controversy concerning the neo-Darwinian motivations of what DeGraff (2003, 2005) calls ‘Creole exceptionalism’, which he defines as follows:

Creole Exceptionalism is defined as a set of beliefs, wides-
pread among both linguists and nonlinguists, that Creole languages form an exceptional class on phylogenetic and/or typological grounds. It also has nonlinguistic (e.g., sociological) implications, such as the claim that Creole languages are a handicap for their speakers, which has undermined the role that Creoles should play in the education and socioeconomic development of monolingual Creolophones. (DeGraff 2005: 533)

Several claims have been formulated to the effect that creolisation is a peculiar process yielding peculiar results, including Bickerton’s (1984) Bioprogram Hypothesis, McWhorter’s (1997, see previous section) proposal of a typological class of Creole Languages and Thomason and Kaufman’s (1988) proposal that creoles are the result of a break in language transmission. DeGraff maintains that the very essence of this view is tinged with notions of racial inequality; in his view, the drive behind linguist’s attempts to separate creoles from their lexifiers cannot be dissociated from ideas of racial and linguistic ‘purity’ on the part of the ‘metropolis’. High-contact varieties which developed in a colonial context and their speakers often find themselves in the uncomfortable position of being looked down on by the speakers of the very languages promoted as the norm. Evidently, it is possible to interpret some speakers’ and researchers’ will to break the taxonomic bonds with the lexifier languages as a reaction to the implications of the observed prestige asymmetry, the very ‘handicap’ DeGraff opposes. Essentially, the ongoing debate concerning the ideological foundations of theories of creole genesis is a powerful reminder of the detrimental potential of genetic taxonomy.

2.3. Ethnocultural factors

The previous considerations on demography introduce a related (but not equivalent) implication of associating a language with a particular ancestor, viz. that linguistic continuity is interpreted as an indica-
tor of ethnocultural continuity; this opens up the possibility of manipulating language to reinforce this perceived cultural link. It is a known fact that e.g. the rise of national identity of a subjugated area is often played out in the linguistic arena, with languages previously considered peripheral varieties (in relation to a metropolis) undergoing a campaign for separate linguistic status. In other cases, the absence of such a separatist drive is equally meaningful, as it is likely to reflect more benign attitudes of two language communities towards one another: language maintenance and a desire for taxonomic integration are strengthened whenever the speech community welcomes the type of ethnocultural relationship thereby reinforced. Crucially, whether or not a perceived ethnocultural lineage is accompanied by demographic continuity is not necessarily relevant. In many cases, the vehicle of acculturation may have taken the form of education, religious conversion or any other, and still have resulted in strong ethno-cultural affinities.

It is precisely here, at the interface between politics, identity and language, that the issue of genetic taxonomy becomes sensitive. This is an issue that may concern linguists inasmuch as linguistic data can be adduced to the debate but it turns out that, more often than not, linguistic data is moulded in the shape of non-linguistic arguments. Mufwene (2007) cautions against the ideological echoes of linguistic classification; the author considers that the decision whether a particular Creole should be classified along with its lexifier or not is a matter that does not depend so much on linguists as on the wishes and cultural allegiances of its speakers:

[...] les linguistes n’ont ni le pouvoir ni l’autorité de stipuler si un parler est un dialecte ou s’il s’agit d’une langue ayant évolué à partir d’une autre, ou encore s’il s’agit d’un parler ayant divergé par rapport à d’autres parlers qui ont évolué simultanément avec lui. La question est à la fois celle de l’“auto-identification” et de l’“identification de l’autre”. Il revient aux seuls locuteurs de décider de leur apparentement à
des locuteurs d’autres variétés.¹ (Mufwene 2007: 61)

Mufwene (p.c.) defends that it is the role of linguists to demonstrate the multi-faceted genetic affiliations of a particular variety without pronouncing themselves of the status of this variety vis-à-vis others. This position undermines the scientific validity, necessity and indeed the impartiality of linguistic taxonomy and is therefore in line with my proposal that such terminology should be clearly kept confined to the realm of pragmatism. Considering the social undertones of language parentage, though, I should add to Mufwene’s idea that the only active role linguists can claim in this matter is that of ensuring, to the extent that this is acceptable, that the decision of the community is informed and free from any type of pressure, be it on the part of governments, neighbouring social groups, institutions, the historical metropolis, etc.

