Exploring variation among the last speakers of Malabar Indo-Portuguese

Hugo C. Cardoso
Universidade de Lisboa, Faculdade de Letras

Despite their ancestrality among the Asian-Portuguese creoles, those of Southeastern India are currently in an advanced state of decline. Collectively known as the Indo-Portuguese creoles of the Malabar, or Malabar Indo-Portuguese, there were once significant speech communities in many towns, but over the centuries, the communities shifted heavily towards English and Malayalam, the dominant language locally. Language documentation was carried out since 2006 with the assistance of the very last fluent speakers of Malabar Indo-Portuguese in Cannanore (6 speakers), Cochin (1 speaker), and Calicut (1 speaker), resulting in a corpus which offers a glimpse of the language at the tail end of a process of loss, when it is no longer used on a daily basis (with the exception of a single homestead) and, in the case of the most geographically isolated speakers, hardly ever spoken at all.

Here, I explore certain instances of variation that can be identified in the corpora, especially with respect to functional morphemes. The specific present circumstance of Malabar Indo-Portuguese – a dwindling minority language, geographically diffuse, and used in highly multilingual settings – poses considerable challenges to a clear interpretation of such instances of variation. While certain cases – especially if they involve the observable interference of the speakers' current dominant languages, English and Malayalam – may reasonably be interpreted either as cases of code-switching, as nonce borrowings, or as the product of language obsolescence, others appear to reflect subgroup preferences, including family-internal specificities and geographical variation, although the small number of informants does not entirely clarify this. In addition, some other instances of variation contradict the obsolescence scenario, such as e.g. the preference for a Malayalam-derived oblique case-marker over a Portuguese-derived one, which is in fact associated with the most fluent creole speakers, rather than those that show the clearest effects of language obsolescence. This is an exercise in the study of variation which simultaneously draws attention to the variability of creole languages (not only across relatively vast areas, but also in closely-knit communities) and highlights the specific challenges of conducting linguistic research among the last speakers of a particular language.

Palavras-chave: Malabar, Indo-Portuguese, endangered languages, language documentation, variation, corpora

1. Introduction[1]

The Asian-Portuguese creoles that are currently spoken in parts of India, Sri Lanka, Malaysia, and Macau are but a small subset of the ones which formed along the Asian shores from the 16th century on, most of which ended up disappearing as their communities shifted to other, locally-dominant languages (Smith 1995; Baxter 1996; Cardoso 2020). In Southwestern India, specifically the current Indian state of Kerala, Portuguese-lexified creoles are still spoken, but by a minute number of speakers, as a result of a process of decline which began at least in the 19th century – if not earlier. Since this was the first region in Asia to witness the establishment of Portuguese settlements, it is likely that those languages, collectively known as the Indo-Portuguese creoles of the Malabar, or Malabar Indo-Portuguese [MIP] (after an early name given to that coastal stretch), were the earliest Asian-Portuguese creoles to be formed. There were once significant speech communities in many towns, including Mangalore [Mangaluru], Cannanore [Kannur], Tellichery [Thalassery], Mahé, Calicut [Kozhikode], Cochin [Kochi], and Quilon [Kollam], to name but a few (see Cardoso, Hagemeijer & Alexandre 2015; Cardoso 2019 for a fuller account). However, over the centuries, as a result of important sociopolitical changes, the communities shifted heavily towards English as well as Malayalam, the dominant language locally, and now MIP runs the risk of extinction in the short term.

Despite this, with the collaboration of the very last speakers of MIP, it has been possible to collect an oral corpus of the language over the past 2 decades, on the basis of which I will explore some instances of variation in this study. The aim here is to reflect on how linguistic variation can be approached and interpreted in a corpus composed of data provided by only a few speakers who constitute the totality (or, at best, near-totality) of the speech community and what constraints that places on the type of analysis that can be carried out. The matter is made even more challenging by two facts: a) that the language was almost entirely undescribed by the time the corpus began to be constituted, which means that, with no reliable source with which to compare the corpus, all the data available for linguistic description is the somewhat variable evidence provided by the last speakers; and b) that it is no longer used equally frequently or fluently by all speakers, which raises the possibility that some features of

their speech may be the result of language attrition and/or show the impact of the language the community is shifting to.

The manifold challenges of working with "last speakers" have, of course, been identified before. Evans (2001), for instance, lists a number of issues that may be involved, ranging from the very identification of a language's "last speakers" (which rests on considerations about which linguists and community members may not necessarily agree, such as the definitions of "language" and "speaker", or notions of ownership of a language and authority to speak) to the classification of their fluency, the shifting nature of language command and use within a given community, and potential barriers to accessing the full extent of a speaker's knowledge of their language. One reflection which is especially relevant here is the degree of uncertainty about the extent to which descriptions based on the speech and knowledge of such speakers represent what the language was like before it began its process of contraction:

> "Especially where little or nothing is known about the language, large domains of one's description are potentially open to three types of interpretation: that the speaker has simply failed to master the full complexities of some grammatical, phonological, or semantic domain; that the variety they mastered had already undergone simplification through contact with the replacing language; or that the language was in fact like that all along, and the speaker has actually mastered it as perfectly as speakers one hundred years before." (Evans 2001: 261)

The identification of instances of intrapersonal or interpersonal variation in the corpus relates to this in that variation may either be interpreted as a long-standing feature of the language or as a consequence of the process of language loss, associated either with transfer from the dominant language or with some other processes (such as e.g. overgeneralization of certain features, paradigmatic reduction and stylistic shrinkage; see Campbell & Muntzel 1989) that may accompany language obsolescence.

In fact, encountering variation while conducting any documentation of an undescribed or underdescribed language presents an additional set of challenges, as has also been amply discussed before (see e.g. Nagy 2009; Mansfield & Stanford 2017; Meyerhoff 2017). These derive from circumstances such as the paucity of data (synchronic and/or diachronic) and incipient nature of the grammatical analysis, the small number of speakers providing data, or the fact that the researcher, typically an outsider to the language and the community, "lacks the linguistic and ethnographic knowledge that Sankoff noted is needed to accurately describe and interpret variability that is sociolinguistically constrained" (Meyerhoff 2017: 528). As a result of these constraints, as

Mansfield & Stanford (2017: 116) note, "many intriguing instances of variation end up as a footnote or parenthetical comment in descriptive grammars, rather than receiving the focused attention they deserve".

However, these authors agree that, despite the necessary caveats, addressing variation during language documentation and description is not only possible, but also desirable. It is from this perspective that I set out to explore a few instances of variation identifiable in the corpus of MIP. I start by taking a quick tour of the history of MIP, emphasizing its formation, regional impact, and subsequent decline, in section 2. Section 3 introduces the MIP corpus and its characteristics, while, in section 4, I analyse the variability it displays in a number of linguistic domains, searching different ways of making sense of it. Finally, in section 5, I discuss the study's results.

