

CHAPTER 5

The development of variable NP plural agreement in a restructured African variety of Portuguese

Alan N. Baxter
Universidade de Macau

This research examines the growth of variable plural (PL) agreement in the noun phrase (NP) of the restructured Portuguese of the Tongas, descendants of Africans contracted on the Monte Café plantation of São Tomé in the 19th and 20th centuries. I first provide a sketch of the sociohistorical background of the speech community, stressing that its Portuguese was acquired in a special contact situation along with simultaneous acquisition of Umbundu. Subsequently I present the results of a VARBRUL analysis of plural NP items based on data collected from three age groups, the oldest age group consisting of first- and second-generation speakers born on the plantation. Four conditioning factors are examined: (1) morphophonological saliency of the plural word; (2) the following phonological context; (3) the form class of the plural item and its position within the NP; and (4) whether informants had at least one African parent, or locally-born parents. Particular attention is paid to the development of PL marking within the NP structure. The left position adjacent to the head is found to be prominently marked, and is the point of entry of PL in the oldest age group. The analysis presented raises the possibility that an Umbundu transfer may underlie this configuration, and suggests that the growth of PL marking across the three age groups is closely linked to a growth in NP structure. The appearance of variable PL marking in L1 Tonga Portuguese is attributed to the past role of L2 Portuguese data in L1 acquisition.

1. Introduction*

This paper concerns the Portuguese of the Tongas of São Tomé (West Central Africa), the descendants of Africans indentured during the Portuguese colonial system. First described by Rougé (1992), this fascinating variety of Portuguese has much to contribute to the study of contact situations, especially those involving Portuguese. Whereas Standard Portuguese has rigid agreement rules applying to items within plural NPs, Tonga Portuguese (henceforth TP)¹ has variable plural agreement. Thus, examples such as (1), in which plural marking occurs on all items

in the NP, occur in variation with examples such as (2),² in which it is incomplete:

- (1) Os ôtros disse
 the[M]-PL other-M-PL say-PRT-3S³
 ‘The others said (so).’ H3DUA
- (2) issos coesa tudo.⁴
 this[NEU]-PL thing-F-S all[NEU]
 ‘all these things.’ H3DUA

This paper has two aims. Firstly, it sketches the socio-linguistic background of the Tongas, concentrating on the context for language acquisition. Secondly, it identifies the factors whereby variable plural agreement develops and spreads. This is addressed through a discussion of the profile of variable NP plural marking in the TP of the Monte Café plantation.

2. Background: indentured labor on são tomé plantations

TP has its origins in the substantial socioeconomic and demographic changes that swept São Tomé in the second half of the 19th century.⁵ The introduction of large-scale cacao and coffee planting coincided with the abolition of slavery on the island, and this event provoked a severe labor shortage. In 1854, slaves indigenous to São Tomé, were liberated.⁶ Foreign slaves were eventually freed in 1876 (Neves 1929: 55). Foreseeing these events, some planters began importing indentured labor from Angola in the early 1850s, (Nascimento & Gomes Dias 1989). In due course, they drew on English and French colonies and the Portuguese colonies of Guiné, Ajudá (in modern Benin)⁷ and, eventually, Mozambique and the Cape Verde islands. Part A of Table 1 (based on Neves (1929: 63) summarizes worker numbers and origins for 1876–1879, showing a clear Angolan prominence. Neves (1929: 63–4) also reports an Angolan majority of 7,419 in 1881, and 10,411 during 1885–1892. Part B of Table 1 shows that the Angolan prominence persisted into the early 20th century. The Mozambicans, however, began arriving only in 1901 (Neves 1929: 67).

Table 1. Indentured laborers entering São Tomé

A. Period of 1876–1879							
Angola	Guiné	Ajudá	Cameroon/ Gabon	Gold Coast	Sierra Leone	Liberia	TOTAL
5,249	1	24	3,229	738	39	516	10,341
B. Period of 1901–1928							
Angola	Mozambique	Cape Verde	Other origins				TOTAL
50,444	43,053	6,305	19				99,821

Two related facts are especially relevant to the origins of Tonga Portuguese. Firstly, prior to 1909, no Angolans were repatriated, although workers of other origins were (Cadbury 1910: 45–6).⁸ Secondly, while most workers were repatriated in the 20th century, their São Tomé-born offspring, the Tongas, became the *de facto* property of the plantations.⁹ After decolonization in 1974, many Tongas remained on the plantations where they were born.

2.1 Languages used on the plantations

The Tongas today speak their own variety of Portuguese and display varying degrees of maintenance of their (often koineized) African languages (Rougé 1992). Knowledge of São Tomé Creole Portuguese (henceforth Sãotomense) is not widespread among the Tongas (Rougé 1992: 173). This is especially true for those who reside on outlying plantations.¹⁰ Thus, among the informants interviewed during this project, only two younger speakers had some knowledge of Sãotomense. The Tongas' linguistic repertoire stems from their particular socio-linguistic history.

The language of communication between overseers and workers on the plantations was not Sãotomense. It was Portuguese (Rougé 1992: 173), presumably with elements of African languages.¹¹ Sãotomense appears to have been absent from the plantations for at least five interconnected reasons:

- (i) Following abolition in the 19th century, according to Mantero (1910: 17), former slaves simply abandoned the plantations *en masse*. Subsequently it appears to have been a widespread policy that the local Creole population was not encouraged to work on the plantations. This is noted in Cadbury (1910: 17–18, 26) and in the testimonies of our oldest informants. In the 20th century, until the late 1950s, most plantation owners seem to have chosen not to employ the locals, and even to have prevented their entry to the plantations.¹²
- (ii) Most overseers seem to have been European. This is reflected in the observations of Cadbury (1910: 30) and, for the Monte Café plantation, both in ANON (1895) and in the testimonies of elderly informants. During the 20th century, however, some overseers were Tongas.
- (iii) The huge numbers of indentured workers continued to speak their ancestral languages. The maintenance of African languages among their descendants, the Tongas, related to the nature of plantation administration. Workers were housed separately, and assigned to work-groups according to their origins.
- (iv) The hermetic nature of the plantation system did not favor outside contact. Until the mid-1940s, the indentured workers were largely confined to the plantation precincts. The worker's day was strictly regimented, beginning with a roll call, followed by supervised work, and ending with a curfew. Many plantations had their own shops, and until the 1940s, the workforce had only one half-day off per week. Indentured workers and Tongas who were associated with the administration of the plantation would have had greater access to

off-plantation contexts. Yet, the average indentured worker prior to the 1940s required a permit to leave the plantation area.¹³

- (v) Finally, geographical factors could disfavor off-plantation contacts. Such factors are relevant in the case of the Monte Café plantation, situated on the shoulder of a mountain in an area that was covered with forest until 1854 (Mello 1864). Even today Monte Café lies beyond the main Creole settlements.

Thus, on many plantations Sãotomense may not have been a significant factor for the development of Tonga Portuguese, unless (i) it was present during the establishment of the new plantations in the 19th century; or, (ii) a plantation had employed local Creoles at any subsequent stage. For the Monte Café plantation, neither possibility can be completely discounted. However, this plantation was founded in 1854, and its owner began to import Angolans in the same period. Although no mention of the use of local labor has yet been located in surviving documents, the possibility remains open.¹⁴ Subsequently, for the 20th century, to judge from plantation pay-sheets, and the testimonies of old Tongas, local Creoles were not employed as field laborers until the mid-1950s, and even then, they did not reside on the plantation.

If Sãotomense did not play a direct role in the development of Tonga Portuguese in the early stages of the plantation system, could the local São Tomé variety of Portuguese have exerted some influence on TP? São Tomé Portuguese is strongly influenced by Sãotomense and has features in common with TP (see Section 2.2.). Thus, Sãotomense could have indirectly influenced TP via local São Tomé Portuguese. Indeed, any communication between plantation workers and outsiders would have been more viable in Portuguese than in Sãotomense. A channel for such communication could have been contacts between outsiders and plantation workers who were more closely connected with administration. However, for many plantations in the pre-1940s period, the restrictions on outside contact are also relevant for contacts with local São Tomé Portuguese.¹⁵ After the 1940s, with greater possibilities of off-plantation movement of workers, and the presence of local workers on the plantations as of the mid-1950s, São Tomé Portuguese came into more frequent contact with Tonga Portuguese. Also in this period, Tonga children came to have general access to off-plantation education in Portuguese.¹⁶

2.2 The speech community on the monte café plantation

On Monte Café, Angolan workers (from the Benguela region) were imported from the early 1850s until the 1870s (Nascimento & Gomes Dias 1989: 52, 68), and from the 1880s (ANON 1885)¹⁷ into the 1890s again largely from the Benguela region (Curadoria 1895–1896).¹⁸ This strong Angolan presence was to continue in the first half of the 20th century, as is evident in Table 2.