2.4. Geographical factors

Linguistic labels are usually fastened to geographical terminology, and conveniently so for clarity. However, it must be pointed out that often there is no one-to-one correspondence between the physical boundaries defined by the label and linguistic boundaries. Recall, as an example, that German has national language status not only in Germany in its present-day shape but also in Austria, Switzerland, Belgium, Luxembourg and Liechtenstein. One may contend that all these territories are adjacent, making it possible to define the entire region as the ‘locus’ of German. Consider, however, the existence of significant German-speaking populations around the globe (e.g. Transylvania or Paraguay) and instances of what could according to some views be seen as non-metropolitan varieties of German, such as Pennsylvania

¹ ‘linguists do not have the power nor the authority to determine whether a variety is a dialect or whether it is a language which evolved from another, or still whether this is a variety divergent in relation to other varieties which evolved simultaneously. The matter is at once one of “self-identification” and “identification of the other”. It is the prerogative of the speakers alone to decide on their relation with speakers of other varieties.’ [Transl. H.C.]
Dutch. In fact, geographical labelling becomes less linear when we are dealing with the transplantation of languages from their perceived area of origin, which is generally the case with migratory movements and very much so with colonial expansion.

This type of population/linguistic movement often results in the development of varieties seen as ‘peripheral’ in relation to a ‘metropolitan’ variety; the reference centre may or may not be dislocated in space in the process. When dealing with high-contact varieties, it is not uncommon to invoke this type of vertical classificatory structure. Linguistic analyses that stress out the typological closeness between creole languages and their (mostly European) lexifiers (e.g. Chaudenson 1992) may not intend to perpetuate a hierarchical view of things, but their arguments run the risk of being interpreted in that light. Whereas one may deem this a harmless position from a scientific point of view, it is not devoid of significance for both the speakers of the ‘peripheral’ and the ‘metropolitan’ varieties; whether or not such a hierarchical structure is perceived as threatening or at least undesirable is a function of the relationship between the ‘metropolis’ and the ‘periphery’, and can only be decided by the speakers themselves. One common reaction to this asymmetric situation, as mentioned earlier, takes the form of a campaign for the taxonomic dissociation of the creole from the language seen as its main precursor, often accompanied by the selection of local norms, the extension of the language’s domains of usage and even local standardisation. Taking as an example the Portuguese-lexified creoles of West Africa (spoken in Cape Verde, Guinea-Bissau, São Tomé e Príncipe, Equatorial Guinea) and the increasing social acceptability of some of these varieties on a national scale (in all of the countries mentioned, with the exception of Equatorial Guinea), one begins to understand how the process by which the creoles cease to be considered peripheral dialects of a metropolitan language can be beneficial to their maintenance. In other words, if it becomes established that a particular creole is more than just a dialect of (European) Portuguese, it is free (if relevant) to assume its status as a national, Gulf of
The Meaning of ‘European’

Guinea, West African and African language – and taxonomy is free to reflect this sociopolitical reality.

Granting a high-contact variety a separate entry in the language family tree will force a segmentation between language A and language B, even if (as in the case of the Portuguese-lexified creoles of Africa) they continue to coexist in a given geographical unit. Such is also true of certain French-lexified creoles which co-exist to this day with standard French and a variety of French known locally as patois. Posner (1985: 171) claims that:

Overseas French is of particular interest to the historical linguist precisely because a distinction is frequently made between patois, used principally by white speakers whose language loyalty is, broadly, to French, and creole used largely by non-whites who identify with a non-French culture in some, at least, of their activities.

