

**** NOTE:** This is a pre-final version. The published version of this paper can be found in:

- Cardoso, Hugo C. & Patrícia Costa. 2021. ‘Synchronic variation in Sri Lanka Portuguese personal pronouns’. *Journal of Pidgin and Creole Languages* 36(1): 81-113.
<<https://www.jbe-platform.com/content/journals/10.1075/jpcl.00070.car>>

Synchronic variation in Sri Lanka Portuguese personal pronouns¹

Hugo C. Cardoso and Patrícia Costa

Universidade de Lisboa – Faculdade de Letras, Portugal

Abstract

This paper presents and discusses the instances of synchronic variation attested in the personal pronoun paradigm of modern Sri Lanka Portuguese, an endangered Portuguese-based creole spoken by relatively small communities scattered across Eastern and Northern Sri Lanka. Although Sri Lanka Portuguese has a long history of documentation dating from, at least, the beginning of the 19th century, only a few studies have explicitly reported cases of synchronic variation. This study aims, therefore, to fill that gap, by contributing to the description and explanation of patterns of variation relating to the personal pronoun paradigm as encountered in documentary data collected between 2015 and 2020, over several field trips to the districts of Ampara, Batticaloa, Jaffna, and Trincomalee. The nature of the variation observed in the data ranges from phonetic alternations to strategies of paradigm regularization and stylistic shrinkage, often revealing the effects of diachronic processes of variant competition and substitution. Combining the observed patterns of variation with surveyed linguistic trends of language shift, we propose that obsolescence may be responsible for some of the variability encountered in modern SLP personal pronouns, especially that associated with certain socially- or geographically-defined subsets of the speech community (viz. the younger generations and the speakers from Jaffna) characterized by advanced language loss.

¹ This study is the result of research activities supported by the *Endangered Languages Documentation Programme* (grant: MDP0357), the *Fundação para a Ciência e a Tecnologia* (grant: IF/01009/2012), and the University of Lisbon (through a doctoral scholarship attributed to Patrícia Costa in 2018). The authors would like to thank the many Sri Lanka Portuguese speakers who contributed generously and patiently to our documentation of the language, as well as Donald Winford and the editors of this special issue, Isabelle Léglise, Bettina Migge, and Nicolas Quint. We are also thankful to the participants of the 2019 joint ACBLPE and SPCL meeting in Lisbon and to three anonymous reviewers whose comments helped us improve our study significantly. Finally, we wish to acknowledge Mahesh Radhakrishnan and Rui Pereira, fellow members of the *Documentation of Sri Lanka Portuguese* project team, who collected and made available to us some of the relevant data for this study.

Keywords: Sri Lanka Portuguese, personal pronouns, variation, endangered languages, language obsolescence.

1. Introduction

While sociolinguistic considerations, by necessity, are never far from the study of creole languages – be it because of questions related to their formation and development, endangerment, coexistence with other languages, or others –, studies of community-internal variation among these languages are not equally distributed. As Sippola (2018:96) pointed out, sociolinguistic studies of variation have been less common for Romance-based than for English-based creoles, and those that do exist have tended to focus on some of the creoles with the largest numbers of speakers. The Portuguese-lexified creoles of South Asia, currently spoken by relatively small communities, are a group of languages which have been somewhat neglected in this respect, even if descriptive works occasionally report instances of synchronic variation and a few studies have explicitly tackled such cases (e.g. Clements 1990; Cardoso 2007; see Cardoso 2014 for an overview and discussion).

This article contributes towards that descriptive effort by exploring the various instances of variation concerning personal pronouns in modern Sri Lanka Portuguese (henceforth, SLP), the Portuguese-lexified creole which formed in the 16th century on the island then known as *Ceylon* and, though endangered, continues to be spoken in a few discrete pockets in the Eastern and Northern parts of Sri Lanka. We base this study on a recently-collected corpus of SLP (Cardoso 2017), part of an ongoing documentation effort (see Cardoso et al. 2019), which records considerable variation in various domains of modern SLP, well beyond pronouns.

This study of personal pronouns is a step towards systematizing the variation encountered in these documentary materials and ensuring that the description of SLP takes stock of it. Several authors have highlighted not just the mutual advantages but also the challenges of combining language documentation and sociolinguistic analysis (Meyerhoff 2017), and of making sure grammatical descriptions do not gloss over variation (Nagy 2009). Potential issues arise from traditional differences in the scope and methodologies of variationist research, on the one hand, and of language documentation and language description, on the other. These challenges are amplified in the case of under-documented or under-described languages – as opposed to languages with a considerable history of research, for which large datasets may already be available and in which patterns of variation may already have been identified –, since the identification and interpretation of patterns of variation often has to proceed in tandem with the

delineation of the language's basic grammatical properties (Meyerhoff 2017); and especially so if the language is endangered, in which case there may be the additional challenge of a small pool of speakers from whom to obtain linguistic data (Nagy 2017). Mansfield & Stanford (2017:119) point out that the practical difficulties faced by outsider researchers interested in making sense of how a language they are documenting interacts with sociolinguistic variables derive from three factors: a) the fact that they are "cultural outsiders" (which poses challenges to the very collection of reliable and representative data); b) the fact that they must identify "variables with limited prior knowledge" (since, when dealing with under-described languages, the incremental nature of knowledge-building is especially evident); and c) the fact that they must conduct "sociolinguistic analysis with limited data". We are aware that all of these apply in the case of our study – although, as we explain in section 2, our analysis of SLP personal pronouns could rely on prior descriptions –, which justifies the cautious nature of some of our conclusions, but still allows progress in making sense of modern SLP variation.

Our first objective in this study, then, is to report on the full range of variants within the SLP personal pronoun domain that can be identified in the corpus, in terms of their form and function, much of which is absent from the previous literature. We will furthermore attempt to identify any patterns in the distribution of variables which may provide insights into the language's social and geographical diversity, but also into its diachrony and the impact of Tamil, the dominant language of the region. Even though, as a creole, language contact between Portuguese and the languages of Sri Lanka (primarily Tamil, in the case of the varieties studied here, but also potentially Sinhala, in earlier stages) is part of the formative history of SLP, the contact dimension is especially relevant to our synchronic analysis because, as explained in section 2, SLP is a minority non-official language and, as such, SLP speakers are at least bilingual.

The SLP speech community's multilingualism is not alien to another fact which informs our analysis of variation, viz. the language's endangerment. Situations of language obsolescence such as that observed in SLP are said to favor particular types of structural change of the non-dominant languages, not because they are exclusive of obsolescing languages, but because they are intensified or sped up in those contexts. These include (to name but a few): overgeneralization of specific (marked or unmarked) features, loss of phonological contrasts, morphological reduction, preference for analytic over synthetic constructions, paradigm levelling, loss of allomorphy, loss of certain grammatical categories, and stylistic shrinkage (see Campbell & Muntzel 1989; Sasse 2001; Palosaari & Campbell 2011; Aikhenvald 2020). In addition, it is also hypothesized that obsolescence can make a receding language particularly permeable to the structural influence of the community's dominant language. Obsolescence may produce variability at the individual level, through the flexibilization of earlier categorical rules (Sasse 2001:1671), or at the communal level, if its effects do not apply homogeneously across

the entire speech community. Given that the degree of SLP language loss is not socially or geographically uniform, we will discuss whether (some of) the patterns of variation identified may relate to language obsolescence – approached here in its dimension of gradual communal (as opposed to individual) language loss.

The paper is organized as follows: section 2 provides necessary background information concerning the object of research, the sources and methods of data collection, and provides a brief sketch of past (19th-century) stages of SLP, collated from early descriptive sources; section 3 describes the modern SLP personal pronoun paradigm, contrasting previous descriptions with the scenario that emerges from an exploration of the recently-collected corpus; section 4 discusses the observed instances of variation, abstracting possible generalizations concerning the distribution of particular variants; section 5 discusses, on the basis of the previous observations, the relevance of language obsolescence to interpret certain observed patterns of variation; finally, in section 6, we make some final remarks concerning the implications of this study for the description of modern SLP and the practice of language documentation.

2. Background

The development of Sri Lanka Portuguese relates to Portuguese colonial presence on the island, which lasted from 1505 to 1658. Over time, this creole acquired a relevant position within that territory and across different communities, with an importance which justified a flurry of 19th-century publications in or about this language, and Dalgado's (1900:xxii-xxiii) assertion that it was particularly prominent "entre os dialectos portugueses coloniaes" ('among the Portuguese colonial dialects'). However, the language subsequently contracted, to the extent that it is now exclusively associated with one subsection of Sri Lanka's population: the *Portuguese Burghers*, i.e. Sri Lankans who claim an Asian-Portuguese ancestry (see McGilvray 1982). By all accounts, SLP is currently endangered (Nordhoff 2013; Pereira 2019; Cardoso et al. 2019).

At present, pockets of SLP speakers² (see Fig. 1) are located in the Eastern Province, in the cities of Trincomalee and Batticaloa (which host the largest concentrations of speakers) and in nearby locations (e.g. Eravur, Valachchenai, Kalmunai, and around Akkaraipattu) in the districts of Trincomalee and Batticaloa, but also of Ampara. In addition to this, there is a small number of speakers in the Northern Province, in the city of Jaffna.