For language acquisition, the consequences of this overwhelming African presence, and longstanding Angolan connection, were exceedingly complex. Table 2 indicates that, throughout this selection of monthly pay sheets for 1935–45, the over-

Table 2. Monte Café workforce 1935, 1944 paysheets

	Angola		Mozambique		Cape Verde		Ajudá		Tonga		Children	
	F	M	F	M	F	M	F	M	F	M	F	M
May 1935	174	37	234	3	1	1	0	2	52	59	53	43
April 1937	153	37	234	3	1	1	0	2	68	109	59	61
April 1939	291	36	208	3	0	1	0	2	65	107	60	54
July 1942	366	42	186	1	8	15	0	1	63	111	83	91
May 1945	197	24	168	1	14	14	0	1	55	97	88	91

Source: *Curadoria: Folhas de pagamento — roças* 'Office for welfare of indentured workers: paysheets — plantations', São Tomé and Príncipe Historical Archive.

whelming majority of the adults were continental Africans. It may be assumed that, alongside the African languages, a continuum of varieties of L2 Portuguese were current, as were L1 varieties of TP. Informants born in the early 1900s testify to several <home> linguistic contexts according to parental origin: <Angolan + Angolan>, <Angolan mother + other African father> (for example from Ajudá, Cape Verde or Mozambique),¹⁹ <Tonga + Angolan>, <Tonga mother + other African father> and <Tonga + Tonga>.

On the one hand, the first language acquired by the offspring of Angolan parents was a Bantu language, probably a variety of Umbundu. This language has been retained in a koineized form among the Monte Café Tongas. Portuguese was acquired mainly through contact with other adults and children. This occurred in three main scenarios: (i) in the field (as mothers generally took their infants to work); (ii) in the precinct of the administrative area where children conducted yard duties supervised by an indentured worker or a Tonga²⁰); and (iii) in the context of the housing complex. In each situation, children would have heard Tonga L2 and L1 Portuguese, and they would also have heard L1 European Portuguese. In the field and in the housing contexts they would have also heard and used Umbundu. In the plantation yard and housing contexts, children could also have heard and used Cape Verde Creole.

The children of African parents of *different* origins acquired the Angolan mother's language (which in some cases was the father's L2), and they acquired 'Portuguese' based on the L2 Portuguese spoken by the parents, and also through their exposure to Portuguese in the contexts beyond the immediate influence of the parents in the <home> context. None of the first generation Tongas interviewed acquired African languages other than Umbundu from their parents.

Finally, the offspring of Tonga couples, or of Tonga plus African couples, appear to have acquired both the Umbundu koiné and Portuguese simultaneously. In this case one principal source of primary linguistic data (henceforth PLD) was Tonga L1 Portuguese. However, just as with the children of African couples, the contexts

beyond direct parental influence must also have affected the acquisition of both Portuguese and Umbundu.

With regard to the other languages present among the workforce (from Ajudá, Mozambique and Cape Verde), it is interesting to note that the Tongas on Monte Café display only minimal evidence of acquisition of these languages. For the population that provided the data for the present study, only two informants of the middle age group claimed to have limited L2 knowledge of a Mozambican language (spoken in Quilimane), acquired as young men. Furthermore, no informant in the population sample studied here speaks Cape Verde Creole.

The historical context for the acquisition of Portuguese by the oldest of the three generations in the database to be analyzed may be summarized as follows:

- (i) the predominant model of Portuguese among the huge worker population was workforce Portuguese L2, which, considering the constant cyclic nature of the indenturing process (with constant new arrivals of African monolinguals contracted for three to five years), must have always existed in a continuum of interlanguage varieties;
- (ii) children of Generations 1 and 2 learned the predominant Umbundu-based koiné prior to learning Portuguese;
- (iii) children learned Portuguese in the field and in the plantation compound, mainly through contact with workers' L2 Portuguese, less so with workers' L1 Portuguese, and marginally through contact with European overseers.

The L1 Tonga Portuguese of first generation Monte Café Tongas displays substantial and consistent morphosyntactic restructuring in the NP (in referencing, number, and gender representation), in the VP (tense-aspect marking), and in the representation of case relations and negation. Those features of L1 TP display many similarities with the grammars of Creole Portuguese varieties (e.g., Cape Verde Creole Portuguese and São Tomé Creole Portuguese; see Baxter 2002). In other respects, especially in the leveling and development of agreement rules as variable rules, TP resembles the vernacular varieties of Portuguese of Brazil, Angola, Mozambique and Macao. Variable NP number agreement is also found in São Tomé Portuguese, especially among speakers from lower socioeconomic backgrounds. The following sections discuss the development of NP plural marking across three generations of Tonga Portuguese.

3. The data for the present study

Fieldwork in São Tomé was conducted during February-May, 1998. Three fieldworkers were involved: the author of this paper, co-researcher Dr. Dante Lucchesi (Universidade Federal da Bahia) and Nara Barreto (Universidade de Macau). Contact with the Tonga communities was established through the kind assistance of local Tonga politician Albano de Deus. Our Tonga field assistant at Monte Café, Fran-

cisco Paulino, conducted invaluable work with the interpretation of the interviews and transcription. Research assistants Joseli Querino (Universidade Federal da Bahia) and Dr. Norma Lopes (Universidade do Estado da Bahia) helped with the codification and initial analysis of the data.

At the Monte Café plantation, Labovian-type socio-linguistic interviews of 1.5 hours average were recorded with 20 Tongas in three age groups (Group 1 = 8 speakers) 61+ years,²¹ (Group 2 = 6 speakers) 41–60 years, and (Group 3 = 6 speakers) 20–40 years, equally distributed by sex. The interviews covered a range of topics relating to plantation life, including specific incidents in the history of the plantation, and the lives of the informants. In addition, a brief ethnographic questionnaire was administered. All informants live on Monte Café and most also work there.²² Only the informants in the youngest age group, and two from the middle age group, had education at the elementary level.²³ Henceforth, these three age groups are referred to as Generations 1 (60+ years), 2 (41–60 years) and 3 (20–40 years).

4. Methodology and preliminaries to analysis

As a vast body of research has been conducted on variable NP plural agreement in Brazilian Portuguese, it was decided to follow the analytic framework developed in that area. This facilitates substantial comparisons and, in many respects, avoids “re-inventing the wheel”. The current study builds on the works of Guy (1981), and especially on that of Scherre (1988) and of Lopes (2001), all of which were conducted within a variationist framework using the VARBRUL programs (Pintzuk 1988).

Following Lopes (2001), all NPs with plural reference were codified for the dependent variable <presence of the plural marker>, and for a range of independent factor groups: (i) saliency of the morphophonological form in terms of the type of plural formation and stress, (ii) the phonological context following the place of plural insertion, (iii) the presence of plural markers prior to the item being coded, (iv) the grammatical class of the pluralizing word, (v) the functional class of the pluralizing word and its position relative both to the head of the NP and to its linear position within the NP, (vi) age of speaker, (vii) sex of speaker, (viii) African parentage, and (ix) education.²⁴

The reader should note carefully the nature of the complex factor groups (i) and (v). On the one hand, factor group (i) <saliency> permits the analysis of all the different Portuguese morphophonological plural types, and all major stress patterns. My approach permits a wide range of analytical comparisons of the differential effect on plural marking according to the form of the word — for example, the effect of metaphony, or whether a singular word ends in a vowel or a consonant.²⁵ On the other hand, factor group (v) combines three qualities. It addresses the effect of linear position in the NP while simultaneously addressing the effect of the functional class of the pluralizing word (as NP head or non-head) and, for non-nuclear items, the effect of their position relative to the NP head. This factor group is thus strongly

syntactic in nature and facilitates a wide range of insights into plural marking, including the interaction of plural marking with bare NPs. Discussion of the rationale behind this type of combined factor group for analyzing Brazilian Portuguese data may be found in Scherre (1988, 1998) and Lopes (2002).²⁶

For the present paper, three factor groups were not examined. Firstly, <education> was not considered because it was categorical for the youngest age group. Secondly, in initial analyses the <grammatical class of the pluralizing word> was identified as not exerting a significant influence on the dependent variable <presence of plural marking>. Thus, it was excluded from the further analyses conducted here. The third factor group omitted is the <presence of plural markers prior to the item coded>. This factor group tested as exerting a significant effect on the plural marking of subsequent pluralizing items in the same NP. However, it is not discussed here because this paper concentrates on the effects relating to the immediate structural points at which plural marking is introduced.