The author then goes on to adduce linguistic differences between the two varieties to justify the distinction, insisting on the fact that these differences (e.g., the existence of nominal gender and verbal inflection in patois) place patois closer to ‘metropolitan popular and dialectal French’. In this excerpt, the classification of both varieties as ‘overseas French’ indicate that the author considers them both as varieties of French, as does the allusion to a possible linguistic continuum linking both creole and patois in places such as Réunion, Guadeloupe and the Virgin Islands (Posner 1985: 171). Some authors (e.g. Chaudenson 1992, Mufwene 2007) are inclined towards this position while others (e.g. Thomason and Kaufman 1988) tend to advocate a taxonomic cleavage between creole and patois/French; in the end, though, the question is essentially extra-linguistic.

2.5. Social factors

As alluded to earlier, there may be a mismatch between the factors that have been taken in as classificatory and the possible ‘readings’ of
such a classification. One particularly sensitive issue, of relevance to a fair number of high-contact varieties but also any variety with a history of spacial displacement, is the potentially alienating effect of applying a geography-bound label to a language spoken in a different region. To avoid harm to these speech communities, one should make sure no classification detracts from the actual social roles performed by the language. It may be the case that a particular linguistic community wishes to imply its link with a foreign land, but the opposite may also be true. Once more, any decision on the matter will be mostly sociopolitical, and linguists would be wise not to engage in the discussion.

A similar point may be made for all other types of labels. In the context of Goa, for instance, one may feel tempted to apply an ethnocultural label on the use of Portuguese by claiming, as Goans often do, that ‘Portuguese is the language of the Catholics’. While patterns of language loss seem to reflect this, such a claim would be rather simplistic in that it would exclude a large number of non-Catholic speakers; in addition, the issue may be politicised against this particular ethnocultural community if ever the sociopolitical atmosphere should become adverse to all remaining colonial echoes.

The truth of the matter is that language interacts with social labels in very complex ways in order to create, maintain or alter social identity. LePage and Tabouret-Keller allude to the myriad social roles of language use, as they notice:

> [n]ational, ethnic, racial, cultural, religious, age, sex, social class, educational economic, geographical, occupational and other groupings are all liable to have linguistic connotations. The degree of co-occurrence of boundaries will vary from one society to another, the perception of the degree of co-occurrence will vary from one individual to another. (LePage and Tabouret-Keller 1985: 248)

Most linguists would agree that it is not their task to interfere in patterns of language use against the wishes of the speech community; delving deeper into the implications of linguistic terminology, I will
propose that linguists should not allow their taxonomic labels to interfere with the processes of identity-construction and/or social (dis)integration. This can only be achieved with some caution and sensitivity when applying linguistic classificatory terminology.

3. The Portuguese-lexified creoles of Asia

At present, a small number of Portuguese-lexified creoles are still spoken in Asia, with an area of implantation much narrower that it once was. The formation of these creoles was set in motion by the contact between Asia and Portugal from the 16th-century onwards, a trading venture that eventually turned into colonial domination in several coastal settlements. Lopes (1936) surveys the references to the use of the Portuguese language found in old European documents and makes it clear that, from at least 1545 onwards, Portuguese or a Portuguese-lexified pidgin were being used as media for communication between Europeans and Asians. The role of Portuguese as a means of interethnic communication was such that other European nations felt the need to employ Portuguese speakers in their voyages as translators (v. Lopes 1936: 32), corresponded and negotiated treaties with Asian rulers in this language (v. Lopes 1936: 29, 35, 40). It is clear from several of these accounts that the language referred to is not standard Portuguese; the use of expressions such as ‘bastard Portuguese’, ‘patois’ or ‘corrupted Portuguese’ is rather transparent. The accounts in Lopes (1936) are taken from descriptions of virtually all corners of Asia, including India, the Malay Peninsula, China, Sri Lanka, the Comoros, the Persian Gulf, Japan, Nicobar and Indonesia.