2. Malabar Indo-Portuguese: rise and decline

The formation of MIP is the result of the establishment of a Portuguese presence on the southwestern coast of India (the Malabar) in the early 16th century, which inaugurated the process of European colonial expansion in Asia. The first Portuguese overseas journey to Asia, that of the armada of Vasco da Gama, made landfall precisely in this region, close to the city of Calicut, on May 20th, 1498. Two years later, the first Portuguese *feitoria* [trading post] was established in Cochin, to which a fortification was subsequently added, and soon others were set up in Cannanore, Calicut and Quilon. For the first few decades of their Asian colonial engagement, the Portuguese were headquartered in Cochin, until the centre of political power shifted to Goa in the 1530s. As suggested in Clements (2000) and Cardoso (2016, 2019), this may have given Cochin and the Malabar an important role in the development of the network of Portuguese settlements that stretched all the way to Japan and Timor, but also a position of primacy in the history of Asian-Portuguese linguistic contact, potentially determining some linguistic features which diffused and now account for a certain degree of similarity between the various Asian-Portuguese creoles (see also Clements 2009; Baxter & Bastos 2012).

The social composition of these Malabar strongholds of the Portuguese in the 16th century was very complex, including not only people from Portugal and from the Malabar coast, but also from other parts of South Asia and Europe, from Africa, and from elsewhere in Asia (see Cardoso 2019). Processes such as the formation of Eurasian families or conversion to the Catholic religion, attested at that time, surely set the stage for the type of language contact that

underlies the development of MIP, although we have no indication of what early MIP was like or how widespread it was among the population of the Malabar. However, we can surmise that MIP was already firmly established by the mid-17th century, because even though the former Portuguese strongholds of the Malabar were conquered by the Dutch in the 1660s, the language was able to survive the sudden political shift and continued to be used by some sections of the population beyond that point in time. Later comments on the characteristics of "Portuguese" as spoken by the various Indo-Portuguese communities of South Asia (the Malabar and elsewhere) hint at great variability within and across communities (an issue studied in Cardoso 2014a). This description, written in 1733 by the German missionary Nikolaus Dal, is quite eloquent regarding the perception that the linguistic behaviour of these communities was socially (and racially) stratified, and comprised several registers associated with different communicative settings:

> *[…] die Sprache nicht bey allen einerley ist. Denn gleichwie man die Portugiesen in drey Classen getheilet hat, so kann man auch drey Arten von der Sprache setzen, nemlich 1) die rechte, 2) die halbverdorbene, und 3) die gantz verdorbene. 1. Die rechte Portugiesische Sprache reden hauptsächlich die Europäischen Portugiesen, und dann auch, die von ihnen herstammen. 2. Die halb-verdorbene reden insgemein, die von vermischter Abkunft sind. Schwartze aber bedienen sich auch derselben in ihren Briefen. Diese Art von Sprache bestehet hauptsächlich darin, daß sie die <u>verba</u> nicht recht nach ihren <u>Coniugationibus</u> zu gebrauchen wissen. 3. Die gantz verdorbene höret man, im gemeinen reden, zwar auch von Portugiesen vermischter Abkunft, doch vornemlich von denen, die gantz schwartz sind. Diese Art von Sprache unterscheidet sich von der vorhergehenden hauptsächlich darin, daß die Leute gar keine <u>Coniugation</u> gebrauchen, sondern nur die künftige Zeit durch die <u>particulam</u> lo, und die vergangene durch ja aussprechen, und sich dabey des <u>Infinitivi</u> mit Auslassung des <u>R</u> bedienen.*
>   *Zu Briefen wird diese Sprache für läppisch und ridicul gehalten. Man sollte auch dencken, daß sie zum Sprechen auch nicht sonderlich geschickt sey; […] Wie denn aus solcher Ursache auch die schwartzen Portugiesen sich der halb-verdorbenen Sprache bedienen, wenn sie aus ihrem Hertzen ein Gebet zu Gott thun. In abgebrochenen Reden aber kann man sich deren noch ziemlich bedienen: wie sie denn auch insgemein gesprochen wird, wenn Engelländer und Dänen, Holländer und Frantzosen, Europäer und Armenier, u.d.g. hier in Indien zusammen kommen, und einer den andern in seiner Muttersprache zu verstehen nicht vermögend ist. (Dal 1733: 919-920)*

[…] the language is not the same for everyone. For as you have the Portuguese divided into three classes, you can also distinguish three types of the language, namely, 1) the proper, 2) the semi-corrupt, and 3) the entirely corrupt. 1. The proper Portuguese is mainly spoken by the European Portuguese and also by those who descend from them. 2. The semi-corrupt speech is generally spoken by those of mixed parentage. But the blacks also make use of it in their letters. This type of language is characterised mainly by the fact that they are unable to use verb conjugations correctly. 3. The entirely corrupt

is generally heard from the Portuguese of mixed descent, but especially from those who are completely black. This kind of language is different from the previous one mainly in that people do not use conjugation at all, instead expressing future tense with the particle *lo*, and past tense with *ja*, and construct the infinitive by omitting R.

 This language is considered silly and ridiculous in letters. One should think that it is not very appropriate for speaking; This is why the black Portuguese use the half-corrupted language when they pray to God from their hearts. But in incoherent speech this language can still be used: it is generally spoken when the English and the Danes, the Dutch and the French, Europeans and Armenians meet here in India and it is not possible for them to communicate in their native languages.] (author's translation)

The few linguistic traits mentioned, especially the TAM markers *ja* and *lo*, clarify that what was considered "entirely corrupt Portuguese" corresponds to varieties of the Indo-Portuguese creoles, since these markers are reflected in various modern Asian-Portuguese creoles (even beyond South Asia; see e.g. Ferraz 1987: 350-351; Baxter 1996: 302). The notion that the written medium tended to require a register seen as closer to standard Portuguese is not only unsurprising, but also consistent with a particular reference to the Malabar, made in a work published in 1786 by the French scholar Abraham Hyacinthe Anquetil-Duperron:

> *Mais il s'en faut bien que ce soit le Portugais pur, appelé dans l'Inde le* Portugais Reinol. *Celui qui s'écrit en approche d'avantage, surtout à la Côte Malabare, où cette Nation a eu de nombreux Etablissements: le Portugais parlé n'est proprement qu'un jargon, consistant en 150 ou 200 mots, presque sans construction.* […]

> [But this language is far from being the pure Portuguese, called in India *Português Reinol*. The one that is written down comes closest to it, especially in the Malabar coast, where that Nation had numerous Settlements: spoken Portuguese is in effect no more than a jargon of 150 or 200 words, with almost no structure.] (author's translation)

We have a convenient example of this in a letter written in 1885 to Hugo Schuchardt by a member of the Indo-Portuguese community of the Malabar, a gentleman identified as Mr. d'Cruz and residing in Mahé (by then a French colony). Mr. d'Cruz was clearly a proficient speaker of MIP, since he also provided data for that language, but his letter[2] – transcribed below – attests to a relatively accomplished command of Portuguese, despite a few deviations from the European standard (for further details, see Cardoso 2014a; 2015a):

---

[2] Manuscript preserved in the University of Graz, *Hugo Schuchardt Archiv*, with reference nr. 01-02067. Available online at: https://gams.uni-graz.at/o:hsa.letter.1999/sdef:TEI/get?context=context:hsa.letters.1342&locale=de.