5. Overall variability and the profile of development

The data comprise 3,366 NP items with plural reference. Of these, 1,835 items (55%) carry the plural morpheme, which is attached with an input probability of 0.55. The broad profile of the development of the NP plural agreement rule across three age groups is demonstrated in Table 3, showing the results of the analysis for the factor group <age of speaker>.²⁷

Given the socio-historical context of the speech community investigated, it is non-controversial to claim that the corpus collected represents the following development: Generation 1 is minimally disposed towards plural marking. Indeed, the oldest of the speakers within this group display an exceedingly low frequency of marking. It seems reasonable to assume that this reflects the nature of predominant PLD available to this generation (and prior generations) of language learners. That type of PLD would have been heavily influenced by L2 Portuguese acquired by African adults, and there was insufficient input to set NP number agreement

Table 3. The effect of <Age group> on PLURAL marking of NP items^a

Generation	Items with PL marking	Total of plural items	% PL marking	Probability of PL marking
1- (61+ years)	191	779	25%	0.11
2- (41-60 yrs)	524	1118	47%	0.42
3- (20-40 yrs)	1120	1469	76%	0.8

^a In this and subsequent tables, only cells with 10 or more data items were processed by VARBRUL.

^b Only levels of significance below .05 are acceptable.

Table 4. The effect of the <origin of the parents of Generations 1 and 2> on PLURAL marking of NP items

Origin of parents	Items with PL marking	Total of plural items	% PL marking	<i>pr</i> weighting of PL mark
Africa	275	914	30%	.36
Monte Café	440	983	45%	.63

Significance: $p = .029$
Log Likelihood = -611.141

(henceforth Agr). Successive generations, with exposure to PLD of a different nature, and with a growth in contact with L1 varieties of Portuguese as the plantation infrastructure expanded and the colonial society changed, display acquisition of Agr as a variable rule.

Additional support for the relevance of Adult L2 Portuguese in the PLD for Generation 1 is suggested by a further analysis of the data from Generations 1 and 2. Here, an additional factor group was included, classifying speakers according to whether their parents were African born (i.e., adult L2 learner-speakers of Portuguese) or born on Monte Café plantation (i.e., ostensibly L1 speakers of Portuguese, having acquired L2 before the onset of puberty). This factor group was selected by VARBRUL as exerting a significant effect on plural marking. The results are presented in Table 4.

Speakers whose parents were born on the African continent disfavor plural marking, with a probability weighting (henceforth *pr.*) of only 0.36, whereas speakers whose parents were born on the Monte Café plantation favor plural marking, with a *pr.* of 0.63.

As the situation depicted in Table 4 is clearly developmental, involving the acquisition of plural marking by this population, one expects to find in the data evidence of how and where the development of plural marking gradually spread in the grammar of each age group. The following sections discuss three linguistic variables which tested as exerting a significant effect on the use of the plural marker: the morphophonological saliency of plural marking, the phonological context following the plural item, and the relative positions and status of structural elements of the NP.

6. Saliency, and the following phonological context

The role played by saliency in contrasting forms has long been considered relevant to the orientation of variation in agreement rules in Brazilian Portuguese, and it figures prominently in studies of variable plural marking (Guy 1981, Scherre 1988). However, saliency has also been included in the study of acquisitional processes in depidginization (Emmerich 1984), decreolization (Bickerton 1975: 142), and the

areas of first language acquisition (Ingram 1989:437) and second language acquisition research (Bayley 1994:170–72, Wolfram 1985). The latter three cases consider the potentially facilitative effect of saliency in the acquisition process and in the association of form with function.

Saliency was contemplated in the present study both in relation to the phonetic distance between singular and plural forms and in relation to stress-class. The analysis considered six saliency factors relating to the phonetic nature of the plural distinction and stress patterns. Four factors consider the presence or absence of stress on the plural syllable together with an accompanying change in the root of the plural word. A fifth factor concerns unstressed monosyllables. A sixth factor caters for plurals whose corresponding singular ends in an *-s*.²⁸ Table 5 exemplifies these factors and presents the results of the analysis of this factor group in four separate VARBRUL runs: one for the entire Tonga data set, and one each for the three generations.²⁹

Generation 1 stands apart, with a very different pattern of variation than those of Generations 2 and 3. This is expected, as Generation 1 would have received a heavier input from L2 PLD, from adult African L1 speakers. Indeed, it is in generation 1 that one might expect to find reflexes of morphological leveling and of transfer phenomena, which are typical of early stage L2 acquisition by adults. The forebears of Generation 1 were predominantly of southern Angolan Bantu stock. Their L1s were agglutinating languages that indicate plural through prefixal classifiers, in a concordance spreading from the noun classifiers. In such languages, the word structure is predominantly paroxitonal (penultimate syllable stress), and they are open syllable dominant. So, when such speakers acquire Portuguese as L2 through contact, two effects are expected. Firstly, they are likely to opt for unmarked settings in the construction of interlanguage, and so will the subsequent generation in L1 acquisition. Secondly, they are expected to display some transfer effects that impede their acquisition of plural morphology. Table 5 clearly shows that the more salient oppositions in Portuguese play an important role in the development of plural morphology originating from a system that initially had a very limited use of plural marking.

6.1 Saliency factors

6.1.1 *Plurals attaching to a stressed final syllable*

In Generation 1, the entry point for plural marking is through words with a stressed final syllable where attachment of the plural does not involve a change in the root of the word (*pr.* 0.93), e.g. *cafés*.³⁰ This is an optimal perceptual context from an acquisitional perspective. At the same time, the presence of this salient evidence in L2 PLD from adults also makes sense. The predominant L1 of the adults, Umbundu, is strongly penultimate-syllable stressed (Valente 1964:27), so Portuguese words with final syllable stress should stand out for the learner as being markedly different.

In the transition from Generation 1 (*pr.* 0.93) to Generation 2 (*pr.* 0.57), the effect of final-syllable stress in words whose plural does not require a root change is neu-

Table 5. The effect of <saliency of plural formation> on plural marking in the NP

Type of plural	Generation 1			Generation 2			Generation 3			All three generations		
	N PL	% PL	pr: PL	N PL	% PL	pr: PL	N PL	% PL	pr: PL	N PL	% PL	pr: PL
1. Stressed PL + root change - Singular in -#; <i>fácil</i> 'easy (S)'; <i>faceis</i> 'easy (PL)' - Singular in irregular -ão#; <i>caminhão</i> 'truck'; <i>caminhões</i> 'trucks'	0/5	0%	.	22/25	88%	0.93	25/27	93%	0.84	47/57	82%	0.88
2. Stressed PL + no root change <i>café</i> 'coffee'; <i>caféis</i> 'coffees' <i>irmão</i> 'brother'; <i>irmãos</i> 'brothers'	24/111	22%	0.93	50/84	59%	0.57	74/93	80%	0.45	148/288	51%	0.52
3. Monosyllable <i>o</i> 'the (S)'; <i>os</i> 'the (PL)'	103/122	84%	0.44	191/194	98%	0.82	317/319	99%	0.79	611/635	96%	0.61
4. Unstressed PL + root change - Metaphoric plural [ô]vo 'egg'; [ô]vos 'eggs'	0/2	0%	.	7/14	50%	0.51	10/13	77%	0.66	17/29	59%	0.59
5. Unstressed PL + no root change - Word with penultimate syllable stress: <i>casa</i> 'house'; <i>casas</i> 'houses' - Word ending in -r: <i>jantar</i> 'dinner'; <i>jantares</i> 'dinners' - Word with antepenultimate syllable stressed: <i>indígena</i> 'native'; <i>indígenas</i> 'natives'	63/465	14%	0.40	244/753	32%	0.39	653/932	70%	0.40	960/2150	45%	0.46
6. Singular in -s <i>país</i> 'country'; <i>países</i> 'countries'	1/73	1%	0.26	10/48	21%	0.35	41/85	48%	0.28	52/207	25%	0.35

tralized; and in Generation 3 (*pr.* 0.45), this factor is slightly unfavorable to marking. In the transition from Generations 2 through 3, as other saliency classes spread plural marking, the dominant saliency role is assumed by words which have final syllable stress and which require a root change when plural is attached.