The privileged role of Portuguese(-lexified pidgin) as a lingua franca in Asia is not surprising given the fact that the Portuguese were the first European nation to have a sustained presence in Asia, and also that their trading and/or colonising efforts were scattered across the entire region. In several of those locations, we observe the development of Portuguese-based creoles, in particular where the political presence
of Portugal was stronger and where European and Asian-European communities formed. South Asia, the original focus of Portuguese expansion in the Orient, was home to a particularly high concentration of these varieties, extending from the Gulf of Cambay to the Gulf of Bengal; creoles were recorded in places such as Diu, Surat, Bassein, Bombay (Mumbai), Chaul, Mangalore, Cannanore (Kannur), Calicut (Kozhikode), Cochin (Kochi), Nagapattinam, Pondicherry (Puducherry), Vishakapatnam, Pipili, Hoogli, Dacca, Chittagong as well as Colombo and Batticaloa in Sri Lanka, to name but a few. Elsewhere in Asia, these varieties have been documented in the regions of Malacca, Singapore, Batavia (Jakarta), Macao and possibly Timor and Flores. Little or no linguistic data is available for a number of these varieties, while others are more or less well documented. Considerably old documents are known for the varieties of the Batavia region and Ceylon (Sri Lanka); the combined efforts of Hugo Schuchardt and Monsenhor Sebastião Rodolfo Dalgado meant that we now have late 19th-c. and early 20th-c. data for several of these varieties, including Cochin (Schuchardt 1882), Diu (Schuchardt 1883a), Mangalore (Schuchardt 1883b), Mahé and Cannanore (Schuchardt 1889b), Batavia/Tugu (Schuchardt 1890), Sri Lanka (Dalgado 1900a), Daman (Dalgado 1902-3), Bassein/Bombay (Dalgado 1906), Nagapattinam (Dalgado 1917) and the Portuguese of Goa (Dalgado 1900b). 20th-century linguistic accounts of Asian creoles include those concerning Malacca (Baxter 1988, Rêgo 1998), Sri Lanka (Smith 1984), Macao (Batalha 1977) and Korlai (Clements 1996).

It has often been noticed that the Portuguese-lexified creoles of Asia share many structural and lexical traits (e.g. the structure of possessive constructions, part of the paradigm of TAM markers, lexemes such as ada/ade/adi ‘duck’) regardless of the languages they came into contact with, forming a linguistic unit opposed, for instance, to the Portuguese-lexified creoles of West Africa (Ferraz 1987). Dalgado (1917) coined the notion of ‘partial reciprocal transfusion’, according to which the various Portuguese-lexified creoles of Asia kept close contact with
each other in the early colonial period (from 1498 onwards) through a history of population flow and a centralised administrative structure with its seat in Goa, and were therefore able to influence each other. Clements (2000) suggests that the Portuguese pidgin documented across Asia from as early as the 16th century was distinct from (though possibly not unrelated to) its African counterpart and fed into the Asian creoles, therefore accounting for their similarities.

At present, Portuguese-lexified creoles remain in Diu, Daman, Korlai, Kannur and Kochi (India), Malacca (Malaysia) with offshoots in Singapore, the Macao/Hong Kong region and possibly in Sri Lanka. The following section will draw on data from one of these varieties to assess the issue of genetic classification in an Asian context.

3.1. Challenges of Diu Indo-Portuguese for linguistic taxonomy

The term *Indo-Portuguese* (henceforth IP) is generally employed to refer to all South-Asian Portuguese-lexified Creoles. As shown in the previous section, IP therefore refers to a cluster of varieties with somewhat different social histories and different linguistic influences. I shall concentrate for the following discussion on the Diuese variety (cf. Cardoso 2006), whose territory falls unequivocally under the Gujarati cultural sphere.

Considering the discussion in section 2, the task of deciding whether to classify Diu IP as a Romance or an Indic language or possibly something else seems rather futile. Considering that all present-day native speakers of Diu IP are fluent in Gujarati and that in their speech they juggle these two languages with a high degree of proficiency, the description of the main etymological parentage of Diu IP is anything but straightforward. But even if one accepts the claim that the bulk of its lexicon derives from Portuguese, there are several caveats to take into account.