It is important to mention that, while the communities of the Eastern Province are in close contact with each other – with family relations extending across the region, considerable

² Excluding those that resulted from recent migrations to other parts of Sri Lanka, especially the city of Colombo.

mobility and opportunities for gatherings (such as weddings or community events), the Jaffna community in the Northern Province has long been isolated from these dynamics, so much so that the recent identification of speakers in that city came as a surprise to the community in the Eastern Province. In addition, the process of language loss is especially advanced in Jaffna, in which very few people – likely between 10 and 15 – speak SLP, the youngest one (identified so far) being 62 years old.

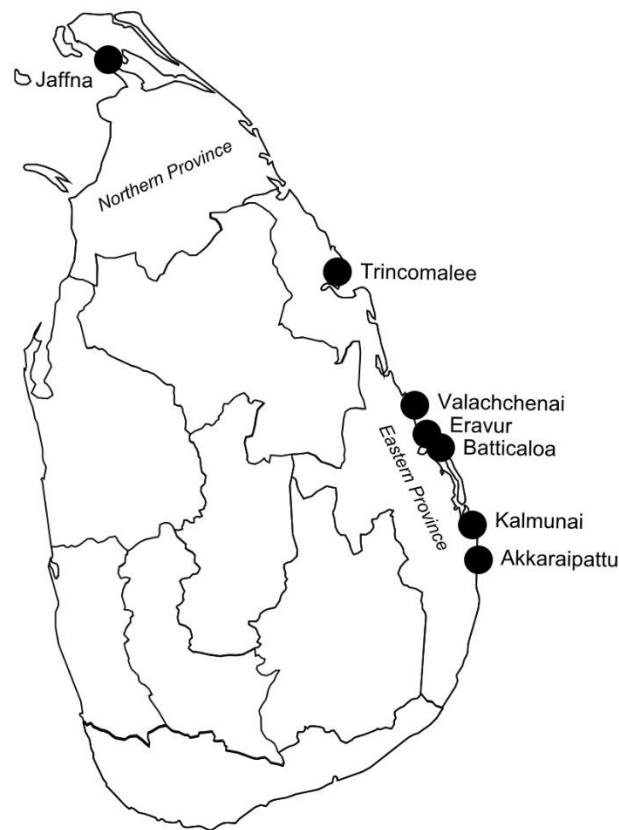


Figure 1: Distribution of the SLP-speaking community in modern Sri Lanka.

When it comes to the Eastern Province, while the overall numbers are much larger, language shift is also a real issue, and affects some regions and sections of the population more than others. A recent sociolinguistic survey conducted among the Portuguese Burgher community of the Eastern Province³ (between October 2017 and May 2018), aimed at ascertaining fluency in SLP and language use among the community, gathered information from 3094 respondents and provides some information in this respect. Overall, 25.53% of respondents declared complete fluency or little difficulty in using/understanding SLP, with an additional 15.68% declaring some difficulty in speaking/understanding it. However, these values vary substantially across districts, with a more robust knowledge of the language in Trincomalee (38.37% and 14.92%,

³ Jaffna could not be surveyed at the time, because the identification of SLP speakers in that city only occurred towards the end of the research project.

respectively), a lower percentage of fluency in Batticaloa (25.56% and 19%, respectively) and much smaller numbers in Ampara (15.60% and 8.07%, respectively) (see Cardoso et al. 2019). Another important conclusion of the survey is that fluency in SLP is unevenly distributed across age groups (see Pereira 2019), with a low percentage of respondents in age groups below 40 declaring any knowledge of SLP at all (with values by decade and across the 3 districts varying between 0% and 34.7%), a somewhat higher percentage in age groups between 40 and 60 (with values varying between 22% and 45.3%) and a much higher percentage among respondents over 60 (with values varying between 43% and 100%). This age distribution of fluency reveals the effects of a progressive break in the transmission of SLP, which can be said to be undergoing a process of “gradual death” (Campbell & Muntzel 1989:184-185), with the younger generations showing an ever more evident shift to a different language – in this case, as also shown by the results of the survey (Pereira 2019), mostly to Tamil.

Contrary to what happens with many other creoles, a significant number of written sources for SLP have been produced, especially in the late 19th and early 20th century (see Tomás 1992; Cardoso, Hagemeijer & Alexandre 2015; Smith 2016). Afterwards, following a period in which the language received little attention (through most of the 20th century), the documentation and description of modern SLP resumed in the 1970s, pushed forward in particular by Ian Smith’s research (e.g. Smith 1977, 1979a, 1979b, 2013). This work brought the language to the forefront of various debates within contact linguistics, especially with reference to the mechanisms and speed of convergence (see e.g. Smith 1979a; Bakker 2006) and resulted in a significant amount of grammatical description.

Throughout this long history of SLP documentation, the various sources do record linguistic diversity, but explicit explorations of variation are rare. Arguably, one reason why community-internal variation has not become more prominent in SLP research is that, until recently, the available spoken data (most prominently that included in Smith’s 1973 corpus) had been produced by relatively few speakers, mostly concentrated in the city of Batticaloa.

However, the recent *Documentation of Sri Lanka Portuguese* project (henceforth DSLP; see Cardoso 2017; Cardoso et al. 2019), hosted by the *Centro de Linguística da Universidade de Lisboa* and funded by the *Endangered Languages Documentation Programme*, has collected a more varied corpus of SLP speech, with language samples from *circa* 150 speakers collected in nearly 50 different locations (towns or town areas) scattered across 3 districts of Eastern Sri Lanka (Ampara, Batticaloa, and Trincomalee) and 1 of Northern Sri Lanka (Jaffna). The DSLP corpus substantially expands the diversity of sociolinguistic profiles for which spoken data are available, as well as the geographical reach of the documentation. In fact, the inclusion of Jaffna in the documentation is a complete innovation. The northern city was originally not

contemplated in the documentation project because the received wisdom was that, as in many other former SLP-speaking strongholds within Sri Lanka, there were no speakers of SLP left in Jaffna. However, a prospective visit to Jaffna in 2018 allowed the research team to locate a few speakers and make the first few recordings of the northern variety. Recently, in 2020, further fieldwork was conducted in Jaffna, resulting in an additional 18 hours of recorded speech (involving narratives, conversation, and elicitation sessions), as a result of which this study is able to make the very first contribution to the description of the Jaffna variety of SLP and to the exploration of the extent to which it coincides with the eastern varieties. Unsurprisingly, given its breadth, the DSLP corpus reveals instances of variation which had not been recorded before, including in the forms and uses of personal pronouns – the focus of the present study.

2.1. Personal pronouns, diachrony and language contact

Cross-linguistically, personal pronouns constitute a universal or quasi-universal (Bhat 2004:30-31; Siewierska 2004:13) heterogeneous category comprising a small and closed set of independent or bound forms whose primary function is to assign the deictic category of grammatical person to the entities involved in a speech act, viz. the speaker, the addressee, and other entities in the linguistic context. When dealing with personal pronouns, there is however a wide range of linguistic issues to observe, since cross-linguistic data (see e.g. Helmbrecht 2013; Siewierska 2013) reveals that, in addition to person (1st, 2nd, 3rd), they can also convey other distinctions, such as number (singular, plural, dual, trial, etc.), gender (masculine, feminine, neutral) and, in some cases, honorificity (degrees of politeness or formality) or clusivity (inclusive or exclusive integration of the addressee). Like nouns, personal pronouns may also convey case distinctions. Additionally, the pragmatic manipulation of a given distinction to convey a non-original purpose (e.g. the well-known use of a second person plural pronoun for a singular honorific referent) is also attested.

Personal pronoun paradigms are usually seen as three-person sets of forms, but theoretical and typological studies (e.g. Benveniste 1971; Lyons 1977; Bhat 2004) based on an analysis of the speech acts and the respective roles performed by personal pronouns have pointed out the distinction between 1st and 2nd person pronouns, on the one hand, and 3rd person pronouns, on the other. According to those studies, while 3rd person pronouns are used for denoting anaphoric referents whose interpretation is dependent on the linguistic context, 1st and 2nd person pronouns are deictic forms which denote the two principal participant/speech roles of the clause in which they occur, viz. the speaker and the addressee. This approach suggests that only 1st and 2nd person pronouns are personal pronouns *per se*, assigning 3rd person pronouns to a different category – Bhat (2004:132),

for instance, following the analysis of a 255-language sample, considers 3rd person pronouns to constitute an intermediate class between ‘personal pronouns’ and the so-called ‘proforms’ (i.e. demonstratives, indefinites and interrogative pronouns), arguing for a particular similitude between 3rd person pronouns and demonstratives, demonstrated by 49% of his language sample, in which third person pronouns morphologically overlap or derive from demonstratives.

From a diachronic point of view, personal pronouns are “often considered to belong to the most conservative part of grammar and [are] regarded as historically stable closed-class items” (Ishiyama 2019:109), though not all forms are said to be equally stable. A concomitant notion to diachronic stability is that it also manifests itself in resistance to replacement by borrowing, which is why some (but not all) personal pronouns normally feature in lists of basic vocabulary. If we compare Swadesh’s 100-strong list of basic vocabulary with the more empirically-constructed Leipzig-Jakarta list (see Tadmor 2009:68ff), we notice that both of them coincide in recognizing the status of 1SG and 2SG as stable items, whereas the Leipzig-Jakarta list – *contra* Swadesh – ascertains the stability of 3SG and disconfirms that of 1PL; 2PL and 3PL pronouns are absent from both lists. This makes the pronominal system an important testing ground for diachronic hypotheses concerning e.g. historical relations between discrete language varieties or particular instances of morphosyntactic change, and will aid us in interpreting the variation we find within the modern SLP pronominal paradigm.