6.1.2 *Monosyllables*

Monosyllabic words are slightly unfavorable to plural attachment in generation 1 (*pr.* 0.44). Why? Two factors are plausibly involved here. Firstly, there is a tendency in generation 1 to omit monosyllabic determiners or to replace them with innovative forms. Definite reference NPs are common without monosyllabic articles, as in example (3):

- (3) Sabe inda dia dumingo elé lanja cumida
 know-PRS-3s still day[M]-s Sunday[M]-s he arrange-PRS-3s food[F]-s
 pa muiere veio.
 for women[F]-PL old-M-S
 ‘You know, even on Sundays he used to get food for (the) old Tonga
 women (i.e., here on Monte Café).’ M3ALB2-569

Furthermore, in Generation 1, the function of definite article can be performed by TP, *isso* ‘this, that (with specific reference [masculine or feminine gender], or with indefinite global reference).’³¹ This is a role that the definite articles *o, a* ‘the’ (respectively, masculine and feminine singular), and *os, as* ‘the’ (respectively, masculine and feminine plural) would perform in Portuguese. Example (4) demonstrates this innovation:

- (4) Mas teve muita confusão aqui?³²{Res.}
 but exist-PRT-3s much-F-s confusion-F-s here.Q
 ‘But was there much confusion here?’
- Muito! Gente no trabaia. Se você tá
 much-M-S.EX people/we[F] NEG work-PRS-3s if you be-PRS-3s
 trabaia venha batê você. {Inf.}
 work-INF come-SUBJ hit you
 ‘A lot! People/we didn’t work. If you were working (they) came (and) beat
 you.’
- Hum. {Res.}
 EX
 ‘Hmm.’
- Isso foro venha batê você. {Inf.}
 this[NEU]-S creole[M]-s come-SU-3s hit-INF you-S
 ‘The forros (i.e., the Sãotomenses) came (would come) and beat
 you.’ M3sIL300-3

Also, in place of Portuguese <preposition+article+number> inflecting compounds, such as *nos/nas* 'in/to the' (< P. *em+o/a+s* [= to/in+the (masculine gender)/the (feminine gender)+plural]), one finds (variably) uninflecting *ni* (or *na*) as a locative/directional (to or from) preposition occurring with NPs without definite articles,³³ as in example (5):

- (5) Oto branco landja fora, ni foro, pa ficare
 other-M-S white-M-S arrange-PRS-3S outside in foro-M-S to stay/be-INF
 cu ele.
 with he
 'Other Europeans arranged (them) outside, among the *forros*, to stay with
 them.' M3SIL1102

Compounds such as *dos/das* alternate with the use of uncontracted *de* introducing NPs without articles.

- (6) Quando contro isso mulato, mulato, mulato,
 when find-PRS-1S³⁴ this-NEU-S mulatto-M-S mulatto-M-S mulatto-M-S
 mulato, é di branco que tinha aqui.
 mulatto-M-S be-PRS-3S of white-M-S that have-PST.IMP-3S here
 'When you find these mulattoes, they are from the Europeans who were
 here.' M3SIL1096

However, a second motive for the slightly unfavorable effect of monosyllables on plural marking might be sought in the adult L2 Portuguese influenced PLD for generation 1: Umbundu has few monosyllables, and functionally, those that are not agglutinating bases tend to be prepositions or adverbs.

In contrast, in Generations 2 (*pr.* 0.82) and 3 (*pr.* 0.79), the monosyllable class assumes notable importance as a significant factor for plural attachment. The explanation would appear to lie in the broad difference in plural marking strategies between Generation 1 and the other two generations. The factors just mentioned, the reliance on alternative article/determiner forms, such as *isso* and *uma*, and the non-agglutinated alternatives to agglutinated Portuguese <preposition + article> forms, are more characteristic of generation 1. The difference in Generation 2 might be interpreted as a reappraisal of the ability of monosyllables to be plural marked as the definite article of Portuguese is acquired.

6.1.3 Unstressed plural syllable + a change in the root

Metaphonic plurals yielded little data and no PL attachment in Generation 1, yet they display acquisitional development from Generation 2—where they are only marginally favorable to plural attachment (*pr.* 0.51)—to Generation 3, where they are reasonably favorable in relative weighting (*pr.* 0.66).

6.1.4 *Unstressed plural syllable + no change in the root*

Across the three generations, this factor is consistently unreceptive to the plural morpheme, with *pr.* weightings of 0.40, 0.39 and 0.40, respectively. Recall that there are four word types involved here: words with penultimate syllable stress, words ending in *-r* and *-l*, and words stressed on the antepenultimate syllable.

In Portuguese, words with penultimate-syllable stress constitute the largest stress-class and the least salient opposition phonetically. In the data base considered here, this class of words accounted for over 93% of the factor <unstressed plural syllable + no change in the root>. It might be argued that the saliency status of this class of word, at least for Generation 1, is further compromised by the fact that Umbundu words are stressed on the penultimate syllable and have open syllables.

Where words ending in *-r* and *-l* are concerned, generation 1 registered no cases of PL attachment. Variable PL attachment occurs only in Generations 2 and 3, with low data numbers. Some speakers, especially in Generation 1, show evidence of treating superstrate words ending in *-r* as ending in a vowel, so that such words are likely to be treated as having penultimate syllable stress. Finally, words with antepenultimate stress registered no data in generation 1, and variation in Generations 2 and 3 with low data numbers.

6.1.5 *Singulars in -s*

Words ending in *-s* are quite disfavorable to plural attachment in the three generations. This negative effect is strongest in Generation 1 (*pr.* 0.26), somewhat weaker in Generation 2 (*pr.* 0.35), and again slightly stronger in Generation 3 (*pr.* 0.28). From an acquisitional perspective, the fact that <*s*> disfavors attachment of plural morphology may be interpreted either as an analogical factor, the speaker associating a final *-s* with plurality, or to haplology.

6.1.6 *Overview of saliency*

The effect of saliency on PL attachment over the three generations is reminiscent of the findings of acquisitional studies, where a growth of inflectional morphology has been observed to initially involve salient oppositions (Bayley 1984: 170–2). The least salient oppositions are acquired last. Consideration of the VARBRUL run of the three amalgamated generational data sets gives an overall view of the relative strengths of the different saliency factors for this speech community. The effect of saliency in acquisitional terms is appreciable. On the one hand, agreement is favored by the morphophonological distinctions of the more marked oppositions, the combination of a stressed pluralizing syllable and a root change associated with PL being the most potent (*pr.* 0.88). Next in strength come the monosyllables (*pr.* 0.61), reflecting the acquisition of the article in Generations 2 and 3, followed by metaphonic plurals (*pr.* 0.59) and words with a stressed pluralizing syllable yet no root change (*pr.* 0.52). On the other hand, it is disfavored by the least salient classes — words with no root change and an unstressed pluralizing syllable (*pr.* 0.46), and words with a singular in *-s* (*pr.* 0.35), the latter severely compromised by their final consonant.

In sum, in analyzing saliency for this population of speakers, one would expect to find some evidence of transfer phenomena, and especially evidence of acquisition guided by saliency. Indeed, this is the case. The most salient oppositions are acquired earlier, whereas the least salient are acquired later, and this process is reflected in the speech community's profile of variation (in the analysis of the data of the three generations together).

7. Following phonological context

Another dimension concerning phonological constraints on the development of plural marking, and on the pattern of variation observed in this community, relates to the phonological context following the plural item. Table 6 (p. 114) shows the relative probability weights of the effect of the <following phonological context> on the dependent variable.

Beginning with Generation 1, it can be said that several of the phonological constraints might be explained in terms of acquisitional tendencies and transfer influence in the L2 Portuguese of African adults who provided PLD to Generation 1. Here, the strongest point of entry for plural attachment is when the plural item is in the final position of the NP voicing group (*pr.* 0.69). The next most propitious contexts are: a following voiceless consonant (*pr.* 0.66), and a following pause within the NP (*pr.* 0.58). On the other hand, plural attachment is strongly inhibited by a following voiced consonant (*pr.* 0.21), although a following a vowel exerts an almost neutral effect (*pr.* 0.46). The cause may be the voicing rule in Portuguese which applies to /s/ prior to voiced consonants and vowels, which Umbundu speakers would have heard from European Portuguese overseers.

Part of this profile may be accounted for by adult L2 input to PLD, because morphological leveling has occurred, and the L1 of these adults displays a strong open syllable phonotactic typology, thus imposing strict constraints on sequences of consonants. Thus, a pause after a plural marker would constitute a perceptually salient phonological context for that marker, especially a pause at the end of the NP voice group. The fact that a following voiced consonant or a vowel strongly inhibits marking, may have some basis in Umbundu phonotactics. In Umbundu, there are no consonant sequences, but rather prenasalized stops. Furthermore, Umbundu does not have [z] in its sound inventory (Valente 1964: 383–4). This may have yielded a restriction on the sequence PLURAL + VOICED CONSONANT OR VOWEL in the Portuguese interlanguage of L1 Umbundu speaking adults. In acquiring L2 Portuguese, Umbundu speakers would experience more difficulty in inserting /s/ before a voiced consonant or before a vowel. On the other hand, the production of a sequence where the second item is unvoiced appears to have been unproblematic and was evidently strong in PLD input to Generation 1.