First, one can observe many instances of accommodation of the
Portuguese etyma to Indic semantic patterns; one case in point is the Diu IP word *pε*, which designates both ‘foot’ and ‘leg’ where Portuguese has *pé* for ‘foot’ and *perna* for ‘leg’; the same goes for Diu IP *mãw* ‘hand, arm’ where Portuguese distinguishes between *mão* ‘hand’ and *braço* ‘arm’. Diu IP shares this semantic pattern (common across languages of the Indian subcontinent) with Gujarati, which uses *pag* for ‘foot’ and ‘leg’, and *hāth* for ‘hand’ and ‘arm’.

In addition, it seems that throughout the history of Diu IP, standard Portuguese functioned as a kind of norm, or at least a privileged contributor to the lexical pool. Nowadays, with standard Portuguese and Portuguese-medium education far removed, one finds, as expected, that Gujarati and also English establish themselves firmly as sources of more or less permanent borrowings into the language.

Finally, it is important to recognise that many common words in Diu IP are characteristic of Asian Portuguese rather than standard Portuguese and are ultimately borrowings from Asian languages. One such example is Diu IP *jag* ‘jaggery’, which Dalgado (1900b) related to Konkani *sākar* and later (see Dalgado 1913) to Malayalam *chakkar*; this word is also found in other IP varieties (cf. Clements 1996) and Goan Portuguese, as well as other Asian Portuguese-lexified creoles (v. Rêgo 1942, reprint 1998; Scully and Zuzarte 2004 for Papia Kristang).

As far as the language’s morphological array is concerned, it is striking how little is expressed by morphological means in Diu IP in comparison with both Portuguese and Gujarati. To mention but one example of this, while both of its ancestral languages require the morphological expression of gender (feminine and masculine in both plus neuter for Gujarati) and number (singular and plural) in all nominal heads with agreement carry-over to its dependent elements, Diu IP does not mark gender or number morphologically¹. Therefore, a Diu IP word such as *fil* is unspecified for gender and number, referring either

---

¹ There are certain exceptions to this rule in the pronominal paradigm of Diu IP, in particular the opposition between the 3ps masculine pronoun *el* and 3ps feminine *el*. The truth remains that nouns are not morphologically marked for gender or number in Diu IP.
to ‘a son’, ‘a daughter’, ‘sons’ or ‘daughters’; disambiguation is either contextual or lexical: the word raprig indisputably refers to a ‘girl’ and plurality may be indicated by preposing the element tud to the noun (e.g. mĩ tud raprig ‘my daughters’).

To complicate matters further, there are certain grammatical traits of Diu IP that bring it typologically in line with Gujarati. One of these refers to the order of elements within the noun phrase, in particular the position of the attributive modifier in relation to the nominal head. In Diu IP, the modifier is consistently preposed to the head of the NP, as in the following example:

(1) ikəl kamiz ku verd i amrel flor
DEM shirt with green and yellow flower
‘that shirt with green and yellow flowers’

In (1), the complex modifier verd i amrel is preposed to its head flor. This pattern is in tune with Gujarati in which, in attributive functions such as this, adjectives are always preposed to the nominal head. In Portuguese, by contrast, the order of the adjective in relation to the noun it modifies is optional, with a strong preference (in modern Portuguese at least) for a postnominal position.

Seeing as the typological evidence cannot with clarity support any taxonomic decision, let us now consider whether or not any strong demographic link can be posited between the speakers of Diu IP and a) a traditionally Portuguese-speaking population, or b) a traditionally Gujarati-speaking population. A crucial point to be made here is that, even if nowadays native competence in Diu IP is reduced to the Catholic section of the island’s population, the language is still very much in use among older members of the Hindu and Muslim communities, sections of the population that do not claim any descent from Portuguese settlers. Some among the Catholic community do, however. As noted in Cardoso (2007),

---

2 An adjective can follow a noun only in a predicative function, cf. Cardona 1965; this is also true of Diu IP although (unlike Gujarati, which is consistently verb-final) a copular element may separate the noun phrase from the postnominal predicative adjective.
The fact that they retain Portuguese family names is normally seen as indication of that; there is the danger of some circularity here, however, as it is known that new converts, whether or not of mixed descent, would usually adopt a Portuguese surname.