2.2. Early accounts of SLP personal pronouns

As mentioned above, SLP (earlier known as Ceylon Portuguese) has a relatively long history of written documentation, due to its importance as a lingua franca during the island’s colonial period. Dating back from the early 19th century, bibliographical sources produced in or about SLP included grammars, vocabularies, phrasebooks, bilingual or trilingual dictionaries, and translations of liturgical and biblical texts (e.g. Berrenger 1811; Callaway 1818, 1820, 1823; Fox 1819; Anon 1826; Newstead 1827, 1852, 1871; Anon. 1851; Anon. 1863). As noticed by several authors and discussed in detail in Smith (2016), the language in these sources is different from modern SLP in many fundamental ways, and, since most of them were produced by non-native speakers, significant interference and linguistic tweaking is expected to have taken place. Nevertheless, these are the only available sources referring to a variety of SLP spoken throughout the 19th century, therefore it is interesting to observe what they indicate in terms of the language’s personal pronoun paradigm. In Tables 1 and 2, we have collated the nominative and non-nominative forms of personal pronouns found in some of these early written sources, namely Berrenger (1811), Fox (1819), Callaway (1820) and Dalgado (1900):

Nominative	Berrenger 1811	Fox 1819	Callaway 1820	Dalgado 1900
1SG	eu eu mesmo voss/boss	eu eu mesmo vos/vosso	eu vos/ vosse/vosses/vosse merci	eu tu/vós
2SG	voss mesmo	vossamesmo		
3SG.M	eli eli mesmo	elli ellimesmo	elle	elle
3SG.F	ela/elé ela mesmo	ella ellamesmo	ella	ella
1PL	nossé nossé mesmo	nos nossmesmo	nos	nós/nossotros
2PL	vossé vosse mesmo	vos, vosses vossotros mesmo vossmesmo	volotros (to inferiors) vosotros (to equals) vosse-ellotros (to superiors)	vós/vossotros
3PL	elotros elotros mesmo	ellotros ellotrosmesmo	elles (of a few) ellotros	ellotros
3PL.M				elles
3PL.F				ellas

Table 1: Nominative forms of personal pronouns in early (19th-c. and early 20th-c.) sources for SLP.

This early SLP personal pronoun paradigm retains several forms of (16th-century) European Portuguese. It is characterized by a three-person set of independent forms which vary according to the pronoun's grammatical function in an utterance. Accusative, dative, genitive and ablative pronouns derive by and large from their nominative counterparts (though sometimes also non-nominative forms, such as *mi* in the case of 1SG) and involve recourse to certain prepositions (e.g. *de voss* '2SG.GEN'; *per mi* '1SG.DAT') or, exceptionally, the postposed case-marker *-su* (derived from Ptg. *sua* '3SG.GEN').

Early SLP pronouns also encode a binary distinction of number (singular and plural) and gender (masculine and feminine). In the case of gender distinctions, it is worth noting that they only apply to 3rd person pronouns; all these sources record gender distinctions in 3SG pronouns (along the lines of *elli* '1SG.M' vs. *ella* '1SG.F'), but only Dalgado (1900) registers their extension to 3PL pronouns.

Non-nominative	Berrenger 1811	Fox 1819
1SG.GEN	de eu	de mi
1SG.DAT	ne eu	per mi
1SG.ACC	eu	mi
1SG.ABL	de/por mi ~ pormi	ne, de mi
2SG	de voss	de vos/vosse-su
	ne voss	per vos
	voss	vos
	de voss/por vos	ne, de vos
3SG.M	de eli/d'eli	de elli
	ne eli	per elli
	eli	elli
	de/por eli	ne, de elli
3SG.F	d'ela	de ella
	ne éla	per ella
	ela	
	de/por ela	ne, de ella
1PL	de nossé	de nos
	ne nossé	per/par nos
	nossé	nos
	de/por nossé	ne, de nos
2PL	de vossé	de vosotros-su
	ne vossé	per vosotros
	vossé	vosotros
	de vosse/por vossé	ne, de vosotros
3PL		de ellotros
		per ellotros
		ellotros
		ne, de ellotros

Table 2. Non-nominative forms of personal pronouns in early (19th-c. and early 20th-c.) sources for SLP.

Berrenger (1811) and Fox (1819) include in their paradigms the combination of the nominative forms of personal pronouns with the element *mesmo*, as in *ela mesmo* ‘herself’, suggesting the existence of a reflexive strategy constructed this way.

Finally, these sources also record politeness distinctions, with certain 2nd person pronouns attributed to particular addressees in accordance with their social status, as explicitly indicated by Callaway (1820), who explained that *volotros* was used for “inferiors”, *vossotros* for “equals”, and *vosse-ellotros* for “superiors”. Another instance of honorificity, clearly reminiscent of Portuguese, involves the opposition between the

familiar *t*-form *tu* ‘2SG’, when addressing one individual, and the *v*-form *vos/vós* ‘2SG/2PL’, used by default for more than one entity, but also as a polite form for single individuals; yet, this politeness distinction is only given in Dalgado (1900). In addition, Callaway (1820) includes a noun-based form of polite address, *vosse merci* (< Ptg. *vossa mercê* ‘your mercy’).

2.3. Data and methodology

The DSLP corpus of modern SLP (Cardoso 2017) which provides the primary data for our synchronic study consists of over 49 hours of video and/or audio recordings of naturally occurring speech and musical performances, produced between 2015 and 2020 in several field trips to various locations in the Ampara, Batticaloa, Jaffna and Trincomalee districts of Eastern and Northern Sri Lanka (Cardoso 2017).⁴ Interviews involved single individuals, small groups (e.g. couples, households) or large groups (e.g. extended families, gatherings). The consultants were often invited to address topics related, but not restricted, to their personal experience or to Burgher identity (language, music, cuisine, crafts, occupations, religion, daily life), but there were also sessions of lexical and grammatical elicitation.

Song lyrics constitute an important part of the corpus, because the DSLP project combined linguistic and ethnomusicological research. However, there is good reason to suspect song lyrics to be particularly conservative and to potentially include linguistic resources which are learnt by rote, which, though interesting and diachronically revealing, should be approached as a separate corpus. Therefore, we have excluded the recordings of songs from our study and focused on naturally occurring speech.

While there is considerable sociolinguistic diversity among the 152 recorded consultants, some profiles (geographical and social) are especially well represented. Table 3 below provides some basic information on the profile of speakers recorded in the DSLP corpus:

	Eastern Province					Northern Province
	Batticaloa	Kalmunai	Eravur	Valaichchenai	Trincomalee	Jaffna
Sex						
F	43	3	2	6	26	1
M	39	3	-	5	20	4

⁴ More detailed information on the corpus (constitution, transcription conventions, workflow and types of archived materials) is given in Cardoso et al. (2019). The materials are made available on the Endangered Languages Archive, at <https://www.elararchive.org>.

Age						
0-14	4	-	-	-	1	-
15-24	3	-	-	-	4	-
25-64	16	2	-	4	21	2
≥ 65	9	1	-	-	13	3
N/A	50	3	2	7	7	-
Total	82	6	2	11	46	5

Table 3. Distribution of speakers in the DSLP corpus (Cardoso 2017) by region, sex, and age group (N=152).

As can be seen in Table 3, the gender distribution only slightly favors women, with 81 female consultants (53%) and 71 male consultants (47%). Consultants vary in age, despite the fact that there is a preponderance of middle-aged or elderly interviewees, as a result of the current sociolinguistic distribution of SLP fluency described above. However, there is a clear predominance of speakers from the urban areas of Batticaloa and, to a lesser extent, Trincomalee. The towns of Eravur and Valaichchenai also belong to the Batticaloa district, whereas Kalmunai is the only town standing for the Ampara district. The number of consultants from Jaffna may seem modest but, according to our current knowledge, it represents a very large percentage of all the speakers in the Northern Province.

The data for this study was gathered by searching the DSLP corpus within multiple ELAN (Brugman & Russel 2004) annotation files, automatically extracting all personal pronouns (and their phonetic and morphological variants) produced by each speaker. The collected sample was then organized in a spreadsheet, filtering out target word forms that occurred in song lyrics and those that had been produced by the interviewers.

3. Modern SLP personal pronouns

Within the modern SLP nominal domain, case distinctions involve the addition of a series of postposed morphemes – which Smith treats as suffixes in the case of the oblique (i.e. accusative + dative), genitive, and locative, and as postpositions in other cases – to the nominative (or, in the case of some postpositions, the genitive) form of the noun (see Smith 2013:113). Yet another productive nominal suffix is the plural marker *-s*. The addition of these morphemes to nouns produces transparent multimorphemic forms containing case and number information (e.g. *luváára* ‘place’ ~ *luváaras* ‘places’ ~ *luváarantu* ‘in (the/a) place’).

When it comes to the personal pronoun paradigm, however, the situation is somewhat different. While the same additional morphemes also occur, they do not account for all pronominal forms, several of which are monomorphemic and suppletive. Such forms occur in the nominative, oblique, and genitive series, which will constitute the core of our analysis in this study.