Mahém aos 12 de Fevereiro de 1885

Exc.^{mo} Snr. Hugo Schuchardt.

Com todo prazer & alegria, tenho a honra de receber pella interposiçaõ do Rev.^{do} Snr. W Schmolck, digno pastor de Chombala, a letra de V. Ex.^{cia} de data de 9 de Janeiro ultimo & perfeitamente percebi o théor d'ella; Em resposta sou à dezer, que seja persuadido que conforme o supplico de V. Ex.^{cia} naõ faltarei de fazer tudo o que depende de mim, nesta circumstancia, para comprazer a V. Ex.^{cia}. & com os meus complimontos

Sou de V. Ex.^{cia} muito alto venerador & criado

J. (?) H. D'Cruz

While this letter does not represent MIP, it was accompanied by a set of translated sentences, which, along with other data sent to Hugo Schuchardt in the late 19th century, paint a picture of some variability in the language, including very contrasting idiolects. However, it is important to note that, at this time, the decline of the various Indo-Portuguese creoles was already well under way, though not at the same pace in all locations – as indicated in the following remarks by the French Missionary Louis St. Cyr, also sent to Hugo Schuchardt in 1883[3] (see also Sousa 2014):

> *Quant à la langue Portugaise, elle est, on peut dire, maintenant inconnue dans ces parages du Maduré, et elle ne s'enseigne nulle part. Les quelques familles dont j'ai fait mention plus haut parlent encore dans leurs maisons une sorte de Portugais; mais l'Anglais et le Tamoul, langue du pays, y sont plus communément employés.*
> *Ces remarques s'appliquent uniquement à notre Mission du Maduré, car nous savons qu'en d'autres parties de l'Inde, comme Madras, Cochin, Bombay, Calcutta etc., l'élément Portugais s'est mieux conservé et est plus marqué.*

> [With respect to the Portuguese language, we can say that it is currently unknown in these parts of Madurai, and it is not taught anywhere. The few families I have mentioned above still speak a kind of Portuguese in their homes; but English and Tamil, the local language, are used more often there.
> These observations apply only to our Madurai Mission, because we know that, in other parts of India, such as Madras, Cochin, Bombay, Calcutta etc., the Portuguese element has resisted better and is more evident.] (author's translation)

Cochin, in representation of the Malabar Coast, is mentioned here as one of the places in which the Portuguese language had "resisted better" than in the Madurai Mission. MIP continued to be used in several locations in the Malabar by certain sections of the population that had by then been classified as "Anglo-Indians" but effectively corresponded, in part, to the earlier "Indo-Portuguese"

---

[3] Manuscript preserved in the University of Graz, *Hugo Schuchardt Archiv*, with reference nr. 01-02219. Available online at: https://gams.uni-graz.at/o:hsa.letter.1688/sdef:TEI/get.

communities. The following testimony of one of the last speakers of MIP in Cochin, collected in 1987 and then published by Jackson (1990: 27-28), refers to the 1940s or 1950s and clarifies the importance of the language to the "Porto-Indian" community of Vypeen, off Cochin:

> Here in the parish of "Nossa Senhora de Esperança" the Porto-Indians have two confraternities. One is the confraternity of Infant Jesus and the other is the Confraternity of the Lady of the Rosary. Every year two feasts are conducted from Monday to Sunday, and one of the members of the confraternity was chosen the 'Presidente', or president. [...] In this gathering, it was a mark of culture to speak only Portuguese, because English was considered the language of the "boutique" keeper, the shopkeeper. The funny thing about it is, they always considered the purely English-speaking people as inferior, but in fact, those English-speaking people were financially very well-off and lived in a much better condition than them. But the tradition was such that they considered only the Portuguese-speaking people as true Indo-Europeans.

Despite the evident symbolic value of MIP, the fact is that it was already in decline, and, over the course of the 20th century, it would cease to be transmitted in the various Malabar locations in which it had earlier been vital. According to Cardoso (2019: 357), transmission of MIP appears to have lasted at least until the 1920s in Tangasseri (Quilon) and Kayamkulam, the late 1930s in Vypeen and Cochin, the 1950s in Cannanore, and ceased sometime in the early 20th century in Calicut and Allepey. Theban (1985) mentions that, in 1973, he was able to contact speakers of MIP in Cannanore, Cochin, Vypeen and Tellicherry. My own research, conducted since 2006, allowed me to identify a few speakers, apparently the very last, only in Cannanore, in Vypeen, and Calicut; the interviews made with these speakers make up the corpus of modern MIP that forms the basis of the present study.

3. The Oral Corpus of Malabar Indo-Portuguese

The documentation of MIP that forms the basis of this study began in 2006 and was carried out, over several visits to Kerala, with the assistance of the last fluent speakers of the language (and a few semi-fluent speakers, or rememberers) in Cannanore [Kannur], in Vypeen (opposite the city of Cochin [Kochi]), and in Calicut [Kozhikode]. This resulted in three regional collections which include the audio recording of free-flowing speech (dialogues involving more than one speaker, one-on-one conversations with the researcher, and narratives) and, in some cases, also of elicitation sessions designed to assist the grammatical description of the language. The three collections are somewhat

divergent in terms of their size, the chronology of their collection, and the number of speakers involved:

- The Cannanore collection (Cardoso 2006-2015) involved the participation of 5 fluent speakers (+ 2 semi-fluent speakers) and the collection of 10h:15m:07s of recordings over 3 visits (in 2006, 2010, and 2015);

- The Cochin collection (Cardoso 2007-2010) involved the participation of 1 fluent speaker (+ 1 semi-fluent speaker) and the collection of 3h:36m:51s of recordings over 2 visits (in 2007, and 2010);

- The Calicut collection (Cardoso 2015b) involved the participation of 1 speaker and the collection of just 22m:54s of recordings over 1 visit (in 2015).