Historical accounts do allow for a considerable number of mixed lineage families (cf. Boxer 1963), but it would be unwise to claim that all Catholic families in Diu have Portuguese ancestry. If we couple this observation with the previously mentioned fact that IP was never exclusive to the Catholic community, one would have to recognise that Diu IP has been spoken throughout its history by a population of overwhelmingly Indian extract, with only comparatively marginal participation of persons of (mixed) European descent and also of African descent.

A social survey of Diu IP usage such as that in Cardoso (2007) makes it clear that the language is essential in defining the boundaries and interactions of sections of the Diuese population; this happens in complex ways and is by no means restricted to the Catholic community. One crucial element to bear in mind is the coexistence, to this day, of both Diu IP and standard Portuguese on the island. The two varieties have very different social connotations and diverse social roles that betray a tension between a reality in which Diu IP is used natively and a prestige structure that assigns normative power to an L2 variety (viz. standard Portuguese) with very limited distribution among the population. Knowledge and usage of standard Portuguese and/or IP in Diu are powerful elements in linking back to Goa, to Europe and to a cosmopolitan past – notice for instance the use of standard Portuguese in administration and Catholic mass, among other social domains. However, it also becomes clear from an observation of present-day patterns of language use that standard Portuguese is considered the target language only when such link is specifically aimed at, while strong local allegiance derives from IP’s ‘vernacular’ status. Some potentially detrimental consequences of the differential in prestige between IP and
standard Portuguese include the fact that it perpetuates social ranking along linguistic lines and also that it prevents IP speakers with limited or no competence in standard Portuguese from interacting freely with outsiders or the highest ranking figures within the community.

4. Conclusion

The previous discussion alerted to the fact that any linguistic label is open to a diversity of interpretative possibilities, not all of which are harmless on a political or ideological scale. It has been argued that, although typology has a word to say in linguistic classification, genetic taxonomy is an enterprise that transcends linguistic facts and in the process may or may not take into account demographic considerations, ethnocultural histories and social realities: crucially, it may or may not be unaffected by political, ideological and identity agendas.

In the particular case of languages actively spoken in Asia which are ultimately seen as transplantations from Europe, I would argue that the application of the ‘European’ label (in particular when not accompanied by a clarification of the label’s intended scope) is problematic; not only may it be harmful, it is also hardly straightforward. The treatment of Diu Indo-Portuguese as a case study has demonstrated how unwarranted some of the possible interpretations of the ‘European’ label may be. Any taxonomy insisting on this language’s ‘European’ status would only partly reflect some of its speakers’ allegiances and ethnocultural traits, as well as part of the typological evidence; but it would also sacrifice the facts concerning the language’s local history, local ethnodemographic links, its social roles and significance on a local sphere and all its Indic (and innovative) linguistic traits. Furthermore, identifying the language and its speakers exclusively with a past colonial period may end up alienating them from their newfound local and national circumstances.

I therefore caution against the indiscriminate use of labels of this
type (and in particular those that allude to a distant ‘metropolis’) when referring to Diu IP. In the same vein, I caution against the indiscriminate application of any genetic label to high-contact varieties. If one subscribes to the axiom articulated by Mufwene (2007) according to which it is not within the realm of linguistic research to decide on the genetic affiliation of these languages, one has to urge linguists to take a sensitive and more spartan approach to language labelling.

References


— (1889a) ‘Beiträge zur Kenntnis des kreolischen Romanisch. V. Allgemeineres über das Indoportugiesische (Asioportugiesische)’.

Zeitschrift für Romanische Philolo-
gie 13: 476-516.