Ian Smith's descriptive work on SLP (Smith 1977; 2013) records the pronominal forms given in Table 4:

	Nominative	Oblique	Genitive
1SG	<i>eev</i>	<i>parim/parmi</i>	<i>mīna</i>
2SG.NHON	<i>boos</i>	<i>boos-pa</i>	<i>bosa</i>
3SG.M.NHON	<i>eli</i>	<i>eli-pa</i>	<i>eli-su</i>
3SG.F.NHON	<i>ela</i>	<i>ela-pa</i>	<i>ela-su</i>
3SG.HON	<i>osiir</i>	<i>osiir-pa</i>	<i>osiir-su</i>
1PL	<i>noos</i>	<i>noos-pa</i>	<i>nosa</i>
2PL	<i>botus</i>	<i>botus-pa</i>	<i>botus-su</i>
3PL.M.NHON	<i>elis</i>	<i>elis-pa</i>	<i>elis-su</i>
3PL.F.NHON	<i>elas</i>	<i>elas-pa</i>	<i>elas-su</i>
3PL.HON	<i>etus</i>	<i>etus-pa</i>	<i>etus-su</i>

Table 4. SLP personal paradigm (adapted from Smith 2013:114).

As can be seen in Table 4, honorificity is an important sociolinguistic factor encapsulated in the pronominal paradigm. In 3rd person, honorific and non-honorific address involve dedicated forms. In the case of the 2nd person, however, no distinction is encoded in plural pronominal forms, but the 2PL form *botus* is also used as a honorific form of address for 2SG (Smith 1977:69), thereby creating a functional syncretism between 2SG.HON and 2PL. In addition, non-honorific 2nd person pronouns may also be avoided by substituting them with a title such as *sinhoor* 'gentleman, mister' (Smith 2013).

In this paradigm, we can also observe that the transparent use of plural suffix *-s* is restricted to non-honorific 3M and 3F pronouns, whereas all other plural pronominal forms are suppletive. When it comes to case distinctions, the oblique suffix *-pa* is recognized in all oblique forms except 1SG, while the genitive suffix *-su* is not applied in any 1st person pronoun nor in non-honorific 2SG.

Smith's description of the SLP pronominal paradigm, given in Table 4, is based on a corpus (Smith 1973) containing approximately 15h of interviews, collected mostly in Batticaloa. With nearly 50h and 152 interviewees, the DSLP corpus is larger and was collected in several locations of the Batticaloa, Trincomalee, Ampara, and Jaffna districts (see section 2.3). Unsurprisingly, this corpus provides alternative forms for several pronouns, resulting in the updated paradigm given in Table 5. Orthography varies slightly with respect to Table 4, reflecting the DSLP team's preferred orthography (for an

explanation of which, see Cardoso et al. 2019:12-16); all forms undocumented in previous accounts of the modern SLP paradigm are underlined:

	Nominative		Oblique		Genitive	
	Forms	Tokens	Forms	Tokens	Forms	Tokens
1SG	<i>eev</i>	2203	<i>páármí</i>	337	<i>minha</i>	1409
	<u><i>ee</i></u>	77	<u><i>paarmi</i></u>	62	<u><i>eev-su</i></u>	22
			<u><i>páámi</i></u>	14	<u><i>paami-su</i></u>	16
			<u><i>paami</i></u>	204	<u><i>minha-su</i></u>	2
			<u><i>pááim</i></u>	26	<u><i>eev-pa</i></u>	1
			<u><i>paaim</i></u>	45		
			<u><i>páármí-pa</i></u>	3		
			<u><i>paarmi-pa</i></u>	3		
			<u><i>paami-pa</i></u>	29		
			<u><i>paaim-pa</i></u>	104		
			<u><i>eev-pa</i></u>	56		
2SG.NHON	<i>boos</i>	497	<i>boos-pa</i>	241	<i>bósa</i>	343
3SG.M.NHON	<i>eli</i>	167	<i>eli-pa</i>	57	<i>eli-su</i>	15
3SG.F.NHON	<i>éla</i>	100	<i>éla-pa</i>	34	<i>éla-su</i>	22
3SG.HON	<i>osiir</i>	815	<i>osiir-pa</i>	193	<i>osiir-su</i>	144
	<u><i>esiir</i></u>	24				
1PL	<i>nóós</i>	2599	<i>nóós-pa</i>	985	<i>nósa</i>	2182
2PL	<i>botus</i>	377	<i>botus-pa</i>	113	<i>botus-su</i>	87
	<u><i>botrus</i></u>	58	<u><i>botrus-pa</i></u>	6	<u><i>botrus-su</i></u>	11
3PL.M.NHON	<i>elis</i>	31	<i>elis-pa</i>	13	[<i>elis-su</i>]	0
3PL.F.NHON	<i>élas</i>	3	<i>élas-pa</i>	1	[<i>élas-su</i>]	0
3PL.HON	<i>etus</i>	1139	<i>etus-pa</i>	384	<i>etus-su</i>	259
	<u><i>etrus</i></u>	333	<u><i>etrus-pa</i></u>	86	<u><i>etrus-su</i></u>	44

Table 5. SLP personal pronouns recorded by the DSLP project, with number of tokens in the DSLP spoken corpus (Cardoso 2017); forms in square brackets were attested in elicitation and fieldnotes but do not occur in the oral corpus.

The nature of the variation encapsulated in Table 5 is not uniform. In some cases, what is at stake are mere differences in the form of particular suppletive pronouns, involving alternations at the level of certain segments (as in 3SG.HON, *osiir* ~ *esiir*), the addition/deletion of certain segments (as in 2PL and 3PL, *botus* ~ *botrus*, *etus* ~ *etrus*, and derived forms thereof), or metathesis (as in 1SG.OBL, *páámi/paami* ~ *pááim/paaim*).

In other cases, variation involves oppositions between monomorphemic⁵ (i.e. suppletive) and bimorphemic (i.e. analytic) forms for oblique and genitive pronouns. In general, across the paradigm, bimorphemic forms consist of the addition of the corresponding case morpheme (*-pa* for oblique and *-su* for genitive) to a base which is equal to the nominative pronoun; as such, and considering that this also holds true of the use of these case-markers with nouns, we can say that it constitutes the regular pattern. However, in those domains in which we observe variation, namely 1SG.OBL and 1SG.GEN, some bimorphemic alternatives follow this pattern (the cases of *eev-pa* and *eev-su*), while others do not (i.e. they select a base form equivalent to a monomorphemic oblique or genitive pronoun); and, as for the latter, either the case associated with the base form and the case-marker coincide (as in *páármí/paarmi/paami/paaim-pa*, and *minha-su*) or they diverge (as in *paami-su*). The one form which falls outside of these possibilities is the single recorded instance of *eev-pa* used in a clearly genitive context, in which the oblique morpheme intrudes in the distributional space of the genitive; we interpret this as a matter of non-canonical case-selection and, therefore, will treat the occurrence as an outlier.

In addition to this – though not evident from Table 5 –, other instances of variation in pronominal selection can be observed in the domain of pragmatics, involving the system of honorific address (in 2SG and 3SG/3PL), and a particular case of number syncretism pertaining to 1st person genitive. These cases of variation in the DSLP corpus will be discussed in the next section.

4. Variation

4.1 Variable forms

When we look at the modern SLP pronominal paradigm (Tables 4 and 5), we identify different pronominal forms for 1SG, 2PL, 3SG.HON, and 3PL.HON. Some of the oppositions involve morphological differences and will be discussed in 4.2; others, however, are of a different nature. Some are purely phonetic:

- the difference in vowel height between *páármí/páámi/pááim* (with a long low central vowel [a:]) and *paarmi/paami/paaim* (with a long near-low central vowel [ɐ:]);

⁵ The question may arise why 1SG oblique forms *páármí/paarmi/páámi/paami/pááim/paaim* are considered monomorphemic in modern SLP, especially when early sources record what appears to be a preposition *par/per/por* with various personal pronouns (see Tables 1 and 2) and nouns, including the 1SG sequences *por mi* and *per mi*. That early preposition but also the modern oblique case suffix *-pa* are indeed derived from the Portuguese preposition *para* ‘for’ or *por* ‘by’. But, while analytical *par/per/por mi* does appear to be the diachronic source of *páármí/paarmi/páámi/paami/pááim/paaim*, we cannot recognise a preposition or prefix here because it occurs nowhere else in the modern SLP nominal system (i.e. it is not productive); a prefix *pa-* does exist, but it is strictly a verbal prefix marking the infinitive (see Smith 2013).

- the difference in vowel backness and roundedness between *osiir* (the most frequent variant, with a high-mid back rounded vowel [o]) and *esiir* (with a high-mid front unrounded vowel [e]).

Some other differences are the result of metathesis:

- the difference between *páámi/paami* and *pááim/paaím* (in the DSLP corpus);
- the difference between *parmi* and *parim* (recorded only in Smith's corpus, see Table 4).

In other cases still, observed differences involve addition/deletion of a segment:

- the difference between *eev* and *ee* (the latter only attested in the Jaffna variety);
- the difference between *páármí/paarmi* and *páámi/paami*;
- the difference between *botus* and *botrus*;
- the difference between *etus* and *etrus*.