Together, they make up the *Oral Corpus of Malabar Indo-Portuguese* [henceforth OCMIP] (Cardoso 2018), All speakers were raised speaking MIP, and narratives such as the one below[4] (retrieved from the corpus; Cardoso 2007-2010: 01.1) clarify the role of the language within their families at the time:

> R: *When you were born, your family spoke Portuguese?*
> I: *Yes, my family, only Portuguese. No other (thing) language.*
> R: *There, in the estate, you said...*
> I: *In the estate, in the estate.*
> R: *Ali nasew, ali jə nase?* [Were you born there?]
> I: *Ali jə nase yo.* [I was born there] *Wayanad* […] *estate dẽtrə.* [Wayanad, in an estate]
> R: *Lõj?* [Is it far?]
> I: *Bõbə lõj aki sə.* [Very far away from here]
> R: *Bos sə famil portuges jẽtə?* [Were your family Portuguese?]
> I: *Papani mamani tudə portəgis. Mamani tudə astãtə poləka saya (tə) lo vota.* [My grandfather, my grandmother, they were all Portuguese. My grandmother and the others used to wear a blouse and skirt.]

However, the corpus also contains abundant descriptions of the break in transmission and gradual decline in fluency within the communities, such as the following dialogue (Cardoso 2006-2015: 01.1), concerning Cannanore at the time of the interviewees' childhood (in the mid-20th c.):

---

[4] In dialogical examples such as this one, <R> indicates the turn of the "Researcher" and <I> that of the "Interviewee". English translations of MIP speech are given in square brackets, immediately after the original.

I$_1$: *muytə jẽti, muytə jẽti tinhə. isti terə ĩter* portuguese*, anglo-indians ĩter* portuguese *namas lo kõbərsa. agə namas jẽti no te. aki dẽt piken krãskrãs tama...* [Many people, there were many people. This land [was] entirely Portuguese, all the Anglo-Indians spoke only Portuguese. Now there is no-one left. Here, the small children also...]
I$_2$: *elot nə tə prẽde.* [They don't learn it.]

As a result of this process, OCMIP offers a glimpse of MIP at the tail end of a process of language loss. At the time of data-collection, most speakers belonged to households in which no other member knew MIP, which meant that the language was no longer used on a daily basis (with the exception of two sisters, who lived together in Cannanore) and, in the case of the most geographically isolated speakers, hardly ever spoken at all. Sadly, since the data was recorded, some of the participants have passed away.

Table 1 provides essential details about the profile of the 7 speakers who participated in conversations in MIP (i.e., excluding those who only provided isolated lexemes or short phrases), whose contributions constitute the basis for the description of MIP; speakers are identified by a letter code, to safeguard their privacy:

Table 1: Personal data for the speakers who contributed substantial amount of data for OCMIP

Code	Gender	Place of residence	Place of birth	Notes
A	F	Cannanore	Cannanore	Siblings
B	F	Cannanore	Cannanore	Siblings
C	F	Cannanore	Cannanore	
D	F	Cannanore	Cannanore	
E	F	Bangalore	Cannanore	Semi-speaker
F	M	Cannanore	Tellicherry	Siblings
G	F	Calicut	Tellicherry	Siblings
H	M	Vypeen (Cochin)	Wayanad	

At the time of data-collection, all speakers were in their 60s and 70s. As can be seen, the group of participants includes 2 pairs of siblings, which is a potentially relevant factor for the analysis of variable data, in the sense that family relations and an expected shared upbringing may be responsible for certain linguistic preferences. But the nature of the corpus, as revealed in Table 1 and the preceding description, also poses a number of additional challenges for the analysis of variation:

a) The low number of speakers makes it difficult to apply statistical methods to disentangle the factors underlying the distribution of the variants;

b) Some speakers are the sole representatives of their respective locations, whether that of residence or that of origin – speaker H is a good example,

as he is the only one to have been born in Wayanad (in the mountainous hinterland of Kerala) and the only one to have resided in and to have been recorded in the Cochin area (more specifically, in Vypeen). Consequently, were there to be a particular association of a linguistic variant with one of these speakers, it would be difficult to tell whether that represents an instance of dialectal, sociolectal, or idiolectal variation;

c) Table 1 also reveals some instances of geographical mobility. In addition to that of speaker H (discussed above), it is also observed in the case of speaker F, and also, quite significantly, of sibling speakers E and F, who resided in different towns (Cannanore and Calicut) but had both relocated from their place of origin (Tellicherry). This factor must also be taken into account when trying to establish whether any instances of variation can be explained geographically;

d) The sociolinguistic circumstances of modern MIP mean that language attrition is a factor to take into account. One of the speakers included in table 1, speaker F, admitted to having a limited proficiency in MIP. Other speakers were certainly more fluent, but they did not all use the language with the same regularity: sibling speakers A and B stood out in this respect, since they lived together and used the language on a daily basis (which resulted in high fluency); and speaker H had not had a chance to use the language for about 3 years when data collection started, ever since the second last speaker in Vypeen, a neighbour, had passed away;

e) One corollary of the previous point is that all speakers represented in the corpus are multilingual. In this particular case, they are all highly fluent in both Malayalam (the dominant and official language of Kerala) and English (that assumes a particularly important role within the "Anglo-Indian" community of which these individuals are seen to be a part of, functioning as the primary language of instruction and as a means of communication in many daily interactions), and perhaps some other languages, such as Kannada in the case of speaker F, who resides habitually in Bangalore. Their multilingualism implies the potential for constant – and variable – transfer of linguistic features, both lexical and structural, from any of these languages.

4. Variability in the corpus

While many linguistic features are essentially consistent across the speech of the various speakers represented in OCMIP, the corpus also contains abundant

instances of variation, from lexical to phonological to syntactic variation. When it comes to lexical and phonological variation, if a particular variant coincides with either Malayalam or English, it could reasonably be interpreted either as a nonce borrowing or a case of code-switching, considering that both these languages are present – and, indeed, dominant – in the speaker's linguistic repertoire; one way to disentangle the issue would be to check for the regularity of these occurrences, but this often becomes difficult because the limited size of the corpus may result in very low numbers of tokens for each variant. The same could be said of morphosyntactic variants, because, as multilingual speakers, there is a real possibility of occasional (as opposed to regular) structural transfer from English or Malayalam onto MIP speech. In this section, however, I will explore some cases of variability which are relatively robustly represented in OCMIP. To that end, I will focus on a few issues within the domain of morphosyntax, to show that, despite the challenges posed by a corpus with the characteristics of OCMIP, some regularities can be found by taking a close look at how variants are distributed in the speech of the various speakers involved.