Table 6. The effect of the <following phonological context> on the use of the PLURAL morpheme in plural NP items

Following phonological context	Generation 1		Generation 2		Generation 3		All three generations	
	N PL	% PL	N PL	% PL	N PL	% PL	N PL	% PL
1. Voiceless consonant <i>pra mias tias</i> 'for my aunts'	98/209	47%	209/324	65%	424/493	86%	731/1026	71%
2. Voiced consonant <i>as loica</i> 'the crockery'	57/194	29%	118/247	48%	270/338	80%	445/779	57%
3. Vowel <i>meas avós</i> 'my grandparents'	23/119	19%	102/194	53%	202/290	80%	327/603	54%
4. Structure-final pause <i>dezesseis anos</i> 'sixteen years'	8/130	6%	58/180	32%	101/149	68%	167/459	36%
5. Structure-internal pause <i>pra mias tias, irmãs dele</i> 'for my aunts, his sisters'	4/122	3%	36/167	22%	120/192	63%	160/481	33%

Significance = .000
Log Likelihood = -137.876

Significance = .005
Log Likelihood = -413.246

Significance = .001
Log Likelihood = -576.261

Significance = .000
Log Likelihood = -1434.972

7.1 The development of PL marking and its overall profile

Several factors operating in Generation 1 carry over to the two subsequent generations — three factors favoring marking, and one strongly disfavoring it. The final pause context continues to be the prominent favorable context in Generations 2 (*pr.* 0.64) and 3 (*pr.* 0.60), although decreasing slightly in the latter. The slightly favorable weighting of the internal pause weakens in Generations 2 (*pr.* 0.53) and 3 (*pr.* 0.54). The context of a following voiceless consonant, the second most propitious context for marking in generation 1, shows a similar decline to just below neutral effect in generation 2 (*pr.* 0.49), and then increases in generation 3 (*pr.* 0.55) to be only slightly favorable to marking. Next, the strongly unfavorable factor of Generation 1, a following voiced consonant, continues to disfavor marking, although it gradually weakens across the three age groups (with *pr.* weightings of 0.21, 0.39 and 0.45 respectively). In contrast, the following vowel context holds a surprise: although it decreases only slightly in weight from Generation 1 (*pr.* 0.46) to generation 2 (*pr.* 0.50), as a neutral effect on marking, in generation 3 (*pr.* 0.41) it has become an unfavorable factor.

8. The structural configuration of the NP

To assess the effect of the structural configuration of the NP on plural attachment, I follow the approach adopted by Scherre (1988, 1998) and Lopes (2001), who both build on the earlier work of Guy (1981). These researchers propose a factor group that contemplates the form class, position relative to the head, and linear position. In the present study, the data were codified for eleven configurational patterns. These are exemplified with Tonga data as follows:³⁵

(7) *Plural item at the left, adjacent head*

Os meus AVÓS nasceram aqui.
 the[M]-PL my-M-PL grandparent-f-pl born-PRT-3PL here
 ‘My grandparents were born here.’ H1JOA

(8) *Plural item at the left, not adjacent head*

Os meu PAI vivia.
 the[M]-PL my-M-s father-M-s live-IMP-3s
 ‘My parents were alive.’ H2ADR

(9) *Plural item is head in first position in NP*

OLHOS abertos assim.
 eye[M]-PL open-M-PL thus
 ‘Eyes open like this.’ H1FRANJU

- (10) *Plural item is head in second position in NP*
 os HOMENS grande.
 the[M]-PL man[M]-PL big-s
 ‘the big men.’ H2JUST
- (11) *Plural item is head in third position in NP*
 os meus VELHOS.
 the.M-PL my-M-PL old-M-PL
 ‘my parents (lit. oldies).’ M2FAT
- (12) *Plural item is head in fourth position in NP*
 os mos primeiros IRMÃOS
 the[M]-PL my-M-PL first-M-PL sibling-M-PL
 ‘my first brothers and sisters.’ H1FRA
- (13) *Plural item is head in fifth position in NP*
 seus nove ou dez ANOS [ININT]
 3POS-M-PL nine or ten year[M]-PL ININT
 ‘her (or his) nine or ten years.’ H1JOA
- (14) *Plural item right of head, second position in NP*
 ELES próprios
 3P-M-PL self-M-PL
 ‘they themselves.’ H1FRA
- (15) *Plural item at right of head, third position in NP*
 COISAS assim desagradáveis
 thing.F-PL thus unpleasant-PL
 ‘unpleasant things like that.’ M1LUC
- (16) *Plural item at right of head, fourth position in NP*
 essas COISA assim naturais
 this-F-PL thing.F-s thus natural-PL
 ‘natural things like that.’ H1JOA
- (17) *Plural item at right of head, fifth position in NP*
 cento e vinte CONTOS³⁶ mensais
 hundred and twenty conto.M-PL monthly-PL
 ‘one hundred and twenty contos monthly.’ H1JOA

Table 7 presents the findings of two analyses of the above configurations in Tonga Portuguese: VARBRUL run A and VARBRUL run B.

In VARBRUL run A, the 11 original configurations (of examples (7) through (17) above) were assigned to seven factor groups. This reduction was motivated by small data numbers representing the 4th and 5th positions in the NP, both for the head and the post-head items. Subsequently, in VARBRUL run B, the seven configurations factors were reduced to five. The latter reduction was undertaken because

low data numbers in Generations 1 and 2 prevented the detailed optional analysis of items to the right of the head. As can be seen in Table 7, comparing VARBRUL runs A and B for each generation, the marginally lower (i.e., better) *Significance* and *Log Likelihood* results in run B, in contrast with run A, for Generations 1, 2 and 3, confirm that the model with five factors provides a better fit of the data for those generations. On the other hand, for the three generations combined, the better *Significance* and *Log Likelihood* results in VARBRUL run A, in contrast with run B, reveal the seven factor model as providing a better fit of the data of the community as a whole. Indeed, as will be seen in the ensuing discussion, VARBRUL run A permits the detection of the differential effect on PL-marking of certain structural facts relating to head position (in Generations 2 and 3) and post-head items, (in Generation 3) that are not visible in run B.

What then do these two analyses demonstrate regarding the effect of the factor group <structural configuration> on PL marking? Overall, they demonstrate that there are different grammars for each of the three Tonga generations, and that certain tendencies introduced in Generation 1 are preserved in later generations.

In Generation 1, PL marking is initially established on the left position adjacent to the head (run A: *pr.* 0.95; run B: *pr.* 0.97). The head itself, when occurring in second position (run A: *pr.* 0.29) or in second or subsequent positions, (run B: *pr.* 0.24) disfavors plural attachment. Post-head positioning (which for Generation 1 mainly involves the 3rd position in the NP) almost categorically disfavors agreement (run A: *pr.* 0.01; run B: *pr.* 0.03). Note that the NP structure is minimal in Generation 1, mainly consisting of one pre-head position,³⁷ preferably PL-marked, a head which is seldom PL-marked, and one post-head position practically unmarked for PL. The basic PL-marking rule of Generation 1 requires PL-marking to be placed in the pre-head position of the NP.

Subsequently, in Generation 2, the NP is structurally more developed, allowing multiple pre-head items (witness run A, factors 2 and 5). The left position adjacent to the head (run A: *pr.* 0.89; run B: *pr.* 0.90) continues as the predominant factor favoring PL marking. However, PL has now spread to the other left positions, where it is also favored, although with a lower weighting (run A: *pr.* 0.72; run B: *pr.* 0.75). The head registers unfavorable results for second, third and fourth positions, (run A: *pr.* 0.30, for 2nd position, *pr.* 0.22 for 3rd, 4th or 5th position; run B: *pr.* 0.28, for 2nd, 3rd, 4th or 5th position). However, the results for run A suggest an incipient distinction between 2nd position heads and heads occurring later in the NP. Post-head positions, nevertheless, are still most unlikely to be PL-marked. In sum, the PL-marking rule of Generation 2 requires PL-marking to be placed (i) on the pre-head position, and (ii) on earlier positions more than later positions.

Finally, in Generation 3, the NP structure is even more complex. It is still the left position adjacent to the head (*pr.* 0.83, runs A and B) that is the most propitious for plural attachment. The next most favorable factor is still the non-adjacent left positions (*pr.* 0.70, runs A and B). Following these, the presence of the head in NP initial position (*pr.* 0.71, runs A and B) is the next strongest factor, attesting the effect of

the principle of preferential marking of earlier positioned items. The positioning of the head in second position (run A, *pr.* 0.27) or in second and subsequent positions (run A, *pr.* 0.26) is similarly unfavorable for PL marking, and the incipient differentiation of these two positions, noted in generation 2 (run A) is neutralized. Indeed, the grammar of Generation 3 would be better described by merging these two positional factors, as was done in VARBRUL run B (new factor 4). However, in run A, a differentiation is evident between the degree of unfavorability of PL-marking of the adjacent post-head position (run A, *pr.* 0.22) and subsequent post-head positions (run A, *pr.* 0.09), the latter being extremely unlikely to bear PL-marking. As a consequence of this further structural growth of the NP, the rules for PL-marking in generation 3 are also more complex than for the previous generations. Here, PL-marking is to be assigned (i) to the left in the NP, (ii) more on the pre-head structure than on the head, and (iii) more on earlier positions.