With the exception of the *eev* ~ *ee* opposition, all cases of deletion refer to the flap, creating an opposition between forms with and without [ɾ]. For ease of reference, we will identify one type as *R+* (i.e. with [ɾ]) and the other as *R-* (i.e. without [ɾ]). Table 6 indicates the percentage of speakers in the DSLP corpus who use 1SG.OBL, 2PL and 3PL.HON pronouns of the *R+* type, of the *R-* type, or alternate between both:

	<i>R-</i>	Both <i>R-</i> and <i>R+</i>	<i>R+</i>	Total
1SG.OBL	19.5% (17)	34.5% (30)	46% (40)	100% (87)
2PL	84.4% (65)	11.7% (9)	3.9% (3)	100% (77)
3PL.HON	77.4% (89)	20.9% (24)	1.7% (2)	100% (115)

Table 6: Use of competing 1SG.OBL, 2PL and 3PL.HON forms (% speakers + number of speakers).

While, in all cases, there is a considerable percentage of speakers who alternate between competing forms, the highest proportion is always found among those for whom one of the alternatives is categorical. However, the preference is not always the same. In the case of 1SG.OBL, *R+* forms are synchronically dominant, with 46% of speakers using those consistently, and an additional 34.5% alternating between *R+* and *R-* forms. In 2PL and 3PL.HON, on the other hand, the preference goes to *R-* forms and is even more pronounced, with a vast majority of speakers using only *R-* forms (84.4% and 77.4%, respectively), and very few only *R+* forms (3.9% and 1.7%, respectively). This particular clustering is interesting in that it coincides with a difference in the phonological context involved in

the *R+* ~ *R-* alternation. Even though all cases involve the presence or absence of a flap, in 1SG.OBL, the segment (when realized) occurs in a syllable coda, whereas, in 2PL and 3PL.HON, what is at stake is an alternation between a simple onset [t] and a complex onset [tr] in the pronoun's final syllable.

A closer look at the distribution of *R+* and *R-* forms reveals that there is a single regional variant in which the competition we just described does not apply. The data collected in Jaffna contains no instances of *R+* forms, not just for 2PL and 3PL.HON pronoun, but also for 1SG.OBL. Elsewhere, the synchronic instability in the pronunciation of these pronominal forms raises the question of whether we are witnessing a process of diachronic change – and, if so, in which direction it proceeds. We would like to argue for a scenario which interprets *R+* forms as older or more conservative, and *R-* forms as more recent. Bearing in mind the distribution in Table 6, this hypothesis implies that the process of change, overall, is less advanced in 1SG.OBL, and approaching its completion in 2PL and 3PL.HON; and completed in Jaffna.

The rationale behind this proposal rests on a few considerations. First of all, the etymological criterion favors the early presence of [tr] in all these cases: the *páá(r)/paa(r)*- element of 1SG.OBL derives from the Ptg. prepositions *para* 'for' or *por* 'by' (see Note 5), and the *-t(r)us* of 2PL and 3PL.HON derives from the Ptg. pronoun *outros* 'others'. As a matter of fact, the corresponding forms in 19th-century sources (see Tables 1 and 2) all reflect the etymological flap. It is true, as already mentioned, that the data in these sources needs to be taken with a grain of salt, but it is significant that all sources surveyed coincide in this respect.

Another fact which reinforces the likelihood of SLP having preserved the flap in the case of 2PL and 3PL.HON (rather than introducing it at a later stage) is that, across its lexicon, we observe a general tendency for similar etymological complex onsets combining a plosive and a flap to have been preserved, and they are to be found in word-initial and word-medial position alike (see Smith 1977:43-52). A few examples include the modern SLP words *létriiya* 'string hopper' (from Ptg. *aletria* 'vermicelli pasta'), *káátru* 'four' (from Ptg. *quatro* 'four'), *lembráá* 'to think' (from Ptg. *lembrar* 'to remember'), *triiya* 'to bring' (from Ptg. *trazer* 'to bring'), *trukáá* 'to (ex)change' (from Ptg. *trocár* 'to (ex)change'), among others. In fact, SLP's acceptance of this type of complex onset has even made it possible for the language to develop a few which were not present in the Portuguese etyma, either as a result of vowel deletion (often involving other segmental transformations), as in *kambráám* 'shrimp' (from Ptg. *camarão* 'shrimp') and *páástru* 'bird' (from Ptg. *pássaro* 'bird'), of metathesis, as in *brumeey* 'red' (from Ptg. *vermelho* 'red'), or of other types of innovations, as in *usprutáál* 'hospital' (from Ptg. *hospital*).

According to this interpretation, then, the *R-* 2PL and 3PL.HON pronominal forms constitute exceptional cases and the resulting simple ~ complex onset alternation reflects

diachronic innovation. However, they are not alone, as there is at least one other case of alternation between simple and complex onsets involving a plosive and a flap, viz. *primeer/prumeer* ~ *pimeer/pumeer* ‘first, earlier’. Considering that this term derives from Ptg. *primeiro* ‘first, earlier’, which contains the complex onset in question, it would also be difficult to argue for the loss and later reintroduction of the flap. As an important adverbial, *primeer* ~ *pimeer* is highly frequent in discourse, just as *botrus* ~ *botus* and *etrus* ~ *etus*, which may be at the root of the motivation for onset simplification, in all three cases. In addition, the *R*- form *pimeer* is much more frequent in the global DSLP corpus (525 tokens) than any of the competing forms (the closest being *primeer*, with 191 tokens), a synchronic predominance of the *R*- form which mirrors what we described for the 2PL and 3PL.HON pronouns; and, consistently with our observations with respect to pronominal forms, *R*+ *primeer/prumeer* is entirely absent from the Jaffna data.

4.2 Monomorphemic vs. bimorphemic forms

In the DSLP corpus, we find competing monomorphemic (suppletive) and bimorphemic (analytic) forms for the oblique and genitive cases of 1SG (see section 3).

For 1SG.OBL, Smith records only monomorphemic *parmi* or *parim* (see Table 4). As mentioned earlier, the DSLP corpus also includes the suppletive oblique form *páarmi/paarmi*, as exemplified in (1), as well as variants *páami/paami* and *páaim/paaim* (Table 5):

- (1) ***páarmi*** *káátru podhiyáás,* *doos mááchi, doos féémia.*
 1SG.OBL four children two male two female
 ‘I have four children, two sons and two daughters’. (Trincomalee; Cardoso 2017:009_1)

However, a few bimorphemic variants are also to be found: a) *eev-pa*, as in (2); and b) a set of forms, as in (3), consisting of a variant of the suppletive oblique pronoun plus the case-marker, viz. *páarmi/paarmi-pa*, *paaim-pa*, and *paaaim-pa*.

- (2) ***eev-pa*** *triinta sees* *"years" ta-fikáá.*
 1SG.NOM-OBL thirty six years PRS-become
 ‘I am turning thirty-six years (of age)’. (Trincomalee; Cardoso 2017:slp071_1)
- (3) ***paaaim-pa*** *ya-ka-iskisa*
 1SG.OBL-OBL PST-PFV-forget
 ‘I have forgotten’. (Jaffna; Cardoso 2017:slp069_4)

Given that, in the oblique series, all forms but those of 1SG are strictly analytic (consisting of the application of the oblique case-marker *-pa*), the formation of bimorphemic 1SG forms can be seen as a case of regularization of the paradigm. However, there is a difference between *eev-pa* and the other bimorphemic forms, in that the former constitutes the full regularization of the oblique pronoun formation strategy by which the oblique case-marker selects a nominative base form of the pronoun, whereas the latter adds that marker to an intrinsically oblique pronominal form, thereby resulting in redundant double case-marking.

The latter strategy (i.e. the selection of a non-nominative base) may be explained by transfer from Tamil. In this language, pronouns (as well as nouns, see Schiffman 1999:27) have two monomorphemic forms, the nominative and the oblique, and case-markers attach to the oblique base. In the case of personal pronouns (see Schiffman 1999:117-119), distinct nominative and oblique forms exist for 1st and 2nd person forms (e.g. *naan* ‘1SG.NOM’ ~ *en* ‘1SG.OBL’; *nii* ‘2SG.NOM’ ~ *on* ‘2SG.OBL’), but not for 3rd person (e.g. *avan* ‘3M.SG.NOM/OBL’); and dative forms of the pronoun involve the addition of the dative case-marker *-(u)kku* to the oblique base, whenever the distinction holds (e.g. *enakku* ‘1SG.DAT’; *onakku* ‘2SG.DAT’), or to the invariable base, in the other cases (e.g. *avanukku* ‘3M.SG.DAT’). This pattern may explain the option of certain SLP-speakers to select a non-nominative base for the oblique suffix in 1SG.OBL, but it should be noted that these SLP forms and the Tamil suffixed personal pronouns are not entirely equivalent. The Tamil oblique actually carries genitive semantics, since the bare oblique form (of a noun or pronoun) can indicate possession, even though there are also genitive markers; and, in SLP, the oblique conflates dative and accusative case. Therefore, the full application of the Tamil model to the formation of SLP bimorphemic pronouns should result in the selection of a genitive base (which does occur, though marginally, in the genitive form *minha-su*; see below).

With 139 occurrences produced by 11 speakers, double case-marked forms constitute a substantial proportion of 1SG.OBL forms in the corpus. However, they are clearly associated with the Jaffna variety: 4 out of 5 speakers interviewed in Jaffna produce double case-marked forms, and their speech accounts for 130 tokens, which represents 81.8% of the 159 instances of 1SG.OBL forms collected in the Jaffna section of the corpus; and globally, while the 4 Jaffna speakers make up just 36.4% of the 11 speakers who use such forms, they produce 93.5% of all occurrences of double case-marked 1SG.OBL pronouns in the DSLP corpus.