4.1. Analytic vs. syncretic pronominal forms

The pronominal paradigm of MIP that emerges from an analysis of OCMIP is given in Table 2. Personal pronouns have three series of forms that correspond to three cases: the nominative, the oblique (i.e., that case associated with objects, whether direct or indirect), and the genitive. The genitive series includes some syncretic forms (1SG, 2SG, 1PL, and 2PL) and some analytic forms (3SG, 3PL) that consist of the addition of the regular genitive marker *-sə*[5] to the nominative form; but they are in complementary distribution within the paradigm. When it comes to the oblique series, however, in all persons but 1SG (syncretic) and 3PL (analytic), the oblique set contains competing analytic and syncretic forms; the analytic option attaches to the nominative form the regular oblique marker, which takes the shape *-pə*, *-pərə* or *-kə*.[6]

---

[5] This is the genitive marker that attaches to nominals in general, as shown in the following example:

(i)   *pãcho-sə   kazɐ*
              Pancho-GEN house
              'Pancho's house'

[6] The oblique morpheme that attaches to nominal objects. See also 4.2. for further details.

Table 2: Pronominal paradigm of MIP, as recorded in OCMIP

	Nominative	Oblique	Genitive
1SG	*yo/ye*	*pærmi*	*minhə*
2SG	*bozə/bɔzə*	*pɔrɔs/pɔrɔzə,* *boz-pə(rə)/kə*	*bɔsə*
3SG	*ɛli/eli* [m.], *ælɐ* [f.]	*prɛli/ɛli-pə(rə)/kə* [m.], *prælɐ/ælɐ-pə(rə)/kə* [f.]	*ɛli-sə* [m.], *ælɐ-sə* [f.]
1PL	*nɔzə*	*pærnozə, nɔs-pə(rə)/kə*	*nɔsə*
2PL	*bozə/bɔzə*	*pɔrɔs/pɔrɔzə, boz-pə(rə)/kə*	*bɔsə*
3PL	*olotə/olotrə*	*olot-pə(rə)/kə*	*olot-sə*

The competition between these two different strategies is exemplified in the following two sentences, produced by the same speaker (speaker H, from Cochin), who selects the analytic 3SM oblique form in (1a) and the syncretic 3SF oblique form in (1b):

(1) a. **eli-pə**   *purtɐge*   *"little"#* *pisĩn pisĩn* *lo-sæva*

  3SM-OBL   portuguese   -   little little   IRR-know

  'He would know [just] a little Portuguese.'

  b. *despos minhɐ irmã̃ maysur, nɔs kazɐ ja-pidi kwa* **præla** *ja-da.*

  after   1S.GEN sister Mysore 1P   house PST-buy   PRF 3SF.OBL PST-give

  'Then my sister [went to] Mysore, we bought a house and gave [it to] her.'

On a first analysis, analytic and syncretic pronominal forms are represented in both of the major collections (Cannanore and Cochin) within OCMIP, which might suggest the kind of competition shown in (1) applies across the board. However, when we establish a correspondence between the use of each variant type with each speaker in the corpus, a different picture emerges.

Table 3: Individual speaker's selection of syncretic and analytic pronominal forms

Speaker	Place of residence	Syncretic	Analytic
A	Cannanore		+
B	Cannanore		
C	Cannanore	+	
D	Cannanore		
E	Bangalore		
F	Cannanore		+
G	Calicut		
H	Vypeen (Cochin)	+	+

As Table 3 reveals, all the speakers whose production in OCMIP includes personal pronouns consistently select one or the other, with the exception of

speaker H. The place of residence of the speakers is not particularly revealing, since, as explained above, speaker H is the sole representative of the Cochin region, and those who reside in Cannanore show a preference for either analytic or syncretic forms. What we can conclude, however, is that, contrary to what we might think just from looking at the language's pronominal paradigm, analytic and syncretic pronominal forms are not actually in competition within the repertoire of individual speakers (with a single exception), but tend to be categorical within a single idiolect. Even if it may not be possible to identify, on the basis of the available data, a particular sociolinguistic variable that may account for the speakers' preferences, this observation highlights the importance of observing the distribution of variants by individuals in corpus linguistics, as opposed to a less detailed approach that might conflate the data and project an idea of intraindividual variation where, in fact, what is at stake is largely interindividual variation.

4.2. The form of the oblique marker

In 4.1., we observed, in the case of analytic pronominal forms, the use of an oblique marker. This morpheme attaches not only to pronominal but also to nominal objects in general – but see Cardoso (2021) for a description of MIP differential object marking –, and does so in one of several possible forms: *-pə* (also realised as *-pɐ* or *-pu*), as in (2), produced by speaker H; *-pərə* (also realised as *-pər*), as in (3), produced by speaker F; or *-kə*, as in (4), produced by speaker A:

(2) *ælfõti-**pə**   jə-faze   tirə*

     elephant-OBL PST-make shot

     '[We] shot elephants.'

(3) *Mr. d'Cruz   igreji-**pərə** fanam jə-da*

     Mr. d'Cruz   church-OBL money PST-give

     'Mr. d'Cruz gave money to the church.'

(4) *eli-**kə**   bõbə rayvɐ je-fikɐ   nos   figə   jə-kumɐ suydɐ*

     3SM-OBL much anger PST-become 1P   banana PST-eat   because

     'He was very angry because we ate the bananas.'

Once again, the individual distribution of these variants, as given in Table 4 is enlightening.

Table 4: Individual speaker's selection of the form of the oblique marker

Speaker	Place of residence	*-pə*	*-pərə*	*-kə*
A	Cannanore			+
B	Cannanore			
C	Cannanore	+		
D	Cannanore	?		?
E	Bangalore			
F	Cannanore		+	
G	Calicut		+	
H	Vypeen (Cochin)	+		

The first observation is that the *-pərə* form is associated with only speakers F and G, who are siblings and were born in Tellicherry (see also Cardoso 2021: 312). As such, this appears to be a variant which is to be associated with either a given family context or, alternatively, with a particular geographical provenance – this being one of those circumstances in which the limited size of the speaker population in the corpus prevents a more concrete reading of the factors underlying variation. In addition, it is interesting that, even though speakers F and G have lived apart and away from their place of origin for a large part of their lives, the form of their oblique marker has remained intact, and has not been replaced with or entered in competition with another variant.