What is especially clear in the above discussion is the prominent role of the left position adjacent to the head. For Generation 1, the left adjacent position seems to play a quantifying role, bracketing the head. Why should this position be the point of entry of plural morphology? And why should it remain a prominent feature?

Three answers to the first question are plausible here, in connection with the type of PLD available to generation 1:

- (i) the evidence for plural marking was so diffuse in PLD that what might be considered to be an unmarked solution was adopted: quantify from the immediate left of the head (which happens to be first position in the minimal NP for generation 1);
- (ii) the PLD contained data from Portuguese L2 with evidence for this same structure coming from simple Portuguese L1 structures (e.g., DEMONSTRATIVE+NOUN); or
- (iii) the PLD contained data from Portuguese L2 with strong L1 transfer effects.

At Monte Café plantation, there are no surviving L2 Tonga Portuguese speakers, so the above three factors cannot be assessed directly. Explanation (ii) might be an option as an initial trigger, but the ratio of Africans to Europeans was very high. Interviews conducted with three European Portuguese, former administrators (including a former *feitor* 'foreman')³⁸ suggest that there was some simplification on the part of the Europeans, in the sense of using more analytic forms (e.g., subject pronouns as object pronouns). Nevertheless, the opinion of these former administrators was that the Tongas were the main originators of divergent forms. Of course, the information supplied by these men does not eliminate the possibility that input at an earlier stage may have contained variation originating in varieties of European Portuguese. Nevertheless, there are good reasons to believe that the main contribution to the development of variable plural agreement has other sources.

The natural developmental argument of explanation (i), involving unmarked settings triggered by a lack of clear evidence, seems plausible. In this scenario, the learner constructs minimal NPs with a quantifying item to the left. However, the

demographics of Monte Café until the 1940s, and some of the phonological considerations raised earlier in this paper, suggest that transfer (explanation (iii)) must be taken into account. Generation 1 is the group that received strongest PLD from the Portuguese L2 continuum, this group having a high proportion of L2 Portuguese speaking parents. L1 influence on L2 continua or interlanguage is well documented in acquisition research, especially for rudimentary stages of L2 development. Certainly the demographic facts of the Monte Café plantation appear to point very clearly to a situation that would tend to maintain a strong presence of early stage L2 Portuguese.

Guy (1981: 300–1) proposed the hypothesis that, in Brazilian Vernacular Portuguese, variable PL-marking, assigned from the left of the NP, could well have its origins in transfer from Bantu-speaking slaves, via noun classifiers and their role in PL assignment in Bantu languages. This hypothesis is worth contemplating with respect to Tonga Portuguese. Umbundu, as other Bantu languages, shows plural on nouns through a prefixal nominal classifier which signals a semantic class membership, and which may also carry deictic functions. Umbundu has a prefixal concordance system that is determined by the classifier of the noun, and extends to adjectives, determiners, and some connectives. Furthermore, the plural indication in the noun classifier is the key morpheme for the assignment of plural classifiers to other NP items. An adult Umbundu speaker in rudimentary stages of Portuguese acquisition could be expected to be more sensitive to plural indication in an adjacent prenominal position. A noun final position would be a rather unlikely place for such a speaker to perceive a plural marker or to attempt to place one. Also, with certain disyllabic classifiers, the second syllable of the classifier may show the plural contrast, e.g., *o-mbwa* ‘(the/a) dog’ vs. *olo-mbwa* ‘(the-PL) dog’; *ovi-ndele* ‘(the/a) white (person)’ vs. *otu-ndele* ‘(the-PL) white (person)’. This could further draw the learner’s attention to the immediate prenominal position. There is some corroboratory evidence in generation 1 Tonga Portuguese that certain Portuguese prenominal determiners might have been heard as classifiers. In the following example, the variable presence of *u-* on the word *uhome* (< P. *o homem* ‘the man’) suggests that the Portuguese definite article may have been reinterpreted phonologically as the classifier of Umbundu class 2 nouns, pertaining to persons, and prefixed with *u-* prior to word initial *-o*.

- ©(18) Agora isso uhome dexô massi³⁹ lá.
 now this[NEU] man[M]-s leave-PRT-3S machete[M]-s there
 ‘Now, the man left a/the machete there.’ F3s-1694

In her work on structural mixing, Myers-Scotton (1997, 2000a, 2000b, 2001) has demonstrated (with extensive data from bilingual aphasia, L2 acquisition, and bilingual code switching) how content morphemes from one language can be slotted into the morphosyntactic frames of another language. Furthermore, when mixing occurs, the syntactic frame acting as a matrix may contain *early system morphemes* — morphemes which are intrinsic to the meaning of the content morpheme slot and

which are generated together with that slot.⁴⁰ This sort of transfer is very common in the initial stages of L2 acquisition. Myers-Scotton's perspective seems relevant here. In other words, in Generation 1, it may well be that the predominant pattern observed relates to the African L1 structure that was strongly evident in L2 PLD available to that generation. From this perspective, transfer in rudimentary adult L2 acquisition would, in part, involve a degree of morphosyntactic frame fabrication involving L1 frames and L2 content morphemes, especially at an early stage in an L2 acquisition situation where L1 is strongly present. In this case, an Umbundu speaking adult might easily associate Portuguese determiners with Umbundu classifiers.

Following this approach, the prominence of the left adjacent position in Generation 1, even if it had a non-transfer source, say as an unmarked developmental solution as mentioned earlier, could be reinforced by the insertion of Portuguese content morphemes within Umbundu nominal frames in rudimentary adult L2, transmitted to generation 1 speakers via PLD. Such nominal frames would require their plural marking to their immediate left, where a classifier would appear in Bantu languages. Subsequently, in Generations 2 and 3, this reinforced trait has been assimilated into the developing Tonga Portuguese NP structure. In the grammar of Generation 3 speakers, who are Portuguese dominant Portuguese-Umbundu bilinguals, the left adjacent position now assumes a characteristic pattern of variable plural marking in an extensively developed NP.

9. Conclusion

The profile of variable plural marking in Tonga Portuguese is the result of cross-generational development. My analysis demonstrates that this change was conditioned by factors rooted in the L2 and L1 acquisition of Portuguese in a contact situation involving a strong Portuguese L2 component already influenced by Umbundu, and simultaneous acquisition of Umbundu. The latter two languages were constantly "renewed" through the indentured labor system. In these circumstances, the role of continuity of possible variation in plural marking in varieties of European Portuguese (Naro & Scherre 2000) would appear to be of minor or null consequence.

Language acquisition in the contact context was the trigger for the development of variable agreement rules. The sources of certain details of this re-configuration, such as the predominant adjacent pronominal position of plural marking, could plausibly result from the choice of an unmarked structure in the face of diffuse models, and the influence of minimal Portuguese L1 NP structures. However, the influence of African L1 transfer in this particular socio-historical context appears to have played a significant role directly and indirectly through Portuguese L2 PLD for L1 acquisition of Portuguese. The findings of this paper lend strong support to Guy's (1981: 300–1) hypothesis that Bantu substrate is a potential source of variable PL-marking, assigned from the left of the NP, in Brazilian Vernacular Portuguese.

Notes

* This paper was prepared as part of a research project on Tonga Portuguese supported by Australian Research Council Grant A59803382. I am grateful to two anonymous reviewers and to John Singler for their helpful criticisms and comments. Special thanks go to Gregory Guy for extensive discussion of the analysis of saliency and noun phrase configuration.