Having established the almost exclusive association of double case-marked 1SG.OBL forms with the Jaffna variety of SLP, let us now consider the use of the remaining forms. Oblique *eev-pa*, the bimorphemic form which represents the simple regularization of the paradigm with respect to 1SG.OBL, occurs 56 times in the DSLP corpus. Table 7

indicates the percentage of speakers who use this 1SG.OBL form only, suppletive forms only, or alternate between both:

<i>Eev-pa</i> only	Both	Suppletive only	Total
6.7% (7)	16.3% (17)	77% (80)	100% (104)

Table 7: Use of *eev-pa* or a suppletive 1SG oblique form (% speakers + number of speakers).

The distribution in Table 7 clarifies that the suppletive strategy is clearly dominant for 1SG.OBL. Nonetheless, nearly a quarter of all speakers in the DSLP corpus employ *eev-pa*, although most of those alternate between that bimorphemic form and a monomorphemic one. Looking at the profile of these speakers who produce *eev-pa*, it is difficult to assign the overall distribution of this form to particular age, gender, or geographical groups. The only possible generalization is that all speakers for whom the use of the bimorphemic *eev-pa* 1SG.OBL pronoun is categorical are under 40 years of age. From an “apparent time” perspective, this distribution is consistent with a scenario of language change in progress, but we believe the particular directionality of change has a deeper meaning (see section 5).

When it comes to 1SG.GEN pronominal forms, the DSLP corpus records a suppletive form *minha* (equivalent to Smith’s *mīna*), exemplified in (4):

- (4) *minha* *pááy* *rábaana* *lo-dááy*.
 1SG.GEN father rabana HAB-play
 ‘My father used to play the rabana⁶’. (Batticaloa; Cardoso 2017:slp040_3)

In addition, similarly to the oblique forms, the corpus also contains three bimorphemic forms (plus a single occurrence of *eev-pa* in a genitive context, which, as explained in section 3, we interpret as an instance of non-canonical case-selection, and therefore disregard here): a) *eev-su*, as in (5), in which the regular genitive case-marker *-su* attaches to a nominative pronominal base; b) *paami-su*, as in (6), in which the case marker selects an oblique base form of the pronoun; and c) *minha-su*, as in (7), in which the genitive case-marker attaches to the suppletive genitive pronoun; however, this form – which constitutes yet another case of redundant double case-marking and, as explained above, closely matches the Tamil model for the formation of non-nominative pronouns – is only produced twice, by a single consultant from Jaffna.

⁶ A *rabana* is a type of traditional drum used in Sri Lanka, including in the musical traditions of the Portuguese Burghers.

- (5) *eev-su kázmeentu "two thousand nine, August second" ya-macháá.*
 1SG.NOM-GEN wedding two thousand nine, august second PST-happen
 ‘My wedding took place on August 2nd, 2009.’ (Batticaloa; Cardoso 2017:slp020_6)
- (6) *paami-su kriyaansa ya-ka-kazáá.*
 1SG.OBL-GEN children PST-PFV- marry
 ‘My children have gotten married.’ (Jaffna; Cardoso 2017:slp064_1)
- (7) *minha-su santáá fáátu.*
 1SG.GEN-GEN sit thing
 ‘[This is] my chair.’ (Jaffna; Cardoso 2017:slp2010_1)

The form *paami-su* is somewhat parallel to the oblique forms *páármí/paarmi-pa*, *paami-pa*, and *paaim-pa* discussed above in that a case-marker selects an oblique base form of the personal pronoun. In this case, however, it does not result in redundant double case-marking. In fact, with only 16 occurrences in the DSLP corpus, *paami-su* is also rather marginal. Nonetheless, the distribution of these few cases is relevant, because they are all produced by 3 speakers from Jaffna who are part of the group of 4 whose speech accounted for almost all instances of double case-marked 1SG.OBL forms (see above). This reveals a certain consistency in the grammar of these speakers from Jaffna and, at the same time, reinforces the link between non-nominative base selection in case-marked 1SG pronouns and the Jaffna community.

Genitive *eev-su* constitutes a more canonical case of paradigm regularization, not only because it mirrors the bimorphemic constitution of most other genitive pronominal forms, but also in that it makes the genitive case-marker attach to the expected nominative base form of the pronoun. With only 22 occurrences in the corpus, it is less marginal than *paami-su*, but also not particularly frequent. Table 8 below indicates the proportion of speakers who make categorical use of this form, of the suppletive form *minha*, or alternate between both:

<i>Eev-su</i> only	Both	<i>Minha</i> only	Total
4,2% (5)	3,3% (4)	92,5% (112)	100% (121)

Table 8: Use of *eev-su* or the suppletive 1sg genitive form *minha* (% speakers + number of speakers).

When we compare Tables 7 and 8, we notice that, in the case of the genitive, the use of the bimorphemic form – both categorically and variably – is less prevalent than in the

oblique case. Looking at the sociolinguistic profile of the speakers, it is also more difficult to associate this strategy to a particular age, sex, or geographical provenance, but it is noteworthy that 4 out of the 9 speakers who produce this form are under 40 – 2 of whom, in fact, under 16 years of age. In addition, another fact also reinforces the connection between bimorphemic 1SG pronouns and lower age groups: only 2 consultants (both from Batticaloa) make exclusive use of both *eev-pa* for 1SG.OBL and *eev-su* for 1SG.GEN, one aged 12, and the other in her late 20s.

4.3 Honorificity

As we noticed in Tables 4 and 5 above, the full system of honorificity encoded in the pronominal paradigm of modern SLP includes different forms (HON = ‘honorific’ and NHON = ‘non-honorific’) for 2nd and 3rd person, which we will explore here in turn. In terms of their distribution in discourse, by and large, NHON pronominal forms are commonly associated with referents younger than the speaker – especially if the referents are children or teenagers – and HON pronominal forms are used in all other cases.

In 2nd person, there is a 2SG.NHON form *boos* (with a suppletive genitive form, *bósa*) used for familiar addressees (children especially), and a 2PL form *botus* or *botrus* (see 4.1) which, in addition, functions as a 2SG.HON form of address, creating a functional syncretism between 2PL and 2SG.HON, as previously described by Smith (1977:69). Therefore, in reality, honorificity distinctions in 2nd person only surface in the singular. Examples in (8) and (9) demonstrate the use of both forms with singular addressees:

- (8) *eev kum dáádha kum ta-vii boos mee nun-teem áki.*
 1SG and father and PRS-come 2SG.NHON FOC NEG-be here
 ‘Me and my father came, but you [= familiar addressee] weren’t here.’
 (Trincomalee; Cardoso 2017:slp077_1)

- (9) *mesa-papiyáá botus.*
 OBLG-speak 2SG.HON
 ‘You [= respected single addressee] must speak.’ (Kalmunai; Cardoso 2017:slp027_1)

As demonstrated in Table 5, 2SG.NHON forms (*boos*, *boos-pa* and *bósa*) alone constitute the majority of 2nd person pronouns in the DSLP corpus. In addition, many occurrences of *botus*, *botus-pa* and *botus-su* also have a singular reference, and are therefore 2SG.HON pronominal forms. The fact that 2SG pronouns outweigh 2PL pronouns in the DSLP corpus is perhaps not surprising and derives from its very nature, considering that it contains a

large number of recording sessions involving a single interviewer and/or a single interviewee.

In 3rd person, as shown in Table 5, NHON pronouns code gender distinctions in both singular and plural, and their plural forms are transparently constructed with the plural suffix -s. HON pronouns, on the other hand, have suppletive forms for singular and plural and do not encode gender. Examples (10) and (11) demonstrate the use of HON and NHON 3SG pronouns:

- (10) *"portuguese" podhi unga ya-tinha, páármí eli-pa*
 portuguese boy one PST-PST.be 1SG.OBL 3SG.NHON-OBL
nuku-sava.
 NEG-know
 'There was a Portuguese boy, I didn't know him.' (Batticaloa; Cardoso 2017:slp037_1)
- (11) *avóóra osiür-pa indung siinku áánu teen osiür-su prenda*
 now 3SG.HON-OBL still five year EXS 3SG.HON-GEN studies
kaváá-pa.
 finish-PURP
 'He still has another five years to go to finish his studies' (Trincomalee; Cardoso 2017:slp005_1)

Table 9 below shows the distribution of HON and NHON 3rd person pronouns in the corpus, revealing, first of all, that NHON pronouns are much less common than HON pronouns:

	SG	PL	Total
NHON	89,2% (395)	10,8% (48)	100% (443)
HON	34,5 % (1176)	65,6% (2245)	100% (3421)

Table 9: Distribution (number of occurrences) of 3rd person HON and NHON pronouns in the DSLP corpus (Cardoso 2017).