When it comes to the other two forms, it is important to clarify, first of all, that the question marks for speaker D in Table 4 result from the fact that the only (very few) occurrences of an oblique marker in her speech are in elicited sentences and reveal a certain hesitation between *-pə* and *-kə*; I do not find the data sufficiently clear and robust to make a call in the case of speaker D, but, at least, the *-pərə* option appears not to be a part of her idiolect. Speakers C and H use *-pə* categorically, and only speaker A uses *-kə* in the corpus (see also Cardoso 2021: 311), also categorically (as does her sister, speaker B, although this has not been captured in the recordings that integrate OCMIP).

The etymology of these forms is relevant here. Both *-pə* and *-pərə* derive from Ptg. *para* 'for, to' (or perhaps *por* 'by'), and must be old forms in the language, since the Portuguese language lost much of its influence on the Malabar coast after the Dutch take-over of the Portuguese strongholds, in the mid-17th c. The morpheme *-kə*, on the other hand, is most likely modelled after the Malayalam dative marker *-kkə*, (for a more detailed exploration of this

hypothesis, see Cardoso 2021).[7] Considering that Malayalam, unlike Portuguese, has remained a potential source of influence throughout the history of MIP, this transfer of a Malayalam functional morpheme could have happened at any time. It might have been a recent introduction, which is not entirely unlikely, given that the few early records of MIP available (those collected by and for Schuchardt in the late 19th c.) do not record an oblique marker similar to *-kə*; but we should beware of jumping to conclusions, since such data are also relatively limited and may not have captured the full extent of the variation in the language at the time.

Either way, we might speculate that the use of a Malayalam morpheme could be a consequence of the attrition of MIP. In that case, however, we should expect this form to occur in the production of speakers in a more advanced state of attrition (which might account for the hesitation in the production of speaker D, whose proficiency is indeed limited), but that is not necessarily the case, since speaker A is one of those who still use the MIP on a daily basis and, as a consequence, she has a particularly fluent command of the language. In addition, in speaker A's idiolect, the use of *-kə* is not simply a wholesale substitution of *-pə*, but implies a reinterpretation of *-pə*. In fact, speaker A still uses a case-marker *-pɐ*, but, in her case, only with addressee arguments of the verbs *rɛza* 'to pray' and *fala* 'to say', as in:

(5) *kɔrsɐ-dæ̃ntrə tæ̃   tudu dews-pɐ   po-fala, yo agɔrɐ   bɔsɐ   pærtə*
     heart-LOC   EXS   all   god-SOC   can-say 1S now   2S.GEN SOC

     kilay tə-kɔ̃mbərsa, akɐ   mɔdɐ nos dret dret   dews-sə   pærtə lɔ-kɔ̃mbərsa.
     how   PRS-talk   DEM way   2P   right right god-GEN SOC   IRR-talk

     '[We] can say to God everything which is in the heart, the way I'm talking with you now, the same way we'll talk directly with God.'

In MIP, addressee arguments are typically marked with *-(sə) pærtə*, which is also a locative marker (see Cardoso 2014b). So, in the idiolect of speaker A (but not that of the other speakers, as far as the OCMIP allows us to observe), the

---

[7] Clements (2009: 60-63) also highlights Malayalam *-kkə* and its functional range as potentially relevant to understand the case-marker paradigm of the Asian-Portuguese creoles. In MIP, an alternative etymon for oblique *-kə* would be the Portuguese preposition *com* 'with', which did result in object markers in other Asian-Portuguese creoles (including those of Malacca and Macau); however, as explained in Cardoso (2021: 312), this is a less likely chronological explanation, not only because the distribution of the MIP oblique case owes much to that of the Malayalam dative, but also because Ptg. *com* resulted in a different MIP form (the coordinator *ku*).

absence of an oblique *-pə* is accompanied by its reassignment to the domain of the valency of *verba dicendi*.

4.3. Position of subordinator *kã̄də*

The syntactic profile of MIP is very much in tune with what we might expect of a strongly head-final language (like Malayalam), as described in Krajinović (2015, 2018) and Cardoso (2021). In addition to features such as a basic SOV word order, prenominal relative clauses, or the use of postpositions, this also manifests itself in the prototypical placement of subordinates at the right edge of the subordinate clause. This rule has only a few exceptions in OCMIP, but one particular temporal subordinator, *kã̄də* 'when' (frequent in clause chaining) is almost equally well-represented in clause-final (i.e., postverbal) position as in preverbal position (see also Cardoso 2021: 306-307): out of 34 instances of *kã̄də*, 16 are preverbal – as in (6), produced by speaker H –, while 18 occur at the end of the clause – as in (7), produced by speaker C:

(6)   *ali   **kã̄də**   fika nɔs-pɐ tudə   tinhɐ acha.*

        there TEMP stay 1P-OBL all   PST   obtain

        'When we lived there, we would get everything.'

(7)   *oʎa-**kã̄də**,   tə-gostɐ sɐnɐ,   da   kaza.*

        look-TEMP PRS-like if   give marry

        'After seeing [her], if [they] like [her], [we] give [her] away in marriage.'

These figures might suggest an almost equal prevalence of both placements of *kã̄də*, but the distribution of each occurrence by speaker, given in Table 5, paints a different picture.

Table 5: Individual speaker's placement of the temporal subordinator *kã̄də*

Speaker	Place of residence	Preverbal	Postverbal
A	Cannanore		+
B	Cannanore		
C	Cannanore	(+)	+
D	Cannanore		+
E	Bangalore		
F	Cannanore		+
G	Calicut		+
H	Vypeen (Cochin)	+	

As shown in Table 5, the preverbal placement of *kãdə* is to be found only in the production of speakers C and H – and, in the case of C, preverbal *kãdə* accounts for only one instance, as opposed to several of postverbal *kãdə* (hence the brackets in the table). For every other speaker in the corpus, this subordinator is strictly postverbal, in line with the position of other subordinators in general.

In the case of speaker H, the preverbal placement of *kãdə* is not only contrary to the linguistic behaviour of all other MIP speakers in the corpus, but it is also categorical; once again, whether this constitutes a particular idiolectal feature or reveals the significance of any of the factors associated with his life and upbringing is impossible to say, given the scarcity of speakers with similar sociolinguistic characteristics. With this in mind, it becomes clear that the numerical overrepresentation of preverbal *kãdə* is justified by the fact that speaker H, the sole responsible for the Cochin corpus, contributed a very substantial amount of the data in OCMIP.

4.4. The past form of verb *tæ*

In MIP, the verb *tæ* (from Ptg. *tem* 'has') is highly syncretic and multifunctional (see Krajinović 2015, 2018). As an independent (main) verb, it functions as a copula, as an existential verb, as a possessive verb, and as a locative verb; when used after the main verb of a clause, it can be interpreted as a light (explicator/vector) verb that constructs the meaning of a perfect; but it can also occur preverbally, as an aspectual marker of imperfectivity (used e.g. for the habitual or progressive).