1. The Tongas, according to the traditional Colonial Portuguese use of the term, referred to the descendants of continental African indentured workers.
2. TP also displays variation in subject-verb agreement and gender agreement of NP items (Baxter 2002).
3. The morpheme glosses correspond to the grammatical forms of Standard Portuguese. The following abbreviations are used in this article:

[]	lexical gender	L1A	first language acquisition
{Inf.}	informant	L2	second language
{Res.}	researcher	L2A	second language acquisition
3	third person	M	masculine gender
Agr.	agreement	-	morpheme boundary
AHSTP	<i>Arquivo Histórico de São Tomé e Príncipe</i> 'the Historical Archives of São Tomé and Príncipe'	NEU	neutral gender
		PL	plural
		POS	possessive
		pr.	probability weighting
EX	exclamation	PRS	present indicative tense
F	feminine gender	PRT	simple perfective indicative
IMP	preterite imperfect indicative		preterite
INF	infinitive	Q	question
ININT	unintelligible word in transcription	S	singular
		SU	subjunctive
L1	first language	.	links EX intonation to word

4. Note that TP often uses the neutral gender form *tudo* 'all' where Standard Portuguese requires the gender- and plural-inflecting forms *todas* 'all (feminine gender plural)' or *todos* 'all (masculine gender plural)'.
5. The discussion in Sections 2, 2.1 and 2.3 supersedes the account in Baxter (2002).
6. São Tomé has two indigenous Portuguese-based creole languages: São Tomé Creole Portuguese (Sãotomense) and Angolar, both of which developed in the 16th century. Angolar is a maroon creole which split off from São Tomé Creole Portuguese in the 16th century (Lorenzino 1998).
7. The minuscule territory of Ajudá (Ouidah), or São João Baptista de Ajudá, was under Portuguese control until 1961.
8. Until the 20th century, the contractual conditions often amounted to a continuation of slavery. An official repatriation system was instituted in 1908. Prior to this, among the Angolans on São Tomé were people who had been sold into slavery or simply deported for petty crimes (Swan 1909). These facts were confirmed in the testimonies of elderly Tongas. For a general overview of the São Tomé labor scandal of the late 19th and early 20th centuries, see Hodges & Newitt (1988).

9. According to the census cited by Tenreiro (1961: 191), by 1950 there were 2,696 Tongas in the province of São Tomé and Príncipe. This number seems small, and perhaps only refers to Tonga adults in the workforce.
10. Lorenzino (1998: 48) refers to Tongas who are sixty and above who are bilingual in Umbundu and Sãotomense. It would seem likely that these are Tongas from plantations close to Creole communities.
11. My oldest informants testify to the fact that some European overseers acquired a minimal knowledge of Umbundu. This type of situation probably also existed in earlier periods.
12. These comments are based on the testimonies of elderly Tongas at the Monte Café, Agostinho Neto and Praia das Conchas plantations, including three Tongas who had been overseers.
13. These circumstances were reported by the oldest Tongas interviewed at the Monte Café, Agostinho Neto and Praia das Conchas plantations, including Tongas who were former workforce overseers.
14. Two linguistic features of TP suggest an early connection with Sãotomense: *ami* 'I', and *ni* 'to, at, in, from'. However, the former is also used in the Umbundu koiné and is a likely Bantu transfer in TP. See note 32 for further discussion of *ni*.
15. Plantations that relied on road transport to the public docks, in contrast with those that had their own narrow gauge railways and jetties, would have required frequent off-plantation presence of numbers of workers (Cadbury 1910: 28–9). Monte Café was not in the former category, as it had its own railway.
16. The language of education is the European Standard, represented in colonial times mainly by European teachers. Broadly speaking, educated São Tomé Portuguese approximates the European Standard, although, as expected, it has certain local characteristics in phonology, morphosyntax and lexicon. However, São Tomé Portuguese displays a wide spectrum of varieties that reflect socioeconomic status and acquisition type — whether Portuguese is a first or second language.
17. This document reports an Angolan predominance among the some 500 African workers (and 40 Portuguese administrators).
18. Correspondence that I located between Monte Café administrators and the *Curadoria* 'Office for the Affairs of Contracted Workers' in 1876 also attests the presence of Kru, and Mina workers (Curadoria 1876, box 82, file 2, folder 4). Thus, although it may be said that the Monte Café Tongas descend mainly from Bantu-language speakers, principally from Angola, it must be kept in mind that Kru and Kwa elements were also present in the 19th century. This point establishes an interesting comparative link between the Tongas and certain Afro-Brazilian communities.
19. In the sample considered here, there is only one couple involving a Cape Verdean (father). The offspring of this couple do not speak Cape Verde Creole.
20. According to Cadbury (1910: 40), many plantations had creches for younger children, under the care of *enfermeiras pretas* 'black nurses'. The creche at Monte Café, may have opened at a later date, since our middle-aged and oldest informants made no reference to it. Rather, they mentioned that they accompanied their parents to the field, or conducted yard duties under the supervision of a contract worker. The details provided by our elderly informants

on the induction of children into the labor force through light work in the plantation administration compound are reminiscent of the observations made by Cadbury (1910:40) regarding the situation elsewhere on São Tomé. With extremely rare exceptions, children did not have access to formal education in the earlier periods of the indenturing system (witness Cadbury 1910:37, and testimonies of elderly informants). At the age of 14 years, children were contracted under the same conditions as their parents (Cadbury 1910:40).

21. The oldest speakers in this group were aged 90+. A future analysis will divide the (60+) group, creating a fourth age group.

22. Men above 65 years of age and women above 60 are retired, although they may continue to do some work for the plantation.

23. In the case of the youngest generational group, all speakers had attended a minimum of five years of primary education. However, the two informants from the second generational group had attended only the year one class, sporadically.

24. One reviewer of this paper made two valid points regarding the range of the independent linguistic factors analyzed. Firstly, in view of the complex semantic structure of the Bantu classifier system, it would be of interest to evaluate whether this system is reflected in any way in Tonga Portuguese, perhaps in terms of animacy privileges with respect to plural marking. This remains a goal for future research. Secondly, as Portuguese NP items which undergo plural marking are also marked for gender, it would be interesting to evaluate the degree and nature of possible correlations between these two agreement systems. This is also a goal for future research. So far, the only work relating to this issue involves dialects of Brazilian Portuguese such as that of Helvécia (Bahia), where variable gender agreement operates somewhat differently from plural marking. Gender marking is assigned from the head outwards in the NP, whereas plural marking is assigned from outside the head, from the beginning of the NP (Lucchesi 2000:256–8). However, pre-nuclear positions did tend to favor the presence of gender marking, much as Scherre (1988) had found in the case of the plural marker (Lucchesi 2000:261–2). Finally, there was considerable interdependence between the rules, such that, if number agreement was categorical, gender agreement was also categorical (Lucchesi 2000:263–4).

25. This framework did not contemplate the potential differential effect on plural marking that might be caused by words with a singular form ending in a non-nasal vowel (for example *casa* 'house') versus those ending in a nasal vowel (for example *ordem* 'order'). This is a goal of future research.

26. The rationale is that an analysis which considers the form class of a pluralizing item alone will fail to account for the fact that the position of the form classes is a major determinant of plural marking. Furthermore, consideration of the position alone of a pluralizing item will not account for the fact that the form class and function of the item interplay with its position. For further discussion see Scherre (1998).

27. Note that in a VARBRUL analysis the probability weightings are to be read in interpreting the results, *and not the percentages*. In a study involving unevenly distributed amounts of data from each informant, the percentage is not a reliable indicator of the effects of factor groups. For a discussion of the inadequacy of percentages and the rationale for using weighting probabilities, see Sankoff (1988).

28. Only data items with stress patterns matching those of Standard Portuguese were codified. Nevertheless, few items were excluded.

29. These results are based on VARBRUL run B, for each generation, explained in Section 8.

30. As Greg Guy has suggested (p.c.), a clarification is essential here. It is important to interpret the results in Table 5 with respect to their distribution across NP positions (see Table 7). Consider the high *pr.* weight for the pluralization of words with stressed final syllables, with a low % value, and the contrasting lower *pr.* weight of monosyllables, with a high % value. This apparent anomaly arises because most monosyllables are articles, occurring in NP initial position, where they are marked for plural at a rate of 87% to 98% across the three generations (see Table 7). It is in relation to other words that occur in first position in the NP that the monosyllables have a slightly unfavorable *pr.* weight. In contrast, the words with stressed final syllables are NP heads (in the traditional sense), mainly occurring in non-initial positions in the NP. In generation 1, in Table 7, a head in non-initial position is marked at a rate of 4%, overall. So, in relation to other heads occurring in non-initial position, the 14% marking of words with stressed final syllables, for generation 1 in Table 5, is highly favorable, and thus receives a heavy factor weighting.

31. TP *isso* < P. *isso* 'this' — a neutral pronoun.

32. The conversation refers to events of the 1950s, when the colonial government attempted to force the São Tomé Creoles to work on the plantations. The informant recalls that the Creoles had attempted to disrupt plantation work.

33. In the speech of my oldest informants, the preposition *ni* corresponds variably to 'in', 'to' and 'from'. Motivation for the conflation of the three distinct function forms from Portuguese into one form in Tonga Portuguese is found in Umbundu. Here three prepositions express location: (i) locative and directional *ku*; (ii) *mu*, which expresses location (within), and direction (both into and out of an enclosed space), and (iii) *mo* which expresses location (on). The fact that two of the Umbundo prepositions have directional functions while sharing locative functions with Portuguese *em* may have facilitated a transfer of their overall functions onto Portuguese *em*. In addition, in Umbundu-influenced Portuguese, initial [d] would be [nd], and this could facilitate the formation of *ni* from Portuguese *de* [di]. Alternatively, the values of *ni* in TP could be accounted for by appealing to a connection with São Tomé Creole.