Table 9 also makes it clear than NHON pronouns are especially infrequent with plural reference. In this respect, it is interesting that, in addition to the canonical contexts in which we would expect the selection of a NHON pronominal form (i.e. for younger or familiar referents), in the case of 3PL we also see NHON pronouns applied to generic referents, as demonstrated in example (12):

- (12) *nóós nuku-oyáá, elis ta-faláá ‘Malaysia’-su “wood”*
 1PL NEG-see 3PL.M.NHON PRS-say Malaysia-GEN wood
isi faláátu.
 DEM QUOT
 ‘We didn't see it, they [i.e. people] say it's Malaysian wood.’ (Batticaloa; Cardoso 2017:slp021_2)

In the DSLP corpus, the application of pronominal honorificity distinctions functions largely as expected, with one notable exception: the speech of the speakers from Jaffna. The system of honorificity in Jaffna appears to be impoverished in comparison with the speech of speakers from other regions represented in the corpus. In the case of the 2nd person, this is manifested in the fact that Jaffna speakers appear to use mostly *boos*, *boos-pa* and *bósa* (329 occurrences vs. 26 in which the honorific form is used) in contexts where other varieties would select a HON form, thereby contradicting the robust functional syncretism between 2PL and 2SG.HON observed in the other varieties. When it comes to the 3rd person, politeness oppositions also appear to be non-existent, as one singular option and one plural option are never recorded in the available data: in the singular, M.NHON *eli* and F.NHON *éla* occur, but the HON form *osiir* does not; whereas, in the plural, HON *etus* is attested, whereas NHON *elis* or *élas* are not.

Another instance of variation which needs to be mentioned here involves a partial syncretism between 1SG.GEN and 1PL.GEN pronouns, which has not been previously described in the literature. While, as indicated in Tables 4 and 5, *minha* is the canonical 1SG.GEN pronoun (with the alternative forms discussed in 4.2 above) and *nósa* is the canonical 1PL.GEN pronoun, the DSLP corpus contains several instances of *nósa* used in contexts that are clearly singular, such as in example (13) – in which, in a strictly monogamous community, the referent of the pronoun (a spouse) can only be a single person; contrast it with example (4) above.

- (13) *nósa máriidu ja-ka-mura-pa dispoos, eev kustuura*
 1PL.GEN husband PST-PFV-die-OBL after 1SG sewing
mee ya-kusa.
 FOC PST-sew
 ‘After my husband died, I started sewing.’ (Trincomalee; Cardoso 2017:slp003_5)

A study of such cases in the DSLP corpus reveals that this type of variation in genitive 1SG pronominal reference bears some relationship with notions of honorificity, in that this use of *nósa* appears to be restricted to contexts in which speakers refer to their rapport with a relative who commands a degree of honorificity. In the corpus, we only identify

this specific use of *nósa* in sentences that refer to a husband (as in (13)), a wife, an uncle, or adult siblings.

5. Discussion

While, as discussed in section 1, researchers dealing with variation in an endangered or lesser-studied language often face the challenge of having no prior knowledge of it, in our case, Smith's (2013) description of the SLP personal pronoun paradigm assisted us in identifying variation in this domain. In general, our study confirms that the pronominal forms previously described in the literature – i.e. those in Table 4 – are indeed the most robust and most frequent in discourse, and for that reason can be considered canonical, the only exception being the alternate 1SG.OBL form *parim* reported by Smith (2013:114), but absent from the DSLP corpus.

Having said that, our survey also unearthed numerous alternate forms and instances of variability in the expression of personal pronouns. In some cases, the available data does not reveal a clear association between a variant and a given group of speakers. In other cases, however, a distribution pattern emerges, associating a variant with a subset of the speech community, either socially or geographically defined. In most of them, the fact that these groups of speakers are those among which language shift is most prevalent intersects with the fact that the variant in question is consistent with one of the processes of structural change expected to result from language loss (see section 1) to suggest that language obsolescence does underlie some of variability observed in modern SLP.⁷

As described in section 2, the youngest sections of the Burgher community report the lowest levels of proficiency in SLP and frequency of use of the language (Pereira 2019). Two alternate pronominal forms are especially associated with this sociolinguistic profile: the bimorphemic 1SG.OBL *eev-pa* and the 1SG.GEN *eev-su* (see 4.2). The association is especially strong in the case of *eev-pa*, since it was observed that all speakers for whom this form is categorical (i.e. who never use one of the suppletive forms) are under 40; in the case of *eev-su*, the association is based on the fact that, of the few speakers who use it at all (whether or not in competition with the suppletive alternative *minha*), nearly half are under 40; in addition, there are only two consultants

⁷ There are two cases in which a particular variant has a clear geographical association and yet it is not so clear that it necessarily derives from language loss: a) the fact that the data from Jaffna contains no R+ pronominal forms (while R+ and R- compete elsewhere) does result in a reduced level of societal variation in this variety and in this domain, but whether this development coincides with any of the processes impacted by obsolescence cannot be ascertained without a more detailed study of Jaffna SLP phonology; and the same could be said of b) the fact that the reduced form of 1SG *ee* is only identified in speech from Jaffna.

for whom both *eev-pa* and *eev-su* are categorical, and they belong to this age bracket. What unites these two forms, within the framework of the variation described in section 4, is that they represent instances of regularisation by analogy with the structure of nearly all other non-nominative forms in the pronominal paradigm. If we consider suppletive 1SG forms to be older than the suffixed 1SG forms of modern SLP (as suggested by the early data in Tables 1 and 2, and by proposed scenarios of convergence towards Tamil affecting word order; see Bakker 2006), we must admit that the preference for *eev-pa* and *eev-su* constitute innovations. So, these speakers engage in paradigm levelling, one of the processes said to be especially prevalent in situations of language obsolescence (see section 1).⁸

However, young speakers are not the only group associated with a preference for bimorphemic 1SG.OBL and 1SG.GEN forms. The small SLP-speaking community from Jaffna also shows a clear preference for bimorphemic pronominal forms, but, in this case, the base selected by the case-suffix is not nominative: it is most often oblique (in 1SG.OBL *paami-pa/paaim-pa* and 1SG.GEN *paami-su*), and, on two occasions only, genitive (in 1SG.GEN *minha-su*). What is interesting in this case is that, to some extent, it also constitutes a case of paradigm levelling (in the sense that the bimorphemic nature of most SLP non-nominative pronouns is extended to the 1SG domain), but introduces an innovation: the selection of a non-nominative base. This innovation, we argued above, is not random and reflects the influence of Tamil, a case of adstrate transfer which appears to impact the Jaffna community more than speakers elsewhere, including the under-40s mentioned in the previous paragraph. The fact that SLP language loss is particularly advanced in Jaffna is perhaps implicated in this development, as it combines two of the types of change said to be reinforced by obsolescence (see section 1): paradigmatic regularization and permeability to the structural influence of the community's dominant language.

This brings us to yet another specificity of the Jaffna community: its impoverished system of honorificity as encoded in personal pronouns. To be clear, in the absence of any records prior to our own documentation, it is impossible to determine whether Jaffna SLP ever had a system of honorificity as robust as what we find in the Eastern varieties. However, there are reasons to suppose it did: on the one hand, of course, the prevalence of such a system in the other modern varieties of SLP, but also the hints of honorificity in the 19th-century descriptions of SLP – which were produced at a time when the language was more vital in Jaffna and across the island; on the other hand, the fact that

⁸ The fact that one of these speakers is a 12-year old raises the question of whether language acquisition could be involved in this type of paradigm regularisation – the rationale being that, in the naturalistic acquisition of SLP, the introduction of a suppletive form in the paradigm may be preceded by a stage in which the regular, more transparent bimorphemic form is dominant. However, given that a 12-year-old is not a young child, and in the absence of longitudinal studies of SLP acquisition, this question cannot be resolved.

the more limited set of 3rd person pronouns used in Jaffna contains forms which, in the Eastern varieties, are associated with both HON and NHON; and, finally, the fact that honorificity distinctions are a conspicuous feature of the major languages of Sri Lanka, Tamil (Schiffman 1999:115-118) and Sinhala (Chandralal 2010:267-272). If this is correct, then the Jaffna variety effectively *lost* its previous system of honorificity, which constitutes an instance of stylistic shrinkage, also associated with situations of language obsolescence (see section 1). Interestingly, however, the influence from the dominant language invoked above in connection with the formation of non-nominative 1SG pronouns is not seen to operate in this case, since the Tamil model has not prevented the loss of honorificity encoding in SLP.

6. Concluding remarks

Our study of modern SLP personal pronouns has revealed that even such a narrow domain of grammar can be characterized by plenty of variability. In the case of the data included in the DSLP corpus, the nature of the observed variation ranges from phonetic alternation (as in the case of segmental differences between pairs of forms such as *osiir* and *esiir*) to the effects of diachronic processes of variant competition and substitution (as in the case of the *R+* and *R-* forms of 2PL and 3PL pronouns); from strategies of paradigm regularization resulting in morphologically different pronominal forms (as in the cases of the opposition between suppletive and analytic pronouns) to differences in pragmatic resources and choices (as with the encoding of honorificity).

As expected, not all variants encountered in the corpus are equally prevalent. However, it is important not to ignore those that have a limited distribution, not only in the interest of descriptive accuracy, but also because they can be enlightening. In the case of our study, some of the more circumscribed instances of variation analyzed have turned out to be associated with groups of speakers characterized by significant language loss, which, in our interpretation, allows us to observe the impact of obsolescence on the structure of endangered languages and on the degree of variability of documentary data.

In this study, one such group was defined by the variable of age (the younger generations of SLP-speakers), while another was defined geographically (the SLP-speakers from the city of Jaffna). The fact that this is the first study of SLP to include data from Jaffna, and that the Jaffna variety has, in recent times, developed in conditions of isolation from the wider SLP-speaking community and of advanced language shift, resulted in new forms and new insights that complexify the overall description of modern SLP personal pronouns. However, it is important not to lose sight of the fact that much of the variation discussed here was observed not among the most isolated or least fluent sections of the speech community, but in the speech of highly fluent speakers. This

conveniently demonstrates the extent to which descriptive homogeneity is often a by-product of data limitations and draws attention to the fact that, if possible, any language documentation and description endeavor should aim to constitute a sociolinguistically-diverse and geographically-encompassing pool of consultants, regardless of the actual size or perceived homogeneity of the speech community.

Abbreviations

1 = first person; 2 = second person; 3 = third person; ABL = ablative; ACC = accusative; DAT = dative; DEM = demonstrative; DSLP = Documentation of Sri Lanka Portuguese; EXS = existential F = feminine; FOC = focus; GEN = genitive; HAB = habitual; HON = honorific; M = masculine; NEG = negative; NHON = non-honorific; OBL = oblique; OBLG = obligative; PFV = perfective; PL = plural; PRS = present; PST = past; PURP = purposive; QUOT = quotative; SG = singular; SLP = Sri Lanka Portuguese.

References

Aikhenvald, Alexandra Y. 2020. Language contact and endangered languages. In Anthony Grant (ed.), *The Oxford Handbook of Language Contact*, 241-260. Oxford: Oxford University Press.

Anon. 1826. *O Livro de oração commum e administração de os sacramentos, e outros ritos e ceremonias de a Igreja, conforme de o uso de a Igreja de Inglaterra: Juntamente com o Psalterio, ou Salmo de David*. London: G. Ellerton & J. Henderson.

Anon. 1851. *O evangelho conforme de Santo Mattheos. Traduzido ne Indo-Portugueza*. London: R. Clay.

Anon. 1863. *O livro de orações usado ne greyas de Wesleyanos ne Ilha de Ceylon*. Colombo: Officio Wesleyano.

Bakker, Peter. 2006. The Sri Lanka *Sprachbund*: The newcomers Portuguese and Malay. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic Areas: Convergence in Historical and Typological Perspective*, 135-159. New York: Palgrave.

Benveniste, Emile. 1971. *Problems in general linguistics*. Coral Gables, FL: University of Miami Press.

Berrenger. 1811. *A grammatical arrangement on the method of learning the corrupted Portuguese as spoken in India*. Colombo: Frans de Bruin.

Bhat, D. N. S. 2004. *Pronouns*. Oxford: Oxford University Press.

- Callaway, John. 1818. *A vocabulary; with useful phrases, and familiar dialogues; in the English, Portuguese, and Cingalese Languages*. Colombo: Wesleyan Mission Press.
- Callaway, John. 1820. *A vocabulary, in the Ceylon Portuguese, and English languages, with a series of familiar phrases*. Colombo: Wesleyan Mission Press.
- Callaway, John. 1823. *A Ceylon-Portuguese and English Dictionary*. Colombo: Wesleyan Mission Press.
- 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. 2007. Linguistic traces of colonial structure. In Eric Anchimbe (ed.), *Linguistic identity in postcolonial multilingual spaces*, 164-181. Newcastle: Cambridge Scholars Publishing.
- Cardoso, Hugo C. 2014. Factoring sociolinguistic variation into the history of Indo-Portuguese. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 5. 87-114.
- Cardoso, Hugo C. 2017. *Documentation of Sri Lanka Portuguese*. London: SOAS, Endangered Languages Archive. hdl.handle.net/2196/a542c4b1-8c36-4fd5-ae43-e777f87f5983. (26 July, 2019.)
- Cardoso, Hugo C., Mahesh Radhakrishnan, Patrícia Costa & Rui Pereira. 2019. Documenting modern Sri Lanka Portuguese In Pinharanda-Nunes, Mário & Hugo Cardoso (eds.), *Documentation and Maintenance of Contact Languages from South Asia to East Asia* (Language Documentation & Conservation Special Publication nr. 19), 1-33. Honolulu: University of Hawai'i Press.
- Cardoso, Hugo C., Tjerk Hagemeyer & Nélia Alexandre. 2015. Crioulos de base lexical portuguesa. In Maria Iliescu & Eugene Roegiest (eds.), *Manuel des Anthologies, Corpus et Textes Romans*, 670-692. Berlin: Mouton de Gruyter.
- Chandralal, Dileep. 2010. *Sinhala* (London Oriental and African Language Library 15). Amsterdam/Philadelphia: John Benjamins.
- Clements, J. Clancy. 1990. Deletion as an indicator of SVO → SOV shift. *Language Variation and Change* 2. 103-133.
- Dalgado, Sebastião Rodolfo. 1900. *Dialecto Indo-Português de Ceylão*. Lisbon: Imprensa Nacional.
- Fox, William Buckley. 1819. *A Dictionary in the Ceylon-Portuguese, Singhalese, and English Languages*. Colombo: Wesleyan Mission Press.
- Helmbrecht, Johannes. 2003. Politeness Distinctions in Second Person Pronouns. In Friedrich Lenz (ed.), *Deictic Conceptualization of Space, Time and Person*, 185-202. Amsterdam/Philadelphia: John Benjamins.

Ishiyama, Osamu. 2019. *Diachrony of Personal Pronouns in Japanese. A functional and cross-linguistic perspective* (Current Issues in Linguistic Theory 344). Amsterdam/Philadelphia: John Benjamins.

Lyons, John. 1977. *Semantics*, vol. 2. Cambridge: Cambridge University Press.

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.

McGilvray, Dennis. 1982. Dutch Burghers and Portuguese mechanics: Eurasian ethnicity in Sri Lanka. *Comparative Studies in Society and History* 24(2). 235-263.

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.

Nagy, Naomi. 2017. Documenting variation in (endangered) heritage languages: How and why? In Kristine A. Hildebrandt, Carmen Jany & Wilson Silva (eds.), *Documenting Variation in Endangered Languages* (Language Documentation & Conservation Special Publication 13), 33-64. Honolulu: University of Hawai'i Press.

Newstead, Robert. 1827. *Um curto catichismo da Biblia, disposto em quarenta divisaões, todas as repostas per as perguntas sendo ne as palavras de a Santas Escrituras*. Transl. of W.F. Lloyd. London: J.S. Hughes.

Newstead, Robert. 1852. *O Novo Testamento de Nossa Senhor e Salvador Jesus Christo*. Transl. of W.F. Lloyd. Colombo: Wesleyan Mission.

Newstead, Robert. 1871. *Cantigas per adoração publico, em lingua portugueza de Ceylon*. Colombo: Wesleyan Mission.

Nordhoff, Sebastian. 2013. The current state of Sri Lanka Portuguese. *Journal of Pidgin and Creole Languages* 28(2). 425-434.

Palosaari, Naomi & Lyle Campbell. 2011. Structural aspects of language endangerment. In Peter Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 100-119. Cambridge: Cambridge University Press.

Pereira, Rui. 2019. *Sri Lanka Portuguese: A sociolinguistic perspective*. Paper presented at the 19th Annual Conference of the Association of Portuguese and Spanish-lexified Creoles & Summer Conference of the Society of Pidgin and Creole Linguistics. Faculdade de Letras da Universidade de Lisboa, June 17th.

Sasse, Hans-Jürgen. 2001. Typological changes in language obsolescence. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language*

Typology and Language Universals; An International Handbook, vol. 2, 1668-1677. Berlin/New York: Walter de Gruyter.

Sippola, Eeva. 2018. Collecting and analysing creole data. In Wendy Ayres-Bennett & Janice Carruthers (eds.), *Manual of Romance Sociolinguistics*, 91-113. Berlin: De Gruyter Mouton.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

Siewierska, Anna. 2013. Gender Distinctions in Independent Personal Pronouns. In Matthew Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Schiffman, Harold F. 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.

Smith, Ian. (coll.). 1973. *Indo-Portuguese (Sri Lanka) (IRS01)*. PARADISEC. <http://catalog.paradisec.org.au/collections/IRS01> (27 July, 2019.)

Smith, Ian. 1977. *Sri Lanka Creole Portuguese phonology*. Ithaca, NY: Cornell University PhD dissertation.

Smith, Ian. 1979a. Convergence in South Asia: a creole example. *Lingua* 48- 193-222.

Smith, Ian. 1979b. Substrata vs. universals in the formation of Sri Lanka Portuguese. *Papers in Pidgin and Creole Linguistics* 2. 183-200.

Smith, Ian. 1984. The development of morphosyntax in Sri Lanka Portuguese. In Mark Sebba & Loreto Todd (eds.), *York papers in Linguistics*. York: University of York.

Smith, Ian R. 2013. Sri Lanka Portuguese. In Susanne M. Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.) *The Survey of Pidgin and Creole Languages*, Vol. II (*Portuguese-based, Spanish-based, and French-based Languages*), 111-121. Oxford: Oxford University Press.

Smith, Ian. 2016. The earliest grammars of Sri Lanka Portuguese. *Papia* 26(2). 237-281.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the World's Languages: A Comparative Handbook*, 55-75. Berlin: De Gruyter Mouton.

Tomás, Maria Isabel. 1992. *Os Crioulos Portugueses do Oriente; Uma Bibliografia*. Macau: Instituto Cultural de Macau.

Authors' addresses

Hugo C. Cardoso
Universidade de Lisboa
Faculdade de Letras
Alameda da Universidade
1600-214 Lisboa
Portugal

hcardoso@letras.ulisboa.pt

Patrícia Costa
Universidade de Lisboa
Faculdade de Letras
Alameda da Universidade
1600-214 Lisboa
Portugal

patriciacosta1@campus.ul.pt