In its preverbal position, *tæ* often occurs with a reduced form *tə* or *tɐ*, as seen in the first clause of (8a). In this section, however, I will focus not on such non-past forms of *tæ*, but in their past counterparts, since considerable variation is to be found in their shape. In the corpus, these verbal forms occur as a) *tinhɐ* or *tinhi*, the closest to its Portuguese etymon *tinha* 'had', as in (8a) and (8b), produced by speaker A; b) as *tinhærɐ*, as in (9), produced by speaker F; or c) in a reduced form *tin* or *ti* in preverbal position, as in (10), produced by speaker A:

(8) a. *akɐ mulhærə   fana   tə-furtæ,   ælɐ minhɐ kambrad   tinhɐ.*

  DEM woman   money IPFV-steal 3SF 1S.GEN   friend   be.PST

  'That women who stole the money, she was my friend.'

b. *bõbɐ*   *pikæn*   **tinhi**.

   very   small   be.PST

   "It was very small."

(9)   *boz*   *õti*   *Kochi-dæ̃trə*   **tinhærɐ**.

   2S   yesterday   Cochin-LOC   be.PST

   'You were in Cochin yesterday.'

(10)   *akɔ̃ndə*   *kɔnkani*   *jẽti*   **tin**-*fika*.

   there   konkani   people   IPFV-live

   'Konkani people used to live there.'

Upon closer analysis, the distribution of these three formal variants is not random. Table 6 shows which of the three past forms of *tæ* can be identified in the speech of each speaker.

Table 6: Individual speaker's use of past forms of *tæ*

Speaker	Place of residence	*tin/ti*	*tinhɐ/tinhi*	*tinhærɐ*
A	Cannanore	+	+	
B	Cannanore			
C	Cannanore	+	+	+
D	Cannanore	+	+	
E	Bangalore			
F	Cannanore		+	+
G	Calicut		+	+
H	Vypeen (Cochin)	+	+	

Only speaker C makes use of all three variants in her speech; the rest of the speakers select two of them, although not all of them select the same two. The pattern in this respect does not appear to be geographical since speakers from Cannanore (A and D) and from Cochin (H) show a similar set of forms (*tin/ti* + *tinhɐ/tinhi*); but, once again (see 4.2. above), we can recognise a similar behaviour in the pair of siblings from Tellicherry, speakers F and G (*tinhɐ/tinhi* + *tinhærɐ*).

If, in addition to this, we also verify the syntactic position of each of these competing forms for each of the speakers, we notice another interesting regularity: the longer form is always that associated with the main or postverbal uses of the verb, and the shorter form is the one reserved for the preverbal position. In other words, the speakers share a rule despite the fact that the actual

forms employed are not the same: for some speakers, *tinhærɐ* is the appropriate form for main and postverbal uses, while *tinhɐ*/*tinhi* is the form used in preverbal position; for other speakers, *tinhɐ*/*tinhi* is the appropriate form for main and postverbal uses, while *tin*/*ti* is the form used in preverbal position. Speaker C does not follow exactly the same pattern, in the sense that she uses all three forms, but, even in this case, *tinhærɐ* does not occur in preverbal position, and *tin*/*ti* does not occur in main and postverbal uses.

5. Final remarks

Whenever the only corpus available to conduct the description of a language is limited in size and/or number of speakers, as is often the case if the language in question is in an advanced state of decline, understanding and explaining variation can be especially challenging. In the case of OCMIP, the high level of multilingualism, the geographical dispersion and mobility of the speakers – not to mention the fact that the language was entirely undescribed before the corpus began to be collected – only added to the challenges. Nevertheless, and even though we cannot avail ourselves of statistical tests to clarify certain issues, we have seen that an observation of the distribution of variants across the various speakers who contributed to the corpus does allow us to observe some regularities on which to build hypotheses concerning the roots of the observed variation.

We have observed four different cases of variability: a) the option for syncretic or analytic pronominal forms; b) the form of the oblique marker; c) the syntactic placement of the temporal subordinator *kãdə*; and d) the past form of the multifunctional verb *tæ*. In our discussion, the potential effect of different individual degrees of language attrition does not emerge as significant. One reason for this may be that the least fluent speakers actually contributed less material for the corpus, and are therefore less likely to be considered in analyses that require substantial data (see e.g. the case of speaker E). But, even in the case of the use of the oblique marker *-kə*, derived from Malayalam and, so, potentially associated with language attrition in MIP, the hypothesis does not hold, since the speakers who employ this form are not among the least fluent in MIP.

In most cases studied here, a clear geographical patterning does not occur, since variability is observed within a particular location (cf. the behaviour of the Cannanore residents with respect to the selection of personal pronouns or the form of the oblique marker), and, in addition, certain variants extend across

geographical boundaries (cf. the individual selection of past forms of *tæ*). The distribution of some variants is not incompatible with a dialectal reading, but this cannot be shown with any degree of finality, either because a particular variant is associated with a speaker who is the sole representative of a given location (cf. the behaviour of speaker H with respect to the syntactic position of the subordinator *kãdə*) or because there is a competing variable that could equally well underlie a given commonality. The similarities observed between speakers F and G exemplify the latter case: since they are siblings and also the only two speakers hailing from Tellicherry, their shared linguistic behaviour could either result from a family-internal specificity or from the retention of a trait previously widespread in their hometown. In either case, the observation of their similarities – even after years living apart – constitutes a relevant regularity in these particular instances of variation.

Our study of variability in OCMIP is an exercise in variation which at once draws attention to the prevalence of variation in creole languages (not only across relatively vast areas, but also in closely-knit communities), highlights the specific challenges of conducting linguistic research among the last speakers of a particular language, and proposes that, in similar circumstances, observing the distribution of variants among individuals allows some progress towards disentangling the roots of variation and avoiding misrepresenting a given type of interindividual variation as intraindividual.

References:

Anquetil-Duperron, Abraham Hyacinthe. 1786. Des recherches historiques et géographiques sur l'Inde. In Jean Bernoulli (ed.), *Description historique et géographique de l'Inde*, vol. II. Berlin: Imprimerie de Pierre Bourdeaux.

Baxter, Alan N. 1996. Portuguese and Creole Portuguese in the Pacific and Western Pacific Rim. In Stephen A. Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds.), *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, vol. II.1, 299–338. Berlin/New York: Mouton de Gruyter.

Baxter, Alan N. & Augusta Bastos. 2012. A closer look at the post-nominal genitive in Asian Creole Portuguese. In Hugo C. Cardoso, Alan N. Baxter & Mário Pinharanda Nunes (eds.), *Ibero-Asian creoles: Comparative perspectives*, 47–79. Amsterdam/Philadelphia: John Benjamins.

Campbell, Lyle & Martha C. Muntzel. 1989. The structural consequence of language death. In Nancy Dorian (ed.), *Investigating Obsolescence: Studies in Language Contraction and Death*, 181–196. Cambridge: Cambridge University Press.

Cardoso, Hugo C. 2006-2015. *Corpus of Cannanore Indo-Portuguese speech*. Unpublished, Centro de Linguística da Universidade de Lisboa.

Cardoso, Hugo C. 2007-2010. *Corpus of Vypeen [Cochin] Indo-Portuguese speech*. Unpublished, Centro de Linguística da Universidade de Lisboa.

Cardoso, Hugo C. 2014a. Factoring sociolinguistic variation into the history of Indo-Portuguese. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 5. 87–114.

Cardoso, Hugo C. 2014b. The case of addressees in Dravido-Portuguese. *PAPIA* 24(2). 307–342.

Cardoso, Hugo C. 2015a. The correspondence between D'Cruz and Hugo Schuchardt. In Bernhard Hurch (ed.), *Hugo Schuchardt Archiv* (2007-). Web edition available at: https://gams.uni-graz.at/o:hsa.person.1342/sdef:TEI/get?locale=de.

Cardoso, Hugo C. 2015b. *Corpus of Calicut Indo-Portuguese speech*. Unpublished, Centro de Linguística da Universidade de Lisboa.

Cardoso, Hugo C. 2016. O português em contacto na Ásia e no Pacífico. In Ana Maria Martins & Ernestina Carrilho (eds.), *Manual de Linguística Portuguesa*, 68–97. Berlin: Mouton de Gruyter.

Cardoso, Hugo C. 2018. *Oral Corpus of Malabar Indo-Portuguese*, Endangered Languages Archive. http://hdl.handle.net/2196/08153dce-e214-479c-9671-abcaf4f4cad1.

Cardoso, Hugo C. 2019. The Indo-Portuguese creoles of the Malabar: historical cues and questions. In Pius Malekandathil, Lotika Varadarajan & Amar Farooqui (eds.), *India, the Portuguese, and Maritime Interactions*, vol. II [*Religion, Language and Cultural Expression*], 345–373. Delhi: Primus Books.

Cardoso, Hugo C. 2020. Contact and Portuguese-lexified creoles. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 2nd ed. [*Blackwell Handbooks in Linguistics*], 469–488. Hoboken, NJ: Wiley-Blackwell.

Cardoso, Hugo C. 2021. Convergence in the Malabar: The case of Indo-Portuguese. *Journal of Pidgin and Creole Languages* 36(2). 299-336.

Cardoso, Hugo C., Tjerk Hagemeijer & Nélia Alexandre. 2015. Crioulos de base lexical portuguesa. In Maria Iliescu & Eugeen Roegiest (eds.), *Manuel des anthologies, corpus et textes romans*, 670–692. Berlin: Mouton de Gruyter.

Clements, J. Clancy. 2000. Evidência para a existência de um pidgin português asiático. In Ernesto d'Andrade, Dulce Pereira & Maria Antónia Mota (eds.), *Crioulos de base lexical portuguesa*, 185–200. Braga: Associação Portuguesa de Linguística.

Clements, J. Clancy. 2009. *The linguistic legacy of Spanish and Portuguese; Colonial expansion and language change*. Cambridge: Cambridge University Press.

Dal, Nikolaus. 1733. Nachricht von den Portugiesen in Indien. In Gotthilf August Francken (publ.) *Drey und dreyßigste CONTINUATION Des Berichts Der Königlichen Dänischen Mißionarien in Ost-Indien / Worin eine Ausfuehrliche Nachricht Von der gantzen Verfassung der Mißion, Wie auch Die Fortsetzung des Tage-Registers vom Jahr 1731. und einige Briefe der Herren Mißionarien enthalten, Nebst einem Anhang Von der Evangelischen Mission zu Madras*, 917–924. Halle.

Evans, Nicholas. 2001. The last speaker is dead – long live the last speaker! In Paul Newman & Martha Ratcliff (eds.), *Linguistic Fieldwork*, 250–281. Cambridge: Cambridge University Press.

Ferraz, Luiz Ivens. 1987. Portuguese creoles of West Africa and Asia. In Glenn Gilbert (ed.), *Pidgins and Creole Languages: Essays in Memory of John E. Reinecke*, 337–360. Honolulu: University of Hawaii Press.

Jackson, K. David. 1990. *Sing Without Shame. Oral Traditions in Indo-Portuguese Creole Verse*. Amsterdam: John Benjamins.

Krajinović, Ana. 2015. *O sistema verbal dos crioulos indo-portugueses do Malabar*. Lisbon: Universidade de Lisboa MA dissertation.

Krajinović, Ana. 2018. *Malabar Indo-Portuguese: Structure and origins of the verbal system*. Munich: LINCOM.

Mansfield, John & James Stanford. 2017. Documenting sociolinguistic variation in lesser-studied indigenous communities: Challenges and practical solutions. In Kristine A. Hildebrandt, Carmen Jany & Wilson Silva (eds.), *Documenting Variation in Endangered Languages* (Language Documentation & Conservation Special Publication 13), 116–136. Honolulu: University of Hawai'i Press.

Meyerhoff, Miriam. 2017. Writing a linguistic symphony: Analyzing variation while doing documentation. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 62(4). 525–549.

Nagy, Naomi. 2009. The challenges of less commonly studied languages: Writing a sociogrammar of Faetar. In James Stanford & Dennis Preston (eds.), *Variation in Indigenous Minority Languages*, 397–417. Amsterdam/Philadelphia: John Benjamins.

Smith, Norval. 1995. An annotated list of creoles, pidgins, and mixed languages. In J. Arends, P. Muysken & N. Smith (eds.), P*idgins and Creoles; An Introduction*, 331–374. Amsterdam/Philadelphia: John Benjamins.

Sousa, Silvio Moreira de. 2014. Die Korrespondenz zwischen Louis Saint-Cyr und Hugo Schuchardt. In Bernhard Hurch (ed.), *Hugo Schuchardt Archiv* (2007-). Web edition available at: https://gams.uni-graz.at/o:hsa.person.1350/sdef:TEI/get?locale=de.

Theban, Laurentiu (1985) Situação e perspectivas do português e dos crioulos de origem portuguesa na Índia e no Sri-Lanka. In *Congresso sobre a situação actual da língua portuguesa no mundo, 1983*, vol. I, 269–285. Lisbon: Instituto de Cultura e Língua Portuguesa.