34. The variable use of the first person singular to represent the third person singular is a low frequency variable.

35. The plural item is underlined and the head is represented in bold type.

36. The *conto* is a Portuguese measure meaning "one thousand monetary units".

37. In Generation 1 there were only three cases with an item in a non-adjacent position at the left of the head, and all were unmarked on the non-adjacent pre-nuclear items. The items adjacent to the head were determiners, demonstratives, numerals and quantifiers.

38. Interviews with João Ronco (in São Tomé), António Portugal (in Lisbon), António Carvalho Maia (former owner of the Monte Café plantation, in Lisbon), Delfim Pinho (former *feitor de mato* 'field foreman' and *feitor geral* 'general duties foreman', in Viseu), and Elias Correia (in Viseu). These men were associated with Monte Café as of the late 1940s, and there are no surviving administrators from earlier periods. In the views of the latter two interviewees, when they arrived in São Tomé the characteristics of Tonga Portuguese, some of which they were able to imitate during the interview, were already firmly in place.

39. *Massi* 'machete' < *P. machim* 'machete'.

40. Myers-Scotton's 4-M model proposes "that differences in when and how different morpheme types are salient in production is reflected in various surface outcomes" (Myers-Scotton 2001:3). The model categorizes morphemes into four classes: content morphemes and three types of system morpheme. Different classes of system morphemes are activated at different levels of language production. Early system morphemes are activated at the mental lexicon level (the *lemma* level) as a type of satellite to content morphemes, adding particular (core) meanings, such as, for example, plural *-s* in English, or definiteness of reference. Late system morphemes are assigned by structural requirements (Myers-Scotton 2000a). The concept of late system versus early system morphemes may help account for some of the phenomena discussed above, where the plural agreement systems of Umbundu and Tonga Portuguese are concerned. A key issue in this case would appear to lie in the need to distinguish some plural morphemes as being more basic, hence early system, as opposed to others that would be assigned structurally.

References

- Baxter, A. N. 1998. "Morfosyntaxe." In Perl & Schwegler, 97–134.
- Baxter, A. N. 2002. "Semicreolization? — The restructured Portuguese of the Tongas of São Tomé — a consequence of L1 acquisition in a special contact situation." *Journal of Portuguese Linguistics* 1: 7–39.
- Bayley, R. 1994. "Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English." In *Research Methodology in Second-Language Acquisition*, E. Tarone, S. Gass & A. D. Cohen (eds.), 157–81. Hillsdale: Lawrence Erlbaum Associates.
- Bayley, R. 1996. "Competing constraints on variation in the speech of adult Chinese learners of English." In *Second Language Acquisition and Linguistic Variation*, R. Bayley & D. R. Preston (eds.), 97–120. Amsterdam: John Benjamins.
- Bickerton, D. 1975. *Dynamics of a Creole System*. Cambridge: Cambridge University Press.
- Cadbury, W. A. 1910. *Os serviços de S. Thomé*. Lisbon: Livraria Bertrand.
- DeGraff, M. 1999. "Creolization, language change, and language acquisition: An epilogue." In DeGraff, 473–544.
- DeGraff, M. (ed.) 1999. *Language Creation and Language Change — Creolization, Diachrony, and Development*. Cambridge: The MIT Press.
- Emmerich, C. 1984. *A língua de contato no Alto Xingu — Origem, forma e função*. PhD Thesis. Universidade Federal do Rio de Janeiro.
- Gass, S. 1988. "Second language acquisition and linguistic theory: The role of language transfer." In *Linguistic Theory in Second Language Acquisition*, S. Flynn & W. O'Neill (eds.), 384–403. Dordrecht: Kluwer Academic Publishers.
- Guy, G. R. 1981. *Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax and language history*. Ann Arbor: University Microfilms International.
- Hodges, T. & M. Newitt. 1988. *São Tomé and Príncipe: From Plantation Colony to Microstate*. Boulder: Westview Press.
- Ingram, D. 1989. *First Language Acquisition*. Cambridge: Cambridge University Press.
- Lightfoot, D. 1999. "Creoles and cues." In DeGraff (ed.), 431–52.
- Lopes, N. 2001. *Tópicos de concordância*. PhD Thesis. Salvador: Universidade Federal da Bahia.

- Lorenzino, G. 1998. *The Angolar Creole Portuguese of São Tomé: Its Grammar and Socio-linguistic History*. Munich: Lincom Europa.
- Lucchesi, D. 2000. *A variação na concordância de gênero em uma comunidade de fala afro-brasileira: novos elementos sobre a formação do português popular do Brasil*. PhD Thesis. Universidade Federal do Rio de Janeiro.
- Maia, António da Silva. 1964. *Lições de gramática de quimbundo*. Cucujães: Escola Tipográfica das Missões.
- Major, R. C. 1996. "Markedness in second language acquisition of consonant clusters." In *Second Language Acquisition and Linguistic Variation*, R. Bayley & D.R. Preston (eds.), 75–96. Amsterdam: John Benjamins.
- Mantero, F. 1910. *Manual Labour in S. Thomé and Príncipe*. Lisbon: Francis Mantero.
- Mello, J.P. de. 1864. *Província de São Tomé e Príncipe. Produtos naturaes, e da indústria agrícola e fabril designados à Exposição Universal em Londres*. São Tomé: Imprensa Nacional.
- Mendes, B.C. 1985. *Contributo para o estudo da língua portuguesa em Angola*. Lisboa: Instituto de Lingüística da Faculdade de Letras de Lisboa.
- Myers-Scotton, C. 1997. *Duelling Languages: Grammatical Structure in Codeswitching*. New York: Oxford University Press.
- Myers-Scotton, C. & J.L. Jake. 2000a. "Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition." *Linguistics* 38: 1053–1100.
- Myers-Scotton, C. & J.L. Jake. 2000b. "Testing the 4-M model: An introduction." *International Journal of Bilingualism* 4: 1–8.
- Myers-Scotton, C. 2001. "Implications of abstract grammatical structure: Two targets in Creole formation." *Journal of Pidgin and Creole Languages* 16: 217–73.
- Naro, A.J. & M.M.P. Scherre. 2000. "Variable concord in Portuguese: The situation in Brazil and Portugal." In *Language Change and Language Contact in Pidgins and Creoles*, J. McWhorter (ed.), 235–56. Amsterdam: John Benjamins.
- Naro, A.J., & M. Lemle. 1977. "Syntactic diffusion." *Ciência e Cultura* 29: 259–68.
- Perl, M. & A. Schwegler (eds.). 1998. *América negra: panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*. Frankfurt am Main: Vervuert/Madrid: Iberoamericana.
- Pintzuk, S. 1988. *VARBRUL Programs*. Philadelphia: University of Pennsylvania Department of Linguistics.
- Rougé, J.-L. 1992. "Les langues des Tonga." In *Actas do Colóquio Internacional sobre Línguas Crioulas de Base Portuguesa*, E. d'Andrade & A. Kihm (eds.), 171–5. Universidade de Lisboa, Lisboa: Colibri.
- Sankoff, D. 1988. "Variable rules." In *Sociolinguistics — An International Handbook of the Science of Language and Society*, U. Ammon, N. Dittmar & K. Mattheier (eds.), 984–97. Berlin/New York: Walter de Gruyter.
- Scherre, M.M.P. 1988. *Reanálise da concordância nominal em português*. PhD Thesis. Rio de Janeiro: Universidade Federal do Rio de Janeiro.
- Scherre, M.M.P. 1998. "Sobre a influência de três variáveis relacionadas na concordância nominal em português." In *Padrões sociolingüísticos*, G.M. Oliveira e Silva & M.M.P. Scherre (eds.), 85–118. Rio de Janeiro: Tempo Brasileiro.
- Swan, C.A. 1910. *The Slavery of Today; or, The Present Position of the Open Sore of Africa*. Glasgow: Pickering & Inglis.
- Tenreiro, F. 1961. *A ilha de São Tomé*. Lisbon: Junta de Investigações do Ultramar.

- Valente, J.F. 1964. *Gramática umbundu — a língua do centro de Angola*. Lisboa: Junta de Investigações do Ultramar.
- Wolfram, W. 1985. "Variability in tense marking: A case for the obvious." *Language Learning* 35: 229–53.
- Zimmermann, K. (ed.). 1999. *Lenguas criollas de base lexical española y portuguesa*. Frankfurt am Main: Vervuert/Madrid: Iberoamericana.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY