

A CONCORDÂNCIA PLURAL VARIÁVEL NO SINTAGMA NOMINAL DO PORTUGUÊS  
REESTRUTURADO DA COMUNIDADE DE ALMOXARIFE, SÃO TOMÉ  
(Desenvolvimento das Regras de Concordância Variáveis no Processo de  
Transmissão-Aquisição Geracional)  
Vol. 1

por

**Carlos Filipe Guimarães Figueiredo**

Dissertação de Doutoramento em Linguística

2010



FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS

DEPARTAMENTO DE PORTUGUÊS

**UNIVERSIDADE DE MACAU**

A CONCORDÂNCIA PLURAL VARIÁVEL NO SINTAGMA NOMINAL DO PORTUGUÊS  
REESTRUTURADO DA COMUNIDADE DE ALMOXARIFE, SÃO TOMÉ  
(Desenvolvimento das Regras de Concordância Variáveis no Processo de  
Transmissão-Aquisição Geracional)  
Vol. 1

por

**Carlos Filipe Guimarães Figueiredo**

**Orientador:** Professor Doutor Alan Norman Baxter

Departamento de Português

Dissertação em Linguística para obtenção do grau de Doutor.

2010

Faculdade de Ciências Sociais e Humanas

UNIVERSIDADE DE MACAU

Copyright 2010 by Guimarães Figueiredo, Carlos Filipe  
University of Macau

Ao meu pai, que partiu em vésperas de eu me licenciar.

À minha mãe, exemplo ímpar de sacrifício e dedicação.

À Nani e ao Diogo, pela felicidade que trazem à minha existência.

Ao meu irmão Rui e aos meus filhos José Manuel e Ana Cátia.

À minha sogra Maria Ivone.



## **Agradecimentos**

Durante o período de tempo em que decorreu o presente estudo, fui acumulando dívidas, impossíveis de saldar. A primeira destas é para com todas as minhas fontes de informação, as quais permitiram que partilhasse das suas vidas e dos seus conhecimentos. De entre as fontes profissionais, tive o privilégio de cultivar amizade com algumas. Sem estabelecer prioridades, gostaria de referir o grande apreço e profundo sentimento de gratidão para com o meu orientador, Prof. Alan Norman Baxter, que, paciente e tolerantemente, dedicou horas sem fim a revisões, permitindo-me entender os objectivos e caminhos do rigor científico.

A segunda dívida de gratidão é para com a Universidade de Macau, que não só me concedeu a oportunidade para encetar um novo caminho na minha vida profissional mas também me proporcionou apoio e auxílio para levar a cabo a presente pesquisa. De entre o corpo profissional da Universidade, um agradecimento especial é para a Prof. Maria Antónia Espadinha, que permitiu que fosse acolhido no seio desta família e me deu estímulo antecipado, ajudando a lançar este projecto. Os agradecimentos estendem-se ao Hugo Cardoso, pelas sugestões e apoio na revisão de texto, bem como ao Aldino Dias, pelo incentivo e ajuda na revisão de texto. Quero também agradecer a todos aqueles que, de uma forma ou de outra, se interessaram pelo andamento da pesquisa. De entre estes, não posso deixar de citar a Paula Campos, a Sara Gonçalves, a Ana Paula Godinho, a Leonor Seabra, a Zhang Jing, o Yao Jing Ming e a Gao Lili, pela amizade e sentido de cooperação.

Mas as minhas dívidas intelectuais vão para além do corpo docente da Universidade de Macau. Esta pesquisa é uma excrescência do gosto pela Linguística, que nasceu durante o percurso de obtenção da minha licenciatura. O fascínio pelas variedades de português estava no gene, sem que me apercebesse, e palpitava de cada vez que, em Angola, ao abandonar a sala-de-aulas da escola primária, mudava o tipo de discurso para comunicar, de modo “genuíno”, com os colegas, fosse no pátio do recreio fosse nas brincadeiras de rua. As vicissitudes da vida determinaram que tivesse de interromper os estudos, quando aportei em Portugal, integrado no grosso dos “retornados” que buscava, na metrópole, alternativas de sobrevivência, finda que estava a aventura portuguesa da expansão colonial. Lembro-me, então, como me soava a estranho o idioma que

aprendera nas carteiras da escola primária, de cada vez que contactava com falantes das diversas regiões de Portugal. Retomando os estudos, já em fase em que começavam a despontar os primeiros cabelos brancos, o incentivo dos Profs. Malaca Casteleiro e Maria José Grosso foi determinante para que reavivasse em mim o interesse em tentar perceber a riqueza encerrada na “minha” variedade de português e entender que a mesma poderia dar contributo valioso para os estudos da aquisição, em particular, e da Linguística, em geral. Para ambos, deixo registada uma palavra de apreço muito especial.

Ao ser integrado no quadro profissional da Universidade de Macau, estava-me reservada a grata surpresa de poder dar continuidade a dois percursos que me completam pessoal e profissionalmente: exercer a carreira de docente, contactando com alunos que me ajudam a descobrir a profundidade de sentimentos e me animam para me envolver em projectos didácticos; desenvolver trabalhos de pesquisa sobre temas que me fascinam e que podem dar contributo valioso para os estudos em Linguística, para além de permitirem que contacte com especialistas que me possibilitam um cada vez maior crescimento intelectual.

A nível estreitamente pessoal, tenho ainda mais dívidas que nunca poderei saldar. Para com os meus pais, Céu e Amílcar, ímpares nos exemplos de rectidão humana e verdadeiros azimutes para me indicarem o trajecto da persistência, sacrifício e esforço. Para com o meu irmão Rui, sempre na vanguarda para me escudar nos momentos mais difíceis. Para com a minha sogra, Maria Ivone, nunca ausente com o seu estímulo e apoio. Para com o José Manuel e a Cátia, que me acolherem e aceitaram no seu seio como pai, ajudando-me a preencher a minha vida.

Para com a Nani, a minha dívida impagável deve-se ao sofrimento em silêncio, à paciência e à tolerância pelas minhas horas de “ausência”, dedicadas ao presente estudo. Quanto ao Diogo, a dívida tem a ver com a adolescência que não sentiu a minha presença diária, devido também ao tempo votado a este projecto. Vocês são, ambos, a magia da minha vida, o meu porto de acolhimento, o meu tudo.

## RESUMO

A pesquisa analisa a aquisição de marca de plural (PL) no sintagma nominal (SN) de uma variedade de português reestruturado, falada pela comunidade bilingue (crioulo e português) de Almojarife, São Tomé. Numa perspectiva da sociolinguística variacionista, são feitas comparações com variedades africanas de português que têm substratos do grupo níger-congo atlântico (o português vernacular de Angola, o português do Maputo, Moçambique, e o português de uma comunidade de descendentes de serviçais na ilha de São Tomé), o português cabo-verdiano, uma variedade brasileira de português rural e variedades brasileiras de português urbano. O estudo recorreu ao pacote GOLDVARB X, que permite identificar as variáveis linguísticas e extralinguísticas que influenciam a marcação plural nos constituintes do SN do português de Almojarife (PA), determinando a natureza da concordância plural variável (CPL-var) a nível comunitário. A análise estrutural da concordância no SN é sustentada por um quadro teórico heterogéneo da área da aquisição de primeiras línguas (L1's), aquisição bilingue e aquisição de L2's (ASL), enfocando, particularmente, o papel da Gramática Universal (GU) na ASL, bem como os preceitos da Morfologia Distribuída (MD) acerca da marcação variável nos elementos do SN. Comparando o português com o santomense, o crioulo de Cabo Verde (CCV) e certas línguas do grupo níger-congo atlântico, determina-se o impacto das características tipológicas das L1's tanto na interlíngua como na posterior língua nativizada que emerge em situação de transmissão linguística irregular, visando explicar a forma como ambos os sistemas podem ser modelados por aspectos linguísticos, psico-cognitivos e sociohistóricos. Uma das propostas do trabalho, baseada na Teoria dos Princípios e Parâmetros, é a hipótese de ocorrerem paralelismos na aquisição da marca de PL e desenvolvimento da CPL no SN das L1's de falantes monolíngues e bilingues, por um lado, e das L2's, por outro lado. Assume-se também que os falantes adultos de L2's acedem parcialmente à GU e que os parâmetros desactivados após o período crítico da aquisição não são totalmente reinstanciados. Os valores obtidos revelam que a CPL-var é motivada, essencialmente, por condicionantes do tipo estrutural, não se vislumbrando uma influência significativa das variáveis do tipo fonológico e semântico no desenho da mesma. A variável *classe gramatical* revelou que, em SN's de estrutura reduzida ocorre relação entre posição e



categoria gramatical: determinantes e primeira posição (posição pré-nuclear), nomes e segunda posição (núcleo) e adjetivos e terceira posição (posição pós-nuclear). Quanto às variáveis extralinguísticas, identificaram gramáticas em competição, com a dos falantes mais idosos evidenciando proximidade à gramática do substrato e a das gerações mais novas caminhando no sentido de uma ligeira aquisição das regras da gramática da língua-alvo (LA). A comparação dos resultados do PA com os das outras variedades africanas e brasileiras de português permitiu concluir que o português adquirido em situação de contacto por falantes de um crioulo português (CP) atlântico manifesta os mesmos padrões de variação do português adquirido numa situação de transmissão linguística irregular, na qual a língua ancestral não é um CP atlântico.

## ABSTRACT

This study focuses on the acquisition of plural (PL) marking in the noun phrase of a variety of restructured Portuguese spoken by the bilingual (Creole – Portuguese) community of Almojarife (São Tomé). Within a variationist sociolinguist perspective, comparisons are made with African varieties of Portuguese which have Niger-Congo Atlantic substrates (Angola vernacular Portuguese, Maputo Portuguese in Mozambique, and the Portuguese of descendents of indentured Africans in São Tomé), the Portuguese of Cape Verde, as well as one rural and two urban varieties of Brazilian Portuguese. The study of variation in PA makes use of the GOLDVARB X package, which permits the identification of linguistic and extralinguistic variables influencing plural marking of noun phrase constituents, thereby determining the nature of variable plural agreement at a community level. The structural analysis of noun phrase plural agreement is rooted in a heterogeneous theoretical framework within the domain of first language (L1) acquisition, bilingual acquisition, and second language acquisition (SLA), with particular emphasis on the role of Universal Grammar (UG) in SLA, as well as the tenets of Distributed Morphology (DM) with regard to variable marking of NP constituents. Comparisons of Portuguese with Santomense, Cape Verdean Creole (CVC) and certain Niger-Congo Atlantic languages establish the impact of L1 and its typological characteristics on both the learners' interlanguage and the nativised language in a context of *imperfect language shift*, in order to explain how both systems can be shaped by linguistic, psycho-cognitive and sociohistorical aspects. One of the working hypotheses is that the acquisition of plural marking parallels the development of NP number agreement in monolingual/bilingual L1's and L2's, based on the Theory of Principles and Parameters. It is also assumed that adult L2 speakers access UG partially and that parameters deactivated after the critical period are not entirely reset.

The values obtained demonstrate that PLA-var is essentially motivated by structural constraints, since phonological and semantic variables appear not to intervene significantly in shaping it. The *grammatical class* variable revealed that, in simple NP's, there is a correlation between position and grammatical category: determiners and first (pre-head) position, nouns and second (head) position, adjectives and third (post-head) position. Extra-linguistic variables, on the other hand, unearthed competing grammars:

that of the older speakers approaching substrate models, and that of the younger generations moving towards partial acquisition of target-language (LA) grammatical rules. The systematic comparison of PA results with those for other African and Brazilian varieties of Portuguese led to the conclusion that the Portuguese acquired in contact situations by speakers of an Atlantic Portuguese-based Creole displays the same patterns of variation as the Portuguese acquired in an *imperfect language shift* setting for which the ancestral language is not an Atlantic Portuguese-based Creole.

## Índice

### Lista de abreviaturas

### Lista de tabelas

### Lista de figuras

<b>Introdução</b> .....	1
<b>Capítulo 1. O tema</b> .....	11
1.1. Português do Brasil e concordância nominal variável de número .....	11
1.2 Variedades africanas de português e concordância nominal variável de número .....	31
1.2.1. Variedades africanas de português com substrato do grupo níger-congo atlântico ...	31
1.2.1.1. Português dos tongas (Roça Monte Café, São Tomé) .....	31
1.2.1.2. Português vernáculo de Angola .....	38
1.2.1.3. Português de Moçambique .....	40
1.2.2. Variedades africanas de português com substrato crioulo .....	45
1.2.2.1. Português caboverdiano e seu substrato (crioulo de Cabo Verde) .....	45
1.2.2.2. Português de Almojarife (São Tomé) e seu substrato (santomense) .....	59
1.2.2.3. A comunidade de Almojarife .....	85
1.2.2.3.1. Características da fala de Almojarife .....	90
1.3. Conclusões parciais .....	99
<b>Capítulo 2. Fundamentação teórica</b> .....	103
2.1. Os pressupostos teórico-metodológicos sobre aquisições de L1, bilingue e L2 .....	106
2.1.1. Variação e sistematicidade na <i>interlíngua</i> : perspectiva generativista .....	108
2.1.2. Variação e sistematicidade na <i>interlíngua</i> : perspectiva psicolinguística .....	113
2.1.3. Variação e sistematicidade na <i>interlíngua</i> : perspectiva sociolinguística .....	114
2.1.4. Variação e sistematicidade na <i>interlíngua</i> : aplicação da perspectiva heterogénea ao estudo do PA .....	116
2.2. Abordagens teóricas sobre aquisição de L1 .....	123
2.2.1. Modelo generativista .....	124
2.2.2. A teorização em torno da GU .....	125
2.2.2.1. O Modelo de Princípios e Parâmetros ( <i>Parameter settings</i> ) .....	125
2.2.2.1.2. Teoria das Regências e Ligações .....	127
2.2.3. Papel da GU na aquisição da L1 .....	130
2.2.4. GU e estudos de aquisição .....	135
2.3. Abordagens teóricas sobre aquisição bilingue .....	137

2.3.1. Acesso ao léxico por parte do bilingue .....	138
2.3.2. Papel da GU na aquisição bilingue: a Hipótese da Diferenciação da Linguagem .....	141
2.4. Abordagens teóricas sobre aquisição de L2's .....	145
2.4.1. O modelo generativista e papel da GU na aquisição de L2's .....	146
2.4.1.1. GU e teorias em torno da aquisição de L2's .....	147
2.4.2. Aquisição de L2's e modelos psicocognitivos .....	152
2.4.2.1. Conexionismo ( <i>Connectionism</i> ) e modelos de processamento paralelo distribuído .....	153
2.4.2.2. Teoria da "Processabilidade" ( <i>Processability Theory</i> ) .....	154
2.4.2.3. Modelo dos Quatro Morfemas ( <i>The 4M Model</i> ) .....	160
2.4.3. Modelo da mudança evolucionista da língua .....	161
2.5. Variação: pressupostos teóricos .....	166
2.5.1. Variáveis extralinguísticas, variação e mudança .....	169
2.5.2. Variáveis linguísticas e contexto .....	172
2.5.3. Modelos matemáticos para reprodução do efeito global dos dados empíricos .....	174
2.6. Concordância de número PL a nível do SN .....	178
2.6.1. Posicionamento e flexão em número dos constituintes do SN: visões descritivistas .....	178
2.6.2. O quadro teórico da concordância .....	186
2.6.2.1. Concordância e núcleos funcionais .....	187
2.6.3. O SN no português europeu .....	195
2.6.4. Categorias funcionais e categorias lexicais .....	197
2.6.4.1. Aquisição do sintagma determinante .....	201
2.6.4.2. Aquisição de género e número .....	206
2.6.5. Aquisição do sintagma determinante em L2 .....	214
2.6.6. Aquisição da concordância de PL no sintagma determinante .....	218
2.7. Aquisição das categorias funcionais da L2: os casos do santomense e do PA .....	225
2.8. A refixação dos parâmetros .....	229
2.9. Concordância variável ao nível do SN .....	231
2.10. Fossilização .....	239
2.10.1. Fossilização e variação .....	239
2.10.2. Potenciais causas de fossilização .....	242
2.10.3. GU e fossilização .....	244
2.10.4. Fossilização no PA .....	246

2.10.4.1. PA: transferência e fossilização fonológica .....	246
2.10.4.2. PA: transferência e fossilização morfossintáctica .....	250
2.11. Aquisição massiva de L2's: línguas crioulas e teorias de aquisição de L2's .....	253
2.11.1. Universais linguísticos vs. transferência: a Hipótese da Interlíngua ( <i>Interlanguage Hypothesis</i> ) .....	253
2.11.1.1. Teoria da “Processabilidade” e flexão morfológica das línguas emergentes por contacto .....	254
2.11.1.2. Teoria da “Processabilidade” e processamento sintáctico das línguas emergentes por contacto .....	257
2.11.2. Hipótese do Desenvolvimento Moderado da Transferência ( <i>Developmentally Moderated Transfer Hypothesis</i> ) .....	260
2.11.3. Transmissão linguística irregular .....	263
2.11.3.1. Vectores para a emergência da transmissão linguística irregular .....	264
2.11.3.1.1. Sociohistória e transmissão linguística irregular .....	267
2.11.3.1.2. Transmissão linguística irregular: parâmetros linguísticos .....	269
2.11.3.1.3. Transmissão linguística irregular e gramaticalização: heterogénese GU/transferência .....	273
2.11.3.2. Variedades crioulizantes .....	279
2.11.3.3. Transmissão linguística irregular e variação: gramáticas em competição .....	282
2.12. Princípio da coesão estrutural e variação a nível do sintagma determinante .....	294
2.13. Conclusões parciais .....	295
<b>Capítulo 3. A metodologia</b> .....	299
3.1. Aplicação dos métodos quantitativos no estudo da aquisição de línguas .....	299
3.1.1. Recolha e transcrição dos dados do <i>corpus</i> .....	302
3.1.2. Tratamento dos dados .....	307
3.1.3. Os dados dos informantes .....	307
3.1.4. Caracterização dos dados .....	309
3.2. O suporte computacional .....	312
3.3. Descrição da variável dependente e das variáveis independentes .....	315
3.3.1. Variável dependente .....	316
3.3.2. Variáveis independentes .....	317
3.4. Codificação dos dados .....	319
3.5. Variável dependente constituída .....	321
3.6. Variáveis independentes constituídas .....	322

3.6.1. Variável independente <i>saliência fónica</i> e suas subvariáveis .....	322
3.6.1.1. <i>Saliência fónica</i> .....	322
3.6.1.2. <i>Tonicidade (Tonicidade dos itens lexicais singulares)</i> .....	330
3.6.2. Variável independente <i>marcas precedentes ao item analisado</i> .....	334
3.6.3. Variável independente <i>contexto fonológico posterior</i> .....	338
3.6.4. Variável independente <i>classe gramatical do item analisado</i> .....	343
3.6.4.1. Constituição e definição dos factores para a variável <i>classe gramatical</i> .....	345
3.6.5. Variável independente <i>posição do item analisado em relação ao núcleo do SN</i> .....	350
3.6.6. Variável independente <i>posição do item na cadeia do SN (Posição linear)</i> .....	354
3.6.7. Variável independente <i>marcação de género</i> .....	359
3.6.8. Variável independente <i>ordem do item na cadeia dos constituintes flexionáveis em número</i> .....	362
3.6.9. Variável independente <i>grau de concordância de número no SN</i> .....	364
3.6.10. Variável independente <i>traço semântico do SN (Animacidade)</i> .....	365
3.6.11. Variável independente <i>sexo</i> .....	368
3.6.12. Variável independente <i>idade</i> .....	371
3.6.13 Variável independente <i>estadia (Permanência fora da comunidade)</i> .....	372
3.6.14. Variável independente <i>escolaridade</i> .....	374
3.7. Codificação dos itens de acordo com os códigos atribuídos aos factores .....	377
3.8. Quantificação dos dados para selecção das variáveis independentes e dos factores a serem incorporados na análise final .....	378
3.8.1. Iteração entre grupos de factores .....	381
3.8.2. Variáveis independentes a serem incorporadas na análise final .....	388
<b>Capítulo 4.</b> Análise dos resultados: variável dependente e variáveis independentes do tipo estrutural .....	405
4.1. Resultados das variáveis constituídas e respectivas análises .....	407
4.1.1. Variável dependente .....	407
4.1.2. Variáveis independentes do tipo estrutural .....	409
4.1.2.1. Variável independente <i>posição em relação ao núcleo do SN</i> .....	410
4.1.2.1.1. Relação entre <i>classe gramatical</i> e <i>posição em relação núcleo do SN</i> .....	411
4.1.2.2. Variável independente <i>ordem do item na cadeia dos constituintes flexionáveis</i> .....	415
4.1.2.2.1. Relação entre <i>ordem do item na cadeia dos constituintes flexionáveis</i> e <i>posição linear</i> .....	417
4.1.2.3. Variável independente <i>classe gramatical</i> .....	420

4.1.2.4. Variável independente <i>marcas precedentes ao item analisado</i> .....	424
4.1.2.4.1. <i>Marcas precedentes</i> : influência dos factores individuais na marcação PL do elemento analisado .....	435
4.1.2.4.2. <i>Marcas precedentes</i> : contribuição dos factores amalgamados para a marcação PL elemento analisado .....	446
4.1.2.4.2.1. Factores amalgamados para estudo do efeito das <i>marcas precedentes</i> na marcação PL do elemento analisado: análise I .....	447
4.1.2.4.2.2. Factores amalgamados para estudo do efeito das <i>marcas precedentes</i> na marcação PL do elemento analisado: análise II .....	453
4.1.2.4.2.3. Factores amalgamados para estudo do efeito das <i>marcas precedentes</i> na marcação PL do elemento analisado: análise III .....	457
4.1.2.4.3. Relação entre <i>marcas precedentes</i> , <i>posição linear</i> e <i>classe gramatical</i> na marcação PL dos itens do SN .....	462
4.1.2.4.4. Análises alternativas considerando <i>marcas precedentes</i> , <i>posição linear</i> e <i>classe gramatical</i> .....	464
4.1.2.4.5. <i>Faixas etárias</i> e perfil de marcação motivado pelas <i>marcas precedentes</i> .....	470
4.1.2.4.6. <i>Marcas precedentes</i> representadas por marcador semântico .....	475
4.1.2.5. Variável independente <i>grau de concordância de número no SN</i> .....	483
4.1.2.6. Variável independente <i>posição do item na cadeia do SN (Posição linear)</i> .....	491
4.1.2.6.1. Relação entre <i>posição linear</i> e <i>classe gramatical</i> na marcação PL dos itens do SN .....	494
4.1.2.6.2. Relação entre <i>posição linear</i> e <i>posição em relação ao núcleo do SN</i> na marcação PL dos itens do SN .....	503
<b>Capítulo 5.</b> Análise dos resultados: variáveis independentes do tipo semântico e fonológico; variáveis extralinguísticas .....	537
5.1. Variável independente do tipo semântico .....	538
5.1.1. Variável independente <i>traço semântico do SN (Animacidade)</i> .....	538
5.2. Variáveis independentes do tipo fonológico .....	541
5.2.1. Variável independente <i>saliência fónica (Processos morfofonológicos de formação de PL)</i> .....	541
5.2.1.1. Relação entre <i>saliência fónica</i> e <i>escolaridade</i> na marcação PL do PA .....	548
5.2.2. Variável independente <i>tonicidade (Tonicidade dos itens lexicais singulares)</i> .....	553
5.2.2.1. Relação entre <i>tonicidade</i> e <i>escolaridade</i> na marcação PL do PA e do MRJ ....	555
5.2.2.2. Efeito da <i>saliência (saliência fónica x tonicidade)</i> na marcação PL dos itens do SN do PA .....	557
5.2.2.3. Efeito da <i>saliência (saliência fónica x tonicidade)</i> na marcação PL do PA e da NURC .....	561



5.2.2.4. Relação entre <i>saliência (saliência fónica x tonicidade)</i> e escolaridade (PA e NURC) .....	564
5.2.2.5. Efeito da <i>saliência (saliência fónica x tonicidade)</i> na marcação PL do PA e do PT .....	567
5.2.2.6. Relação entre <i>saliência (saliência fónica x tonicidade)</i> e <i>idade</i> na marcação PL dos itens do SN do PA e do PT .....	570
5.2.2.7. Efeito da <i>saliência (saliência fónica x tonicidade)</i> na marcação PL dos itens do SN: 3 variedades de português .....	576
5.2.3. Variável independente <i>contexto fonológico posterior</i> .....	581
5.2.3.1. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo os traços de fonte .....	584
5.2.3.2. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando as consoantes num único factor .....	588
5.2.3.3. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando as pausas num único factor .....	591
5.2.3.4. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo os traços de classe principal .....	593
5.2.3.5. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo os traços coronal e não-coronal .....	596
5.2.3.6. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo o modo de articulação: 6 factores .....	599
5.2.3.7. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo o modo de articulação: 7 factores .....	601
5.2.3.8. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo as zonas de articulação .....	605
5.2.3.9. Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN, englobando consoantes segundo os pontos de articulação .....	608
5.2.3.10. Efeito do <i>contexto fonológico posterior</i> na marcação PL da <i>saliência fónica</i> .....	612
5.2.3.10.1. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo os traços de fonte, para a marcação PL dos itens do SN .....	613
5.2.3.10.2. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo os traços de classe principal, para a marcação PL dos itens do SN .....	627
5.2.3.10.3. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo os traços coronal e não-coronal, para a marcação PL dos itens do SN .....	629
5.2.3.10.4. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo o modo de articulação, para a marcação PL dos itens do SN: 6 factores .....	631

5.2.3.10.5. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo o modo de articulação, para a marcação PL dos itens do SN: 7 factores .....	633
5.2.3.10.6. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo as zonas de articulação, para a marcação PL dos itens do SN .....	635
5.2.3.10.7. Efeito do <i>contexto fonológico posterior</i> na <i>saliência fónica</i> , englobando consoantes segundo os pontos de articulação, para a marcação PL dos itens do SN .....	637
5.2.3.11. Efeito das <i>marcas precedentes</i> na <i>saliência fónica</i> .....	640
5.3 Variáveis extralinguísticas .....	652
5.3.1. Variável independente <i>escolaridade</i> .....	653
5.3.2. Variável independente <i>idade</i> .....	663
5.3.3. Variável independente <i>estadia (Permanência fora da comunidade)</i> .....	671
5.3.4. Variável independente <i>sexo</i> .....	674
<b>Conclusões finais</b> .....	681
Referências bibliográficas .....	711
I. Bibliografia consultada e citada .....	711
II. Bibliografia pertinente consultada .....	753
III. Sítios pertinentes consultados .....	785
<i>Curriculum vitae</i> do autor .....	789



## Lista de abreviaturas

<b>ADJ</b>	Adjectivo
<b>Adj</b>	Adjunto
<b>ADV</b>	Advérbio
<b>AFX</b>	Afixo
<b>AHM</b>	Arquivo Histórico Militar (Lisboa)
<b>AP</b>	Complemento pré-nominal
<b>ART</b>	Artigo
<b>AISL</b>	Aquisição imperfeita de segundas línguas
<b>ASL</b>	Aquisição de segundas línguas
<b>AUX</b>	Verbo auxiliar
<b>CCV</b>	Crioulo de Cabo Verde
<b>cf.</b>	Conforme
<b>CGB</b>	Crioulo da Guiné Bissau
<b>CGEN-var</b>	Concordância variável de género
<b>CGG</b>	Crioulo do Golfo da Guiné
<b>Cod.</b>	Código
<b>COD's</b>	Construções com Objecto Duplo
<b>COMP</b>	Complementador; marcador completivo; marcador comparador
<b>Conc.</b>	Concordância
<b>CP</b>	Crioulo de base portuguesa
<b>CPLP</b>	Comunidade dos Países de Língua Portuguesa
<b>CPL-var</b>	Concordância plural variável
<b>CPrep's</b>	Construções com Preposições
<b>CVS's</b>	Construções com Verbos Seriais
<b>DEF</b>	Definido
<b>DEM</b>	Demonstrativo
<b>DE'S</b>	Descrições estruturais
<b>DET</b>	Determinante
<b>DLP's</b>	Dados linguísticos primários ou <i>triggers</i>
<b>DOC</b>	Construções com duplo objecto
<b>doc.</b>	Documento
<b>DOC.</b>	Documentalista

<b>EP</b>	Espanhol do Panamá
<b>EPR</b>	Espanhol de Porto Rico
<b>Esp</b>	Espanhol
<b>Estrutura-P</b>	Estrutura profunda
<b>Estrutura-S</b>	Estrutura de superfície
<b>ex.</b>	Exemplo
<b>FE-1</b>	Faixa etária 1 (20 – 40 anos)
<b>FE-2</b>	Faixa etária 2 (41 – 60 anos)
<b>FE-3</b>	Faixa etária 3 (+ 60 anos)
<b>Fem.</b>	Feminino
<b>FFFH</b>	<i>Failed Functional Features Hypothesis</i> (Falha do Restabelecimento dos Parâmetros)
<b>Fig.</b>	Figura
<b>fl.</b>	Folha
<b>FL</b>	Forma Lógica
<b>FLEX</b>	Flexão
<b>FL0</b>	Estado inicial do processo de aquisição de uma língua
<b>FLf</b>	Estado final do processo de aquisição de uma língua
<b>FTFA</b>	<i>Full Transfer Full Access Hypothesis</i> (Acesso Total aos Princípios e Parâmetros)
<b>FUT</b>	Futuro
<b>GEN</b>	Género
<b>GU</b>	Gramática Universal
<b>HEL-Ba</b>	Dialecto de Helvécia, Bahia (Brasil)
<b>HI</b>	Hipótese da Interlíngua
<b>I</b>	Sintagma entoacional
<b>i.e.</b>	Isto é
<b>IFX</b>	Infixo
<b>IL</b>	Interlíngua
<b>IMP</b>	Imperfectivo
<b>IND</b>	(Presente do) Indicativo
<b>INDF</b>	Indefinido
<b>ININT</b>	Ininteligível
<b>INF</b>	Informante

<b>INT</b>	Interrogativo(a); interrogação
<b>L</b>	Língua
<b>L1</b>	Língua primeira
<b>L2</b>	Língua segunda
<b>LA</b>	Língua-alvo
<b>LDB</b>	Lei de Directrizes e Bases da Educação
<b>LCV</b>	Língua Caboverdiana
<b>LE</b>	Língua estrangeira
<b>LF</b>	Língua-fonte
<b>LM</b>	Língua materna
<b>LO</b>	Língua de origem
<b>LOC</b>	Locativo
<b>Masc.</b>	Masculino
<b>MD</b>	Morfologia distribuída
<b>MLF</b>	<i>Matrix Language Frame Model</i>
<b>MLU</b>	<i>Mean Length of Utterance</i> (Comprimento Médio do Enunciado)
<b>MRJ</b>	Dialecto do Município do Rio de Janeiro (Brasil)
<b>MSIH</b>	<i>Missing Surface Inflection Hypothesis</i>
<b>MSpec</b>	Especificador de marcação
<b>N</b>	Nome; substantivo
<b>NEG</b>	Negação; negative
<b>NUM</b>	Numeral
<b>NURC</b>	Norma urbana culta (Brasil)
<b>OD</b>	Objecto directo
<b>OInd</b>	Objecto indirecto
<b>OP</b>	Oração pequena
<b>p.</b>	Página
<b>PA</b>	Português de Almojarife (São Tomé)
<b>PB</b>	Português do Brasil
<b>PCV</b>	Português de Cabo Verde
<b>p.e.</b>	Por exemplo
<b>PE</b>	Português europeu
<b>Pe.</b>	Padre

<b>PEH</b>	Princípio do Espelho
<b>PERF</b>	Perfectivo
<b>PF</b>	Forma Fonológica
<b>PEX</b>	Prefixo
<b>PIL</b>	Princípio da Integridade Lexical
<b>PL</b>	Plural
<b>PLd</b>	Português de Luanda (Angola)
<b>PM</b>	Português de Moçambique
<b>PMn</b>	Português de Minas Gerais (Brasil)
<b>PMp</b>	Português do Maputo (Moçambique)
<b>POP</b>	Grupo do português popular
<b>POS</b>	Possessivo
<b>pp.</b>	Páginas
<b>PP</b>	Pronome pessoal
<b>PPA</b>	Português de Porto Alegre, Brasil
<b>pr.</b>	Peso relativo
<b>proc.</b>	Processo
<b>PREP</b>	Preposição
<b>PRET</b>	Pretérito
<b>PRO</b>	Pronome
<b>PROG</b>	Progressivo
<b>PST</b>	Português de São Tomé
<b>PT</b>	Português dos tongas (São Tomé)
<b>Ptg</b>	Português
<b>PtgLA</b>	Português língua-alvo
<b>PtgL1</b>	Português língua primeira
<b>PtgL2</b>	Português língua segunda
<b>PtgLE</b>	Português língua estrangeira
<b>PtgLM</b>	Português língua materna
<b>PVA</b>	Português vernáculo de Angola
<b>PVB</b>	Português vernáculo do Brasil
<b>Q</b>	Quantificador
<b>REL</b>	Pronome relativo

<b>SADJ</b>	Sintagma adjectival
<b>SADV</b>	Sintagma adverbial
<b>SCOMPL</b>	Sintagma complementador
<b>SDET</b>	Sintagma determinante
<b>SFLEX</b>	Sintagma flexional
<b>SFX</b>	Sufixo
<b>Spec</b>	Especificador
<b>SPREP</b>	Sintagma preposicional
<b>SG</b>	Singular
<b>SN</b>	Sintagma nominal
<b>SNN</b>	Sintagma nominal nu
<b>SQ</b>	Sintagma quantificador
<b>SUJ</b>	Sujeito
<b>SV</b>	Sintagma verbal
<b>T</b>	Projecção máxima de informações de tempo
<b>Trad.</b>	Tradução; tradutor
<b>TMA</b>	Tempo, modo e aspecto
<b>UNI</b>	Grupo do português universitário





## Lista de tabelas

<b>Tabela</b>	<b>Título da tabela</b>	
Tabela 1.1	Informação encerrada nos SNN's do santomense .....	72
Tabela 2.1	Hierarquia hipotética da articulação dos processamentos na aquisição da interlíngua .....	158
Tabela 3.1	Articulação dos processamentos na aquisição da interlíngua inglesa .....	159
Tabela 4.1	Taxionomia dos factores motivadores da fossilização .....	246
Tabela 5.1	A assunção: similaridades estruturais configuram transferência .....	258
Tabela 6.1	Os factos: similaridades estruturais podem não representar transferência ....	259
Tabela 7.1	Inversão da ordem frásica na interlíngua alemã .....	259
Tabela 8.1	Informantes de Almoxarife: faixas etárias e seus aspectos sociais .....	309
Tabela 9.1	A variável dependente (análise atomística) .....	317
Tabela 10.1	Variáveis independentes (análise atomística) .....	319
Tabela 11.1	Factores constituídos para a variável dependente .....	321
Tabela 12.1	Factores constituídos para a variável independente <i>saliência fónica</i> .....	329
Tabela 12.2	Factores constituídos para a variável independente <i>tonicidade</i> .....	333
Tabela 12.3	Factores constituídos para a variável independente <i>marcas precedentes ao item analisado</i> .....	337
Tabela 12.4	Factores constituídos para a variável independente <i>contexto fonológico posterior</i> .....	342
Tabela 12.5	Factores constituídos para a variável independente <i>classe gramatical do item analisado</i> .....	349
Tabela 12.6	Factores constituídos para a variável independente <i>posição do item analisado em relação ao núcleo do SN</i> .....	354
Tabela 12.7	Factores constituídos para a variável independente <i>posição do item na cadeia do SN</i> .....	358
Tabela 12.8	Factores constituídos para a variável independente <i>marcação de género</i> ....	361
Tabela 12.9	Factores constituídos para a variável independente <i>ordem do item na cadeia dos constituintes flexionáveis em número</i> .....	363
Tabela 12.10	Factores constituídos para a variável independente <i>grau de concordância de número no SN</i> .....	364
Tabela 12.11	Factores constituídos para a variável independente <i>traço semântico do SN</i> .	368
Tabela 12.12	Factores constituídos para a variável independente <i>sexo</i> .....	370
Tabela 12.13	Factores constituídos para a variável independente <i>idade</i> .....	372
Tabela 12.14	Factores constituídos para a variável independente <i>estadia</i> .....	373
Tabela 12.15	Factores constituídos para a variável independente <i>escolaridade</i> .....	377

Tabela 13.1	Exemplo de cadeia de códigos resultante da codificação dos elementos do SN .....	378
Tabela 14.1	Escala hierárquica decrescente das variáveis independentes seleccionadas como válidas para análise da CPL-var no SN do PA .....	384
Tabela 15.1	<i>Variável dependente</i> : percentuais de marcas plural nos itens do SN do PA .	407
Tabela 15.2	Percentuais de SN's plenamente marcados no PA .....	407
Tabela 15.3	Percentuais de SN's plenamente marcados: 2 variedades de português .....	408
Tabela 15.4	Percentuais de marcas PL nos itens do SN: 7 variedades de português .....	409
Tabela 16.1	Efeito da <i>posição em relação ao núcleo do SN</i> na marcação PL dos itens do SN do PA .....	410
Tabela 16.2	Relação entre <i>classe gramatical</i> e <i>posição em relação ao núcleo do SN</i> : distribuição das classes gramaticais relativamente ao núcleo do SN .....	412
Tabela 17.1	Efeito da variável independente <i>ordem na cadeia dos constituintes flexionáveis</i> na marcação PL dos itens do SN .....	416
Tabela 17.2	Efeito da relação entre <i>posição linear</i> e <i>ordem do item na cadeia dos constituintes flexionáveis</i> na marcação PL dos itens do SN .....	418
Tabela 18.1	Efeito da <i>classe gramatical</i> na marcação PL dos itens do SN .....	420
Tabela 18.2	<i>Categoria gramatical</i> : efeito da <i>classe gramatical</i> na marcação PL no elemento analisado: 3 variedades de português .....	421
Tabela 19.1	Efeito das <i>marcas precedentes</i> na marcação PL dos itens do SN .....	426
Tabela 19.2	<i>Marcas precedentes</i> : contribuição dos factores individuais para a marcação PL do elemento analisado – EPR e 5 variedades de português ....	436
Tabela 19.3	<i>Marcas precedentes</i> : configuração SS_ (item na terceira posição, duas marcas formais precedentes) – Informantes e grau de escolarização .....	443
Tabela 19.4	Efeito das <i>marcas precedentes</i> na marcação PL dos itens do SN do PA [Factores amalgamados: Análise I] .....	450
Tabela 19.5	<i>Marcas precedentes</i> : contribuição dos factores amalgamados para a marcação PL do elemento analisado – 3 variedades de português .....	452
Tabela 19.6	Efeito das <i>marcas precedentes</i> na marcação PL dos itens do SN do PA [Factores amalgamados: Análise II] .....	455
Tabela 19.7	<i>Marcas precedentes</i> : contribuição dos factores amalgamados para a marcação PL do elemento analisado – 2 variedades de português .....	455
Tabela 19.8	Efeito das <i>marcas precedentes</i> na marcação PL dos itens do SN do PA [Factores amalgamados: Análise III] .....	458
Tabela 19.9	<i>Marcas precedentes</i> : contribuição dos factores amalgamados para a marcação PL do elemento analisado – 3 variedades de português .....	460
Tabela 19.10	Efeito das <i>marcas precedentes</i> na marcação PL dos itens do SN do PA: <i>faixas etárias</i> .....	473

Tabela 20.1	Efeito do <i>grau de concordância de número no SN</i> na marcação PL dos itens do SN do PA .....	484
Tabela 20.2	Efeito do <i>grau de concordância de número no SN</i> no percentual de marcação PL dos itens do SN: comunidade e faixas etárias .....	488
Tabela 21.1	Grupo de factores <i>posição do item na cadeia do SN</i> e respectivos factores, para estudo da configuração do SN PL do PA. [Grupo de factores a testar para incorporação na análise de dados] .....	492
Tabela 21.2	Teste de significância estatística dentro do grupo de factores <i>posição do item na cadeia do SN</i> .....	492
Tabela 21.3	Efeito da <i>posição linear</i> na marcação PL dos itens do SN do PA .....	493
Tabela 21.4	<i>Posição linear</i> : contribuição dos factores individuais para a marcação PL do elemento analisado: 2 variedades de português .....	594
Tabela 21.5	Relação entre <i>posição linear</i> e <i>classe gramática</i> : distribuição das classes gramaticais pelas diversas posições do SN .....	595
Tabela 21.6	Efeito da relação entre <i>posição linear</i> e <i>posição em relação ao núcleo do SN</i> na marcação PL dos itens do SN do PA: total de ocorrências e percentuais de marcação .....	504
Tabela 21.7	Efeito da relação entre <i>posição linear</i> e <i>posição em relação ao núcleo do SN</i> na marcação PL dos itens do SN do PA .....	504
Tabela 21.8	Relação entre <i>posição linear</i> e <i>posição em relação ao núcleo do SN</i> na marcação PL dos itens do SN do PA: comunidade e faixas etárias .....	507
Tabela 21.9	Relação entre <i>posição linear</i> e <i>posição em relação ao núcleo do SN</i> na marcação PL dos itens do SN: 6 variedades de português .....	514
Tabela 22.1	Efeito do <i>traço semântico do SN</i> na marcação PL dos itens do SN do PA ...	538
Tabela 22.2	<i>Traço semântico do SN</i> : contribuição dos factores individuais para a marcação PL do elemento analisado – 3 variedades de português .....	539
Tabela 23.1	Efeito da <i>saliência fónica</i> na marcação PL dos itens do SN do PA .....	541
Tabela 23.2	<i>Saliência fónica</i> : contribuição dos factores individuais para a marcação PL do elemento analisado – 7 variedades de português .....	544
Tabela 23.3	Relação entre <i>saliência fónica</i> e <i>escolaridade</i> na marcação PL dos itens do SN do PA .....	550
Tabela 24.1	Efeito da <i>tonicidade</i> na marcação PL dos itens do SN do PA .....	553
Tabela 24.2	<i>Tonicidade</i> : contribuição dos factores individuais para a marcação PL do elemento analisado – 3 variedades de português .....	554
Tabela 24.3	Relação entre <i>tonicidade</i> e <i>escolaridade</i> na marcação PL dos itens do SN do PA .....	556
Tabela 24.4	<i>Saliência</i> resultante do cruzamento entre <i>saliência fónica</i> e <i>tonicidade</i> (todos os dados): grupo de factores constituído consoante metodologia de Lopes .....	559

Tabela 24.5	Efeito da <i>saliência</i> ( <i>saliência fónica x tonicidade</i> ) na marcação PL dos itens do SN do PA: grupo de factores constituído consoante metodologia de Lopes .....	560
Tabela 24.6	<i>Saliência</i> ( <i>saliência fónica x tonicidade</i> ): contribuição dos factores individuais para a marcação PL no elemento analisado – 3 variedades de português .....	562
Tabela 24.7	Relação entre <i>saliência</i> e <i>escolaridade</i> na marcação PL dos itens do SN do PA: grupo de factores constituído consoante metodologia de Lopes .....	565
Tabela 24.8	<i>Saliência</i> resultante do cruzamento entre <i>saliência fónica</i> e <i>tonicidade</i> (todos os dados): grupo de factores constituído consoante metodologia de Baxter .....	568
Tabela 24.9	Efeito da <i>saliência</i> ( <i>saliência fónica x tonicidade</i> ) na marcação PL dos itens do SN do PA: grupo de factores constituído consoante metodologia de Baxter .....	568
Tabela 24.10	<i>Saliência</i> ( <i>saliência fónica x tonicidade</i> ): contribuição dos factores individuais para a marcação PL do elemento analisado – 2 variedades de português .....	569
Tabela 24.11	Relação entre <i>saliência</i> e <i>idade</i> na marcação PL dos itens do SN do PA: grupo de factores constituído consoante metodologia de Baxter .....	572
Tabela 24.12	<i>Saliência</i> resultante do cruzamento entre <i>saliência fónica</i> e <i>tonicidade</i> (todos os dados): grupo de factores constituído consoante metodologia de Jon-And .....	577
Tabela 24.13	Efeito da <i>saliência</i> ( <i>saliência fónica x tonicidade</i> ) na marcação PL dos itens do SN do PA: grupo de factores constituído consoante metodologia de Jon-And .....	577
Tabela 24.14	<i>Saliência</i> ( <i>saliência fónica x tonicidade</i> ): contribuição dos factores individuais para a marcação PL do elemento analisado: 4 variedades de português .....	579
Tabela 25.1	Efeito do <i>contexto fonológico posterior</i> na marcação PL dos itens do SN do PA .....	581
Tabela 25.2	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os traços de fonte) na marcação PL dos itens do SN .....	585
Tabela 25.3	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes num único factor) na marcação PL dos itens do SN .....	589
Tabela 25.4	Efeito do <i>contexto fonológico posterior</i> (englobando pausas num único factor), na marcação PL dos itens do SN .....	592
Tabela 25.5	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os traços de classe principal) na marcação PL dos itens do SN .....	594
Tabela 25.6	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os traços coronal e não-coronal) na marcação PL dos itens do SN .....	598
Tabela 25.7	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo o modo de articulação) na marcação PL dos itens do SN: 6 factores .....	600

Tabela 25.8	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo o modo de articulação) na marcação PL dos itens do SN: 7 factores .....	602
Tabela 25.9	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo as zonas de articulação) na marcação PL dos itens do SN .....	606
Tabela 25.10	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os pontos de articulação) na marcação PL dos itens do SN .....	611
Tabela 25.11	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os traços de fonte) na marcação PL da <i>saliência fónica</i> .....	613
Tabela 25.12	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os traços de classe principal) na marcação PL da <i>saliência fónica</i> .....	628
Tabela 25.13	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os traços coronal e não-coronal) na marcação PL da <i>saliência fónica</i> .....	630
Tabela 25.14	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo o modo de articulação) na marcação PL da <i>saliência fónica</i> : 6 factores .....	632
Tabela 25.15	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo o modo de articulação) na marcação PL da <i>saliência fónica</i> : 7 factores .....	634
Tabela 25.16	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo as zonas de articulação) na marcação PL da <i>saliência fónica</i> .....	635
Tabela 25.17	Efeito do <i>contexto fonológico posterior</i> (englobando consoantes segundo os pontos de articulação) na marcação PL da <i>saliência fónica</i> .....	637
Tabela 25.18	Efeito das <i>marcas precedentes</i> na marcação PL da <i>saliência fónica</i> .....	641
Tabela 25.19	Itens com singular em <i>-s</i> e <i>-r</i> (Marcação fonológico-geracional) .....	644
Tabela 26.1	Grupo de factores <i>escolaridade</i> e respectivos factores, para estudo da configuração do SN PL do PA: grupo de factores para teste de factores a incorporar na análise de dados .....	655
Tabela 26.2	Efeito da <i>escolaridade</i> na marcação PL dos itens do SN do PA .....	656
Tabela 26.3	<i>Escolaridade</i> : contribuição dos factores individuais para a marcação PL do elemento analisado – 4 variedades de portugueses .....	658
Tabela 27.1	Efeito da variável <i>idade</i> na marcação PL dos itens do SN do PA .....	663
Tabela 27.2	SN's plenamente marcados no PA: valores por <i>idade</i> .....	664
Tabela 27.3	<i>Idade</i> : contribuição dos factores individuais para a marcação PL do elemento analisado – 6 variedades de portugueses .....	668
Tabela 28.1	Efeito da <i>estadia</i> na marcação PL dos itens do SN do PA .....	673
Tabela 29.1	Efeito da variável social <i>sexo</i> na marcação PL dos itens do SN do PA .....	674
Tabela 29.2	Marcação PL dos itens do SN do PA: valores por <i>idade</i> e <i>sexo</i> .....	675
Tabela 29.3	SN's plenamente marcados no PA: valores por <i>sexo</i> .....	676
Tabela 29.4	Variável social <i>sexo</i> : contribuição dos factores individuais para a marcação plural do elemento analisado – 5 variedades de portugueses .....	678



## Lista de figuras

<b>Figura</b>	<b>Legenda</b>	
Figura 1	Dialecto de HEL-Ba e PT: desenvolvimento da atribuição de PL .....	36
Figura 2	Marcação plural no CCV (Marcador visível em DET) .....	56
Figura 3	Marcação plural no CCV (Ausência de marcador em DET) .....	56
Figura 4	Estrutura sintáctico-discursiva dos sintagmas nominais nus no santomense ....	72
Figura 5	Estrutura sintáctica do SDET no santomense (marcação de número) .....	76
Figura 6	Índice demográfico de Santana e Almojarife .....	86
Figura 7	Percentagem das crianças com 5-17 anos que têm acesso à escolaridade .....	89
Figura 8	Evolução da população de São Tomé e Príncipe – 1940 a 2001 .....	90
Figura 9	Modelo de Princípios e Parâmetros .....	129
Figura 10	Modelo de aquisição das línguas.....	130
Figura 11	Estrutura X-barra .....	131
Figura 12	Estrutura alternativa gerada a partir da estrutura X-barra .....	132
Figura 13	Modelo de transmissão/aquisição de L1 .....	132
Figura 14	Incrementação estrutural da interlíngua .....	156
Figura 15	Processamento-S e Processamento Frásico .....	158
Figura 16	<i>Spell-out</i> em morfologia distribuída .....	188
Figura 17	Modelo da morfologia distribuída .....	189
Figura 18	Estrutura minimalista da concordância sintagmática: deslocação do objecto para receber caso .....	194
Figura 19	Estrutura do SDET: parte funcional e parte lexical .....	198
Figura 20	Estrutura do SDET: nó funcional SNUM .....	202
Figura 21	Representação funcional da estrutura do SN .....	202
Figura 22	Movimento de N para NUM .....	226
Figura 23	Sintaxe na OP para concordância .....	234
Figura 24	Sintaxe da concordância sujeito/verbo (PA e PB) .....	236
Figura 25	Sintaxe da concordância interna ao SN (PA e PB) .....	236
Figura 26	Sintaxe do possessivo núcleo para a concordância .....	237
Figura 27	Fossilização temporária e fossilização permanente .....	247
Figura 28	Modelo de aquisição em situação de contacto de línguas .....	283
Figura 29	Inglês: ausência de Mov-V .....	288
Figura 30	Português e francês: Mov-V-para-I .....	288



Figura 31	PA: Mov-V-para-I .....	288
Figura 32	CCV: Mov-V-para-I .....	288
Figura 33	Configurações de Mov-V para diferentes línguas .....	289
Figura 34	Estrutura do SN na FLO da aquisição .....	528
Figura 35	Romeno: movimento de N e consequente fixação de DET .....	529

## INTRODUÇÃO

“Não há nada mais perigoso que uma ideia, quando só temos uma.”  
(Paul Claude)

As línguas que emergem em situação de contacto linguístico têm fornecido material de valor inquestionável para o avanço dos estudos de um dos temas mais debatidos na actual linguística: a aquisição. O fascínio pelos vernáculos, que cedo marcou presença em estudos sobre línguas,<sup>1</sup> ganhou dimensão acrescida no Brasil, país no qual a língua portuguesa é o reflexo de uma situação sociolinguística bipolarizada (Lucchesi, 1994, 1996, 1998a, 1999, 2000a, 2001; Baxter & Lucchesi, 1997), dividida entre o padrão mais normatizado e o padrão que exhibe concordância variável, afectando estruturas das variedades de fala mais populares. No caso concreto da CPL-var registada entre os elementos que compõem o SN, a mesma é apontada como um fenómeno geral e categórico<sup>2</sup> dos mais marcantes do português do Brasil (PB), e que o distingue do português europeu (PE). O fenómeno, tem também sido amplamente estudado (p.e. Braga & Scherre, 1976; Braga, 1977; Scherre, 1978, 1988; Ponte, 1979; Nina, 1980; Guy 1981a; Fernandes, 1996; Carvalho, 1997a; Lopes 2001, P. Ribeiro, 2003; Tieppo, 2003; Baxter, 2009), originando um debate aceso acerca das motivações que o determinam. Sucintamente, a discussão em torno destas motivações remete, actualmente, para quatro hipóteses de condicionamentos:

- (i) A hipótese de influências das línguas de substrato africano, levantada por Guy (1981a) e que sugere que essa variação provirá de um ou mais substratos. Assim sendo, o cenário apontará para a possibilidade de o PB se estar a descrioulizar (Guy, 1981a, 1989; Houaiss, 1985; Holm, 1987) a partir de uma língua crioula histórica, influenciada por línguas africanas;
- (ii) A hipótese da transmissão linguística irregular (*imperfect language shift*),<sup>3</sup> postulada por Baxter & Lucchesi (1997), e que propõe que a mudança de língua e reestruturação se ficará a dever a um forte *input* de dados de PtgL2 (*transmissão irregular*) no processo de aquisição de L1, criando uma crioulação leve;
- (iii) A hipótese oposta às primeiras, isto é, a questão da deriva românica, avançada por Naro (1981), Scherre (1988) e Naro & Scherre (1993, 2000, 2007a), ambos

defendendo que as derivações direccionadas de fenómenos históricos internos à própria língua estarão na origem da variação;

- (iv) Por último, a hipótese mais ecléctica avançada por Holm (1998; 2004), que propõe explicar a formação das línguas reestruturadas com uma matriz sociolinguístico-histórica, assimilando as três perspectivas anteriormente apontadas: (1) a deriva; (2) a transmissão linguística irregular, com influências das L1's ancestrais (Winford, 2003a); (3) a influência ou influências de línguas pidgin ou línguas crioulas.

O comportamento do fenómeno da concordância nominal no SN do PB popular é marcado por redução das categorias morfológicas e pode ser conotado ao das línguas africanas (Guy, 1981a:301-302) do grupo níger-congo atlântico (Baxter, 2009:292), que o transmitiram também aos CP's atlânticos e variedades reestruturadas de português africano. Assim, é importante que se considere, neste trabalho, o funcionamento da regra da concordância nominal de número não só em outras línguas resultantes de um processo intenso de transmissão linguística irregular, tenham elas CP's atlânticos (português da comunidade de Almoxarife, São Tomé, e português cabo-verdiano) ou línguas do grupo níger-congo atlântico (português vernáculo de Angola, português do Maputo e português dos Tongas, São Tomé), como substrato directo. Nesta conformidade, foram já levados a cabo estudos comparativos, numa perspectiva sociolinguística, sobre o português dos Tongas (PT) (Baxter 2004, 2009), o PA (Figueiredo, 2008, 2009a, 2009b), o português do Maputo (PMp) (Jon-And, 2008) e o português de Cabo Verde (PCV) (Jon-And, 2009). Nestes trabalhos, tem sido apontado um perfil de variação e funcionamento dos factores condicionantes estruturais semelhantes aos do PB. Portanto, de certa forma é corroborado o pressuposto de Guy (1981a), quanto aos mecanismos responsáveis pela CPL-var no SN, quando o português é adquirido por falantes de uma L1 do grupo níger-congo atlântico ou de um CP. De facto, os ancestrais falantes das variedades de português referidas poderão, durante a fase de aquisição do PtgL2, ter associado os determinantes do português aos classificadores pré-nominais que controlam a pluralização no SN das línguas do grupo níger-congo atlântico. Em resultado de tal, a CPL-var de tais variedades, a exemplo do

que acontece no português vernacular do Brasil (PVB), é fortemente condicionada pelo item colocado na posição imediatamente anterior ao núcleo do SN.

Face às questões expostas, a pesquisa que aqui se propõe levar a cabo pretende aprofundar e complementar os estudos já iniciados por nós sobre a CPL-var no SN da variedade reestruturada de português falada pelos almozarifanos, membros bilingues em santomense e português (reestruturado) de uma comunidade semi-isolada de São Tomé, que descende de antigos escravos resgatados em África e enviados para a ilha, alguns dos quais, possivelmente, remetidos depois para o Brasil. O PA, que tem como substrato um CP do Golfo da Guiné, o santomense, formou-se no bojo do contacto entre falantes deste e do português,<sup>4</sup> que determinou uma situação de transmissão linguística irregular no processo de nativização do PtgL2 falado pelos primeiros.

Um dos aspectos centrais no âmbito dos estudos sobre ASL e reestruturação de línguas emergentes em contexto de transmissão linguística irregular prende-se com a natureza altamente variável evidenciada em muitos dos elementos que compõem a cadeia frásica: utilização de determinantes, aplicação da concordância de género e número nos constituintes do SN, uso de preposições, construções relativas, categorias de Tempo, Modo e Aspecto (TMA), etc. Algumas das questões que determinaram a selecção da variação da concordância de número entre os elementos flexionáveis do SN no PA, como objecto primordial do nosso estudo, foram já avançadas em Figueiredo (2008, 2009a, 2009b). Contudo, continua por se precisar até que ponto certos factores linguísticos e extralinguísticos poderão determinar o actual estado de variação registado no PA. À medida que progredimos na análise dos fenómenos que determinarão a CPL-var no SN, mais convictos ficámos de que o processo de aquisição da marcação e desenvolvimento das regras da concordância PL reflecte uma intrincada correlação de forças não só entre características tipológicas das L1's e línguas-alvo em contacto mas também entre todas estas línguas e factores de aquisição universais, independentes das mesmas.

Desta forma, a nossa análise combina, essencialmente, vários aspectos da gramática generativa (Teoria dos Princípios e Parâmetros da Gramática Generativa), para explicação da aquisição da concordância nominal, e da sociolinguística quantitativa, para apuramento dos pesos relativos dos contextos linguísticos e extralinguísticos na

estruturação do SN e desenvolvimento das suas regras de concordância. Paralelamente, o peso quer de modelos psicolinguísticos e cognitivos, como o Modelo dos Quatro Morfemas (*The 4M Model* – Myers-Scotton & Jake, 2000a, 2000b; Myers-Scotton, 2002), o modelo da mudança evolucionista da língua (Givón, 1998; Croft, 2000; Mufwene, 2001; Clements, 2009) ou o modelo da incrementação dos processamentos linguísticos assente numa hierarquia universal inata (*The Processability Theory* – Pienemann, 1988, 2005b), quer os pressupostos da MD, avançados por Halle & Marantz (1993), Costa & Silva (2006b) ou Noyer (2006), foram também por nós tidos em consideração. O ecletismo que decidimos adoptar foi determinado pelo propósito de se observar até que ponto a variação inerente à fala dos almoxarifanos, um grupo de indivíduos que partilha a mesma L1, reflecte ou não aquisição de acordo com os princípios da GU, e se a variação actual resulta de reestruturação determinada por gramáticas em conflito. Por outro lado, o método utilizado na recolha de dados seguiu as premissas propostas por Labov (1972a, 1983) e Weinreich, Labov & Herzog (2006 [1968]), que apontam o discurso espontâneo de uma dada língua como o *corpus* ideal para os estudos e análises em linguística, visto contribuir de forma incontornável para ajudar a compreender cinco questões problemáticas e fundamentais acerca da variação registada na fala das comunidades: (i) a problemática da restrição (*constraints problems*), que estabelece os limites entre variação e mudança; (ii) a problemática da transição (*transition problem*), que determina a ocorrência de diferentes fases de variação até se consumir o processo de mudança; (iii) a problemática do encaixamento (*embedding problem*), que observa a relação estabelecida entre mudança e factores da estrutura sociolinguística; (iv) a problemática da avaliação (*evaluation problem*), que analisa a atitude dos falantes relativamente à variação; (v) a problemática da implementação (*actuation problem*), que se debruça sobre as potenciais causas conducentes à mudança linguística.

Paralelamente, a observação do desenvolvimento dos factores identificados como condicionadores da aquisição da regra de concordância no processo de transmissão geracional permitirá uma conclusão acerca do perfil da concordância e sua influência na variação registada a nível do SN na fala actual de Almojarife. A comparação com os perfis registados noutras comunidades ajudará também a coligir, por certo, elementos

que permitirão confirmar se o português adquirido em situação de contacto por falantes de um CP manifesta padrões de variação semelhantes aos do português resultante de uma situação de transmissão linguística irregular na qual a língua ancestral não é um CP (Figueiredo, 2008; 2009a). Constatada esta premissa, poder-se-á trazer mais alguma luz a um dos dilemas que tem preocupado a actual crioulistica: terão as variedades parcialmente reestruturadas em situação de contacto sido desenvolvidas por influência de uma língua crioula (ou por línguas crioulas), ancorando a CPL-var no SN em mecanismos sintácticos dos substratos (Guy, 1981a:3001-302), ou, em contrapartida, ficaram a dever-se a processos internos à própria língua, independentes de tal influência, como a deriva românica (Naro, 1981; Scherre, 1988; Naro & Scherre, 1993, 2000, 2007a)?

Tendo como entendimento o pressuposto de que a ASL por adultos implica um acesso parcial à GU (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), debateremos algumas diferenças registadas a nível da aquisição de L1's e L2's, enfocando questões como a aquisição das categorias funcionais. A escolha de uma comunidade específica que produz variação relativamente uniforme na sua fala foi também determinada pelo propósito de se utilizarem os dados levantados como *corpus* de controlo, para ajudar a fornecer resposta à questão que Figueiredo (2008; 2009a) tem vindo a constatar, ou seja, que o português adquirido em situação de contacto por falantes de um CP manifesta padrões de variação semelhantes aos do português resultante de uma situação de transmissão linguística irregular em que a língua ancestral não é um CP. Assim julgamos estarem reunidas condições para melhor se identificarem as características linguísticas da comunidade, uma vez que este tipo de produção tem a vantagem de evitar artificialidades que adulteram os resultados dos dados submetidos a análise. Efectivamente, as limitações impostas à fala em tempo real impedem que, por exemplo, os falantes mais escolarizados possam assumir um controlo metalinguístico nas suas produções.

A análise do desempenho individual de cada um dos informantes foi feita com recurso ao pacote GOLDVARB X, que permitiu identificar não só as variáveis linguísticas e extralinguísticas que influenciam a marcação plural nos constituintes do SN da fala de Almojarife mas também que a variação registada a nível da concordância

dos itens flexionais é sistemática. A aplicação ou ausência das regras de concordância na sintaxe aparente do PAL1 poderá estar relacionada quer com a tipologia do santomense quer com a tipologia dos substratos africanos deste. Não obstante, se esta fosse a única justificação para a variação, não se entenderia o porquê de os falantes almorávidas aplicarem regularmente a regra em determinados contextos. Deste modo, tentaremos verificar não só se a variação em análise resulta da interacção complexa entre gramáticas da L1 e da LA mas também avaliar a forma como os princípios universais e psicolinguísticos subjazem à construção das referidas gramáticas.

O número de variedades de português africano e brasileiro a que recorreremos para estudos comparativos com o PA, o extenso quadro teórico que apoiou o nosso estudo e a metodologia pela qual optámos produziram, em agregado, uma quantidade elevada de tabelas e dados para análise. Desta forma o nosso estudo alcançou dimensão considerável, facto que, reconhecemos, o torna bastante volumoso. Depois de reflectirmos sobre este aspecto, optámos por não reduzir muito a dimensão inicial do trabalho, pois, ao fazê-lo, teríamos que retirar também achados que consideramos importantes para estudo da CPL-var numa perspectiva comparada. Desta forma, elaboraram-se cinco capítulos, cabendo ao primeiro destes estabelecer o pano de fundo sociohistórico da comunidade de Almorávida e apresentar também uma resenha dos contextos sociohistóricos das comunidades cujas línguas foram já observadas em estudos anteriores, os quais servirão para efectuarmos a nossa análise numa perspectiva comparada. Nesta parte do trabalho, dirigiremos a atenção para particularidades sociohistóricas que determinaram a aquisição do português em determinadas regiões do Brasil (Rio de Janeiro, Salvador, Helvécia e Rio de Contas) e de África (Angola, Moçambique, Cabo Verde e São Tomé). A identificação das condicionantes sociolinguísticas que poderão ter determinado o desenvolvimento da CPL-var no SN de todas essas variedades de português, com especial relevância para a da comunidade de Almorávida, objecto nuclear do nosso estudo, permitirá ver que tais condicionantes se tocam em muitas das suas especificidades, sugerindo transferência morfossintáctica, que se manteve depois por transmissão irregular. A fim de melhor se observarem estes fenómenos, o capítulo finaliza apresentando alguns dos traços gerais que caracterizam a fala de Almorávida e listando alguns paralelismos morfossintácticos, a nível do SN, que

os mesmos estabelecem com variedades de português (brasileiras e africanas), CP's atlânticos e línguas africanas, com especial incidência para as línguas do grupo níger-congo atlântico.

No capítulo 2 serão apresentados os pressupostos teóricos do nosso trabalho, elaborando-se não só um panorama acerca das teorias de aquisição de L1, bilingue e L2 mas também da forma como determinados estudos da actual Sociolinguística Variacionista têm procurado estabelecer a relação entre o comportamento linguístico dos falantes que integram uma dada comunidade de fala (língua-E) e a sua competência linguística enquanto usuários nativos de uma determinada língua natural (língua-I). Neste capítulo daremos ainda conta das perspectivas gerais sobre a variação na L2 e causas que poderão determinar fenómenos de fossilização. Estes aspectos implicam que se abordem também as questões em torno da morfossintaxe do SN no PE, PB e PA, observados numa perspectiva comparada. Apresentaremos ainda o quadro teórico da concordância a nível frásico e a nível do SN, abordando os pressupostos acerca da *interface* entre módulos gramaticais, com enfoque nas abordagens dos Princípios e Parâmetros da Gramática Generativa (Chomsky, 1957, 1965), Minimalismo (Chomsky, 1996 [1995]) e MD (Halle & Marantz, 1993; Costa & Silva, 2006b; Noyer, 2006). Sem deixar de levar em conta as diferenças registadas a nível da concordância plural no SN do PE e do PB, comentaremos as perspectivas acerca da aquisição das categorias funcionais em L1's e em L2's e condicionamentos que determinam a CPL-var no SN das segundas. Estes condicionamentos conduzem-nos, então, às questões em torno da transmissão linguística irregular e sua determinação na reestruturação de línguas, nomeadamente no desenho da configuração sintagmática do PA.

Quanto ao capítulo 3, será dedicado à descrição da metodologia aplicada na recolha, selecção, codificação e tratamento dos dados. Neste capítulo apresentaremos e descreveremos ainda o tipo de variáveis linguísticas e extralinguísticas a que recorreremos para postular hipóteses acerca das motivações que poderão determinar a CPL-var no SN do PA. No mesmo capítulo faz-se, ainda, uma descrição acerca do suporte estatístico computacional a que recorreremos para quantificar e analisar os dados.

Os resultados da análise serão apresentados em dois capítulos. No capítulo 4, tratámos os dados e analisámos os resultados das variáveis dependentes e dos grupos de



fatores linguísticos do tipo estrutural, comparando os nossos resultados com os de outros trabalhos sobre a CPL-var no SN de variedades brasileiras e africanas de português. No capítulo 5, recorreremos a idêntica metodologia para tratamento de dados e análise de resultados das variáveis linguísticas dos tipos semântico e fonológico, bem como das variáveis extralinguísticas.

Após o capítulo 5, serão apresentadas as conclusões do estudo. Determinou-se que a variação da concordância de número entre os elementos flexionáveis do SN é sistemática e orientada por contextos linguísticos específicos. Estes aspectos permitem prever em que tipo de estruturas é que os aprendentes do PtgL2 que tenham línguas crioulas como L1 revelam propensão para colocar, ou não, as marcas formais de PL. Os resultados permitiram concluir também que a aplicação da regra de concordância no PA se encontra em estágio de variação estável, fazendo antever uma situação de mudança linguística em progresso. Paralelamente, verificou-se que o português adquirido em situação de contacto por falantes de um CP manifesta os mesmos padrões de variação do português adquirido numa situação de transmissão linguística irregular na qual a L1 ancestral não é um CP (Figueiredo, 2008; 2009a).

Dada a quantidade de páginas produzida, e para facilitar o manuseamento do trabalho, optámos por dividi-lo em duas partes, estando incluídos no primeiro volume a introdução e os três primeiros capítulos. Quanto ao segundo volume, dele fazem parte os capítulos 4 e 5, dedicados à análise dos resultados, bem como as conclusões do estudo e a lista bibliográfica consultada. No final do texto de cada um dos volumes, encontram-se inseridas as notas de fim de texto respeitantes ao mesmo.

Resta apenas referir que, durante o período que decorreu a presente investigação, a mesma permitiu ao autor produzir os seguintes trabalhos:

#### Participação em congressos:

Figueiredo, Carlos Filipe Guimarães. 2006. *O sintagma nominal no português reestruturado de Almojarife (São Tomé)*. Congresso Anual da Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBPLe). Universidade de Coimbra, Portugal, 26-28 de Junho.

—2008. *Encaixamento linguístico da configuração do sintagma nominal plural do português reestruturado da comunidade de Almojarife (São Tomé)*. Congresso Anual da Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBPLe). Universidade Nova de Lisboa, Portugal, 03-05 de Julho.

- 2009. *Concordância plural variável no sintagma nominal do português reestruturado de Almojarife: motivações morfofonológicas*. Joint Summer Meeting of the Society of Pidgin and Creole Linguistics (SPCL) and the Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBPLE). Universidade de Colónia, Alemanha, 11-15 de Agosto.
- 2010. *Paralelismos morfossintáticos em variedades bantu, crioulos atlânticos de base portuguesa e variedades africanas e brasileiras de português: transferência ou trajecto universal de aquisição?* 10ème Colloque International de l'Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBPLE). Llacan, Campus CNRS de Villejuif, Paris, France, 1-3 de Julho.

Publicações:

- Figueiredo, Carlos Filipe Guimarães. 2008. *A concordância variável no sintagma nominal plural do português reestruturado de Almojarife (São Tomé)*. *Papia – Revista Brasileira de Estudos Crioulos e Similares*, 18. 23-43.
- 2009. A configuração do SN do português reestruturado da comunidade de Almojarife – São Tomé [em linha]. *RCBPLE – Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, 1(1). 28-55. Disponível em:  
<http://rcblpe.sftw.umac.mo/doc/A%20Configura%C3%A7ao%20do%20SN%20Plural%20do%20Portugu%C3%AAs%20Reestruturado%20da%20Comunidade%20de%20Almojarife%20-20Sao%20Tom%C3%A9.pdf>.



## CAPÍTULO 1

*“O mais grave nos negros não-cristãos não é serem selvagens. Mas é estarem convertendo em selvagens os portugueses que são brancos e cristãos.”*

(In, Mia Couto, *O Outro Pé da Sereia*, p. 229)

### O tema

O debate em torno da CPL-var no SN do PB tem motivado estudos em sociolinguística variacionista, principalmente no Brasil (p.e. Braga & Scherre, 1976; Braga, 1977; Scherre, 1978; Ponte, 1979; Nina, 1980; Guy 1981a; Scherre 1988; Fernandes, 1996; Carvalho, 1997a; Lopes 2001, P. Ribeiro, 2003; Tieppo, 2003; Baxter, 2009). Alguns destes trabalhos têm procurado determinar se as motivações da CPL-var ancoram nas ancestrais línguas de substrato (Guy, 1981a:301-302), originando reestruturação devido a forte input de PtgL2 por transmissão linguística irregular (Baxter & Lucchesi, 1997), ou se, em contrapartida, têm origem em derivações direccionadas de fenómenos inerentes à própria língua portuguesa (Naro, 1981; Scherre, 1988; Naro & Scherre, 1993, 2000, 2007a). Estes aspectos determinaram que os estudos acerca da concordância variável, numa dimensão comparada, tivessem ganhado alento, analisando-se o fenómeno quer em textos do português medieval (Tarallo, 1996; Naro & Scherre, 2000) quer em variedades brasileiras de português (Lucchesi, 2000a; Lopes, 2001, P. Andrade, 2003, Baxter, 2009) quer ainda em variedades africanas de português (Baxter, 2004, 2009; Figueiredo, 2008, 2009a, 2009b; Jon-And, 2008, 2009).

Neste capítulo faremos uma resenha de alguns trabalhos elaborados para análise do fenómeno da CPL-var no SN em determinadas comunidades linguísticas, observando-se também o pano de fundo sociohistórico desta. Posteriormente, apresentaremos os traços mais marcantes que caracterizam a fala dos membros da comunidade de Almojarife, São Tomé, listando alguns paralelismos morfossintácticos que o PA estabelece, a nível do SN, com variedades de português (brasileiras e africanas), CP's atlânticos e línguas africanas do grupo níger-congo atlântico.

#### 1.1. Português do Brasil e concordância nominal variável de número

O PB apresenta diversidade acentuada relativamente ao PE, sobretudo a nível dos falares populares e rurais, aspecto que, desde cedo, chamou a atenção de dialectólogos e

linguistas. Os estudos científicos sobre o fenómeno, sobretudo na década de 70 e início dos anos 80 do século XX, permitiram concluir que tal diversidade reflecte, em si, a própria sociohistória do país. Alguns linguistas chamam a atenção para a inapropriação das gramáticas normativas no Brasil (Lucchesi, 1994, 1998a; Bagno, 2001, 2002), subjugadas às regras teórico-descritivas europeias: *“Fundamentam-se as regras da gramática normativa nas obras dos grandes escritores, em cuja linguagem as classes ilustradas põem o seu ideal de perfeição, porque nela é que se espelha o que o uso idiomático estabilizou e consagrou”* (Lima, 2002 [1957]:7). Como as produções escritas de carácter literário, tomadas como exemplo, não reflectem o registo oral espontâneo, Bagno radicalizaria o seu discurso, lançando aos pesquisadores o repto de determinarem se no Brasil se fala *“português ou brasileiro”* (Bagno, 2001:19).

De entre os fenómenos morfossintácticos do PB, as concordâncias variáveis nominal e verbal evidenciam distinção acentuada em relação às regras categóricas padronizadas. Dado que a variação apresenta diferenciados graus de aplicação, consoante os falantes sejam usuários da norma culta ou da norma popular, não surpreende que os linguistas a tomem como elemento privilegiado de análise, a fim de determinarem o modo como o fenómeno se encaixa na gramática e na sociedade, ou seja, quais as motivações que o determinam e a forma como ele é percebido socialmente. No concreto, os estudos acerca da concordância variável nominal de número não têm evidenciado consenso sobre as motivações que a impulsionam. Basicamente, alguns estudiosos, defensores da linha contactista, entendem que o desenvolvimento do PL no SN tem, no seu bojo, a intervenção de processos funcionalmente irrestritos de crioulização. Estes, por sua vez, levaram a que a variação na concordância do PB fosse determinada por fenómenos morfossintácticos indiciadores da ocorrência de reestruturações colaterais, impulsionadas pelo contacto massivo entre línguas que marcaram a sociohistória do Brasil (Baxter & Lucchesi, 1997; Lucchesi, 1998a, 2001), mas que teriam uma fonte comum: os ancestrais dialectos africanos falados pelos escravos africanos transplantados (Guy, 1981a, 1989, 2005; Holm, 1987, 1992; Baxter & Lucchesi, 1997, 1998; Lucchesi, 2000a; P. Andrade, 2003, Baxter, 2009).<sup>5</sup> Por seu lado, os defensores da linha derivista sugerem que a variação é consequência de o português estar a trilhar um trajecto próprio no Brasil, demarcando-se morfossintacticamente do PE, quer por incrementar o uso do

pronome na posição de sujeito e deixar de aplicá-lo no lugar do objecto quer por o seu sistema de concordância nominal e verbal estar a ser afectado por mudanças fonéticas naturais (Révah, 1959), determinadas, principalmente, pela deriva diacrónica (Naro & Lemle, 1976; Scherre, 1988).<sup>6</sup> Deste modo, o padrão de concordância variável do SN no PB terá sido fornecido por variedades do PE (Naro & Scherre, 1993, 2000, 2007a, 20007b). Contudo, no Brasil, o fenómeno apresentará uma extensão diferente, por ter sido intensificado e acelerado pelo contacto do português com as línguas indígenas e africanas, ou seja, em virtude de os falantes do PtgL2 terem exacerbado o modelo europeu (Naro, 1981; Naro & Scherre, 1993, 2000, 2007a). Ainda que assim seja, o facto de a aprendizagem do PE (modelo sem variação) por contacto produzir uma L2 com variação, é suficiente, por si só, para concluir que não faz sentido incidir atenção sobre a questão da deriva. Com efeito, e independentemente de este fenómeno ocorrer ou não no PE, o contacto irá originar sempre um produto final marcado por variação.

Por outro lado, é também importante ter em consideração que a noção de deriva latina foi formulada por filólogos numa época em que os fenómenos de contacto eram pouco compreendidos, o que levava a atenção dos estudiosos a centrar-se na língua-I (*I-language*).<sup>7</sup> Por conseguinte, sendo as línguas miscigenadas encaradas como deturpações da LA<sup>8</sup> envolvida no contacto, as questões extralinguísticas não eram consideradas essenciais para os estudos em diacronia. Este aspecto determinou que a fonética ganhasse grande protagonismo no final do século XIX e princípio do século XX, tendo o Positivismo dos neogramáticos introduzido a hipótese da existência de leis fonéticas de carácter universal e absoluto. Com estas leis aplicadas cegamente sobre os sons, as mudanças linguísticas eram explicadas de forma idêntica para todas as línguas, originando uma quase total ignorância acerca dos processos de mudança e aquisição de língua em estudos linguísticos (Baldinger, 1972 [1958]; Vänänen, 1981 [1957]; Williams, 1986 [1938]), até à sexta década do século XX. Contudo, e sendo inquestionável que todas as línguas são afectadas por mudanças de geração para geração, mesmo em contextos monolíngues, a verdade é que tais estudos não negam as interferências fonológicas e adopção de elementos lexicais dos substratos (Branco, 1984:7). É o que acontece, por exemplo, com o fenómeno da lenição (abrandamento) das consoantes oclusivas em posição intervocálica, que afectou e simplificou, na zona da

România Ocidental, o sistema de consoantes duplas do latim falado, muito por força do seu contacto com o substrato celta. Este fenómeno despoletou uma série de outros em cadeia, com vista a evitar situações de homonímia e homografia, e que, nalguns casos, chegou a culminar com o aparecimento de novas consoantes, como acontece, por exemplo, com a consoante /v/, inexistente no latim clássico (p.e. [-pp- > -p- > -b- > -ɸ- > -v-]). Outro fenómeno identificado de interferência substratal ocorreu com a conservação do [l] e do [n] intervocálicos em determinadas palavras no galego-português, em virtude do seu contacto com o moçárabe (p.e. *canito*; *male*) (Branco, 1984: 15). Quanto a incorporações lexicais dos substratos, poder-se-á apontar, por exemplo, o uso do sufixo celta *-occus* no latim, em combinação com um nome, geralmente um deus (p.e. *Indovellicus*) (Branco, 1984:13).

No entanto, cabe aqui questionar se os fenómenos de influências ou transposição se limitaram aos aspectos meramente fonológicos ou lexicais. Thomason & Kaufman (1988:37) identificam duas tendências gerais nos fenómenos determinados por contacto, cuja extensão depende dos factores prevaletentes neste: a transferência e o empréstimo. Para acontecer a primeira, é necessário que os falantes envolvidos no contacto linguístico sejam monolíngues. Estes, à medida que vão aprendendo a LA, transferem estruturas da língua nativa para a língua na qual procuram comunicar (Thomason & Kaufman, 1988:73-74), originando a emergência de uma língua que se distingue quer do substrato quer da LA. Posteriormente, e caso o contacto se mantenha, a nativização da nova língua, que geralmente ocorre entre três e seis gerações (Clements, 2009:34), é também marcada por reestruturação assente em empréstimos, uma vez que os falantes submetidos ao contacto se tornam bilingues (Thomason & Kaufman, 1988:73-74). No caso concreto da transição do latim para o português, apesar da tenaz oposição do povo indígena da Ibéria Central e Ibéria do Noroeste aos romanos, há registo de existirem, nestas regiões, durante o período entre 100 aC e 300, falantes bilingues em variedades celtas e latim (Tovar, 1977:124). A grande maioria destes seria analfabeta, o que minorava a recepção de *input* negativo acerca das produções da LA que evidenciavam variação. A intervenção dos bilingues neste extenso período de contacto linguístico terá sido determinante para a manutenção da morfologia pré-romana nos romances ibéricos central e ocidental. Nas variedades celtiberas, os nomes com vogal temática em *-o*

formam o nominativo PL em *-os*. Este traço entrou, por transferência, no sistema da segunda declinação dos nomes pré-romanos e latinos, com vogal temática em *-us* (*-us, -i, -o, -um, -o; -i, -orum, -is, -os, -is*). Posteriormente, e por empréstimo, os nomes temáticos em *-o* do português passaram a registar o PL actual em *-os* (o sufixo nominativo PL do latim) (Clements, 2009:34). Como se verifica, a variação que configura a mudança não estará apenas dependente de fenómenos inerentes à própria língua, como a deriva românica, já que pode ser também consequência de outros aspectos, como o contexto em que se dá a aquisição, o grau de similaridade estrutural das variedades em contacto, o tipo de falantes envolvidos neste e até a sua atitude comunicativa.

No caso concreto do Brasil, a visão teocêntrica do português quinhentista demonstra, logo desde o achamento, a manifestação de um preconceito que, durante séculos, discriminaria dialectos cujas tipologias fossem distintas da sua L1. Em *A Carta*, por exemplo, é notório o paralelismo que o cronista real Pêro Vaz de Caminha estabelece entre as práticas sociolinguísticas dos ameríndios e o paradigma europeu: “... *não pôde deles haver fala nem entendimento que aproveitasse*” (Seabra, 2000).<sup>9</sup> Por seu lado, situações como a implantação tardia do ensino no país concorrerem para que a unidade construída em Portugal para o seu idioma não ocorresse também no Brasil. Neste, a sociedade é altamente estratificada e regista um *continuum* linguístico bastante marcado, que vai desde as variedades vernáculas presentes nas comunidades rurais isoladas e quase analfabetas até às formas populares urbanas ou culta, usadas nos centros cosmopolitas.

Estudos de finais do século XIX (Coelho, 1880/1881, 1882, 1886) dão-nos conta de diferenças registadas entre o PB e o PE, sem, contudo, apontarem qualquer tipo de variação no segundo. Tais diferenças fazem perceber uma situação de bipolarização no PB, instalada até ao final do século XIX (Lucchesi, 1994, 1996, 1998a, 1999, 2000a, 2001, 2009a; Baxter & Lucchesi, 1997), altura em que ocorreu o súbito processo de industrialização e urbanização do país. A situação de bipolarização ter-se-á iniciado quando, no final do século XVI, o quadro habitacional do Brasil se começou a modificar drasticamente com a instalação dos engenhos de cana-de-açúcar no Nordeste do território. De facto, das diferenças linguísticas nos davam já conta quer Silva Neto (1986



[1957]), que chega a citar o Padre António Vieira por, no século XVII, fazer menção ao avesso e direito do português no território, o primeiro falado pelos naturais e africanos e o segundo pela elite que zelava pelos padrões europeus. Nesta fase, são também comuns os relatos de alguns viajantes estrangeiros que fazem referência aos falares dos índios e escravos negros, impregnados de características crioulas. Posteriormente, uma série de acontecimentos sociohistóricos viriam a afectar a hegemonia e representatividade dos falares populares no Brasil: declínio dos engenhos do açúcar e consequente deslocamento de parte das massas escravas do nordeste para trabalhar nas minas de esmeraldas e ouro do sudeste da colónia; penetração nos sertões brasileiros do surto das bandeiras, composto por grupos paulistas que englobavam indivíduos de condição social baixa, fossem eles brancos, mestiços ou índios, aí erigindo povoados que evoluíram para grandes cidades e nas quais se movimentam também poetas, intelectuais e pequenos burgueses; novo deslocamento de escravos, desta vez para as roças de café que se iam espalhando pelos estados de São Paulo e do Rio de Janeiro, após esgotadas as reservas de minerais da zona de Minas Gerais; abandono das fazendas por parte dos libertos, após assinatura da Lei Áurea (1888) e extinção da escravatura, que se dirigem para os centros urbanos, exponenciando o crescimento vertiginoso destes e distendendo, diatopicamente, o contacto do PVB a contextos conservadores e europeizantes. Observada numa perspectiva sociolinguística, a entrada de escravos africanos no Brasil e posterior abolição da escravatura neste reflectiram-se extensamente no idioma do país, já que a distribuição dos falantes do PVB pelo país, traduzido na fixação e aculturação da população de origem africana um pouco por todo o território, acabou por impor não só as variantes vernáculas do português sobre a língua geral inicialmente falada pelas populações autóctones mas também por converter o modelo defectivo do PtgL2, falado por escravos, imigrantes pobres e colonos afastados dos contextos linguísticos urbanos, em L1 dos seus descendentes, fossem eles pedógamos ou mestiços. Contudo, ao verem dificultada a promoção da sua cidadania, os ex-escravos que aportaram aos centros cosmopolitas acabaram por buscar a ascensão social recorrendo a diversos tipos de estratégias, entre as quais se poderá destacar a abertura à pressão linguística, traduzida na introdução de características da norma culta no português popular. Ainda outro aspecto a ter em linha de conta, enquanto fenómeno originado pelo êxodo dos ex-

escravos para os centros cosmopolitas, foi o facto de a sociedade brasileira deixar de assentar grandemente em padrões de características rurais e transitar, em finais do século XIX, para uma sociedade do tipo urbano.

As mudanças verificadas na sociedade brasileira em finais do século XIX viriam também a ser incrementadas pelo rápido desenvolvimento de estruturas de diversos tipos, as quais afectariam, directa ou indirectamente, todo o seu sistema linguístico-cultural. O ensino básico, apesar de precário, massificou-se e a população menos elitista viu também facilitado o acesso aos meios de comunicação social. Quanto à rede viária, ramificou-se por todo o país e as grandes empresas capitalistas ligadas às áreas da mineração, extracção madeireira e produção de celulose expandiram-se pelo interior, incrementando a exploração dos recursos naturais do país. Submetidas à pressão deste jogo de forças, a norma popular e a norma culta foram estreitando laços, cedendo ambas nos seus padrões mais marcantes e caminhando para um patamar de equilíbrio das suas características linguísticas. A primeira deu passos no sentido de aproximar as suas particularidades às do padrão urbano culto e fazendo com que se diluíssem as marcas mais salientes das alterações resultantes do extenso contacto linguístico ocorrido nos séculos antecedentes. E a este processo de mudança não escaparam também os contextos linguísticos mais afastados e isolados, que viram a sua cultura ser desarticulada e o seu modo de vida radicalmente alterado. No entanto, os estudos empíricos levados a cabo nestas comunidades evidenciam que, apesar de terem perdido as suas marcas mais crioulizantes, as suas falas continuam a revelar características que permitem conotá-las aos processos de transmissão linguística irregular que determinaram a sua emergência.

Quanto à norma culta, sujeita à pressão quer dos falantes das classes imigrantes que se mudavam do campo para a cidade e se integravam na estrutura social quer dos defensores da demarcação em relação às normas europeizantes, via-se democratizada e sujeita a mudanças que faziam com que aproximasse também o seu registo do da norma urbano-popular que se ia implementando. A prova da penetração nos estratos sociais médios e altos de algumas das características resultantes da aquisição/nativização do português falado por ex-escravos e seus descendentes é dada, por exemplo, pela *“variação na concordância de número (tanto verbal, quanto nominal), que se teria originado no processo de transmissão linguística irregular ocorrido na base da*

*pirâmide das sociedades colonial e do Império, e se teria difundido tornando-se hoje um fenômeno que se verifica em todos os extratos da sociedade brasileira”* (Lucchesi, s.d.). Em consequência da pressão exercida, em simultâneo, sobre as normas popular e culta, atenuou-se o quadro marcadamente bipolarizado que havia imperado até então, caminhando-se para o actual panorama registado no país. Ainda assim, o ensino bastante tardio do português não conseguiu evitar que, no Brasil, emergisse uma situação de quase diglossia, cavando-se um fosso entre a língua da população miscigenada e de origem escrava, por um lado, e a das elites escolarizadas, por outro lado. Deste modo, a unidade construída em Portugal para o seu idioma não ocorreu no Brasil, onde a sociedade é altamente estratificada e regista um *continuum* linguístico bastante marcado, que vai desde as variedades vernáculas presentes nas comunidades rurais isoladas e quase analfabetas até à formas populares urbanas ou culta, patentes nos centros cosmopolitas. Como consequência das diferentes políticas de língua adoptadas, a norma do PE é diferente da do PB, e ambas sê-lo-ão também das outras que estão a germinar em todos os outros países de expressão portuguesa. Não obstante estes aspectos, os primeiros trabalhos sobre a concordância nominal de número no SN teimavam em apresentar uma visão tradicionalista deslocada da realidade do país, indicando-a como categórica (p.e. Lima, 2002 [1957]:305-306; Bechara, 1999 [1928]:544-545).<sup>10</sup>

Assim, a concordância variável de género e número no SN de variedades do PB continuou a ser apontada apenas em trabalhos de dialectólogos, como os de Amaral (1920), sobre o dialecto caipira, ou de Monteiro (1933) e Marroquim (1945 [1934]), sobre o português popular nordestino. Quanto aos linguistas, têm privilegiado a sociolinguística quantitativa como modelo de análise, uma vez que esta permite determinar o modo como o fenómeno se encaixa na gramática e na sociedade, ou seja, entender quais as motivações que o determinam e como ele é percebido socialmente. Mais recentemente, análises do mesmo tipo têm sido feitas em variedades africanas de português (Moreno & Tuzine, 1998; Baxter 2004, 2009; Figueiredo, 2008, 2009a, 2009b; Jon-And, 2008, 2009) e PtgLE em contexto de ensino (Godinho, 2005, 2008), abrindo o leque de possibilidades para o estudo do fenómeno da variação no SN numa perspectiva comparada.

No Brasil, o primeiro estudo sistemático sobre a CPL-var no SN no Brasil foi levado a cabo por Braga & Scherre (1976), que observaram o comportamento da concordância PL do PB falado na área urbana do Município do Rio de Janeiro (MRJ). Este trabalho teve como modelo base o estudo de Naro & Lemle (1976) sobre a concordância sujeito/verbo produzida por falantes semi-escolarizados, também da área urbana do Rio de Janeiro, e que frequentavam o curso de alfabetização de adultos. Na análise de Naro & Lemle (1976:240-241), é advogada que a mudança registada no PB é determinada pela deriva natural da língua, começando nos contextos menos salientes e atingindo, posteriormente, as formas mais salientes, por “imitação”. Por seu lado, Braga & Scherre (1976:474) concluíram que os itens de maior *saliência fônica* favorecem mais a inserção da pluralização, por oposição aos elementos terminados em *-s* no singular e que pluralizam com a inserção de *-es* final, que apresentam tendência para inibirem a marcação. A tendência para a não-marcação destes elementos deve-se, segundo as autoras, ao fenómeno da analogia,<sup>11</sup> comum em situações de aprendizagem que envolvem dados divergentes ou insuficientes, uma vez que os itens em questão apresentam já um *-s* final, que poderá ser entendido pelos falantes como uma marca de pluralização.

Seguidamente, Braga (1977) observou a concordância PL do português falado pelas classes média e baixa do Triângulo Mineiro (PMn), voltando a enfatizar que a *saliência fônica* exerce influência na marcação PL e que os itens terminados em *-s* a inibem, mas que estes comportamentos não são uniformes em todos os estratos sociais. Deste modo, Scherre (1978) avaliou a influência das variáveis sociais na marcação PL produzida por falantes de diversos níveis de escolarização do MRJ. Nesta análise, a linguista detectou que a *saliência fônica* propicia mais a concordância nos falantes escolarizados do que nos semi-escolarizados, apesar de ambos favorecerem positivamente a pluralização. Com base nas constatações de todos estes trabalhos, Scherre (1978:127) advogaria que a mudança emerge nos ambientes de menor *saliência fônica*, alastrando-se e alcançando, depois, os de maior *saliência*.

Ponte (1979:264-266), analisando o português de semi-analfabetos de Porto Alegre (PPA), Rio Grande do Sul, concluiu que o SN deste dialecto apresenta uma hierarquia de marcação PL idêntica à dos informantes semi-escolarizados do MRJ. Nina (1980), por

seu lado, observou os percentuais de pluralização realizados por informantes analfabetos da micro-região bragantina, Estado do Pará, tendo concluído que os itens na primeira posição da cadeia sintagmática são os que mais recebem marcas de PL. Constituindo a variável *grau de saliência fônica na oposição singular/plural*, Nina (1980:103) constatou que ela não é relevante para a marcação, mas que ocorre tendência para se pluralizarem os itens regulares e metafônicos. Por seu lado, os elementos que formam o PL com inserção de *-s* final e mudança silábica (p.e. *casal/casais*) apresentavam uma taxa de inibição total (100%).

Com recurso à variável *natureza flexional do elemento precedente*, Nina (1980:144-145) verificou também que os factores que mais favorecem a marcação são o numeral e os elementos sem marca formal de PL, mas com carga semântica de pluralização. Face aos resultados das variáveis linguísticas, a pesquisadora conclui que o português paranaense apresenta tendência para eliminar a concordância redundante, seja ela nominal seja ela verbal. Este aspecto vem ao encontro do pressuposto da hipótese funcionalista das Condições de Distintividade (Kiparsky, 1972:195), que advoga que a inserção de PL na primeira oportunidade restringe a marcação nas seguintes, em virtude de esta se revelar redundante. Por outro lado, as variáveis sociais constituídas por Nina (1980:152) – *sexo e idade* –, levaram a autora a constatar que as mulheres usam mais as marcas e que a regra de concordância se encontra em processo de aquisição pelos escalões etários mais novos.

Naro (1981:76), retomando a questão da possível relação entre maior frequência de marcação em número e maior saliência fônica, propõe que o peso desta na marcação passe a ser observada sob duas perspectivas, levando em conta a diferença fônica, por um lado, e a tonicidade, por outro lado. No trabalho, Naro (1981:97) insiste que a deriva subjaz às mudanças linguísticas, pelo que a variação registada no PB será motivada por tal fenómeno. Em contrapartida, Guy (1981a:189), no seu estudo sobre o português dos falantes semi-escolarizados do Rio de Janeiro, apesar de seguir a linha directriz proposta por Naro para a configuração dos grupos de factores, conclui que a *saliência fônica* não exerce efeito relevante na marcação PL.

Dado que o Brasil foi o principal destino dos escravos resgatados na região afro-atlântica, Guy (1981a), para identificar os contextos linguísticos e sociais que podem

determinar o apagamento do *-s* final, seja ele marca de PL, ou não, teve em consideração o facto de ter ocorrido contacto entre línguas europeias e africanas. De acordo com esta visão, a inserção de marcas é fonologicamente determinada pelas sílabas tónicas e pelo contexto fonológico seguinte, quando este é realizado por pausa ou pelos traços [–vozeado] e [velar]. Na cadeia linear, o item que mais recebe a marcação é o que se encontra na primeira posição, com todas as outras posições revelando tendência para a inibição. Constituindo a variável *informação de plural precedente*, Guy (1981a:179) refere que, caso ocorra um numeral antes do item analisado, a tendência é para inibir a marcação no elemento seguinte. Paralelamente, e tendo como base os estudos de Braga (1977) e Scherre (1978), Guy (1981a:184) observou também o comportamento da *saliência fónica* das categorias morfológicas, estabelecendo dois factores para o efeito: *palavras terminadas em vogal* e *palavras terminadas em consoante*. A conclusão é que as primeiras inibem a inserção de marcas, enquanto as segundas a favorecem.

A fim de testar a hipótese funcionalista das Condições de Distintividade (Kiparsky, 1972:195), Guy (1981a:191) constituiu as variáveis *correlação entre marcas de PL no SN e marcas de PL no verbo* (factores *verbo seguinte com marca de plural* e *verbo seguinte sem marca de plural*) e *efeito do local da informação adicional de plural*. A primeira das variáveis não confirmou a hipótese, mas Guy (1981a:192) justificou o facto referindo que os falantes, por uma questão de estilo, pluralizam quer o SN quer o SV, caso pretendam usar mais a marca de número PL. O predomínio de determinada gramática, em detrimento de outra, pode ser também explicada através do Princípio da Coesão Estrutural (Haiman, 1983:782), que determina que o falante revela tendência para constituir mentalmente grupos que compartilham traços, a fim de empregar itens linguísticos em consonância com a gramática ou código que selecciona para poder interagir nas diferentes situações de fala (cf. Lucchesi, 2000a:141). Quanto à variável *efeito do local da informação adicional de plural*, confirmou a hipótese funcionalista. Assim, existe correlação entre gramática e posição linear, daí a marcação PL revelar tendência para ser inserida na primeira posição da cadeia sintagmática e apresentar inibição da marcação nos lugares à direita desta posição. Já no que concerne às variáveis sociais,<sup>12</sup> Guy (1981a:240) considera que a variável *sexo* é aquela que merece mais

credibilidade, com as mulheres aplicando mais as marcas, mas em valores bastante próximos dos apresentados pelos homens.

Estabelecendo paralelos entre a variação do português e dialectos hispano-americanos, Guy (1981b:179) reafirma que a posição linear constitui a chave para a inserção de marcas de PL e estabelece conexões entre posição e classe gramatical. Com base nas suas constatações deste e de outros trabalhos posteriores, Guy (1981a; 1989; 2005) situa o eixo da marcação PL nos determinantes, estabelecendo um paralelismo configuracional entre o PB e as línguas dos ancestrais substratos bantu e kwa falados pelos escravos transplantados para o Brasil e seus descendentes. Desta forma, é de considerar a possibilidade de o PB se encontrar em processo de descrioulização (Guy, 1981a, 1989; Houaiss, 1985; Holm, 1987), a partir de uma língua crioula histórica, influenciada por estruturas bantu e kwa (Guy, 1981a, 1989, 2005).

Scherre (1988) reanalisa a concordância nominal do MRJ, discordando de Guy (1981a, 1981b) no que concerne à correlação entre gramática e posição linear. Fazendo uma revisão sobre os trabalhos que haviam abordado a questão da CPL-var do SN, a autora analisa, atomística<sup>13</sup> e sintagmaticamente,<sup>14</sup> as estruturas do SN, estabelecendo ainda paralelos com o espanhol de Porto Rico (EPR). Constituindo um considerável número de hipóteses acerca da variação do fenómeno da concordância, Scherre (1988) testa-as, recorrendo às potencialidades concedidas pelo programa computacional VARBRUL. Para além de confirmar que as formas salientes são as mais marcadas, a autora conclui ainda a favor da intervenção do Princípio do Processamento Paralelo,<sup>15</sup> defendendo que a presença de marcas nos itens do SN conduz a mais marcas nos restantes elementos, isto é, que *“marcas levam a marcas e zeros levam a zeros”* (Scherre, 1988:208). Posteriormente, Silva & Scherre (1998 [1996]) e Scherre (1998a [1996]) confirmariam este pressuposto, mas P. Andrade (2003:22) chama a atenção para o facto de as autoras não tomarem em consideração, para as suas conclusões, os valores exibidos pelos factores *ausência de marcas fora do SN* e *zero formal na primeira posição*, isto é, *“justamente nas posições em que, no português do Brasil, preferencialmente se marca o plural”*.<sup>16</sup> Assim, será também de considerar que a marcação no interior do SN possa ser orientada pelo Princípio da Coesão Estrutural

(Haiman, 1983:782; Lucchesi, 2000a:141), e que leva o falante a constituir, mentalmente, grupos que compartilhem traços.

Até ao estudo de Scherre (1988), as pesquisas efectuadas em torno da marcação no SN vinham postulando que a primeira posição deste era o factor preferencial para a inserção da marca de PL, enquanto as restantes posições inibiam a marcação. Ao constituir a variável *relação dos elementos não nucleares em função do núcleo e posição dos elementos no SN*, Scherre (1988:167) considera este pressuposto falacioso, visto ter verificado que a ausência de marca de PL estabelece uma relação inequívoca com as posições antepostas ao elemento nuclear. No entanto, P. Andrade (2003) e Baxter (2004) observariam que o modelo linear pode funcionar, caso a estrutura do SN seja simples, isto é, composta por apenas dois elementos.

Outra questão em que há divergência entre Scherre (1988, 1998a [1996]) e Guy (1981b:179) tem a ver com a relação estabelecida entre posições lineares e categorias gramaticais na influência da marcação PL, uma vez que o segundo defendera que os determinantes surgem, por norma, na primeira posição do SN, os nomes na segunda posição e os adjectivos na terceira posição. Scherre discorda, referindo que os determinantes favorecem a pluralização, enquanto os substantivos e adjectivos a inibem. Por outro lado, Scherre (1988:152-153, 1998a [1996]:94-97) entende também que as posições lineares não podem ser vistas como exclusivas de determinadas categorias gramaticais, já que há uma diversidade de classes gramaticais nas primeira, segunda e terceira posições do SN, com qualquer uma delas revelando tendência para ser marcada, caso surja no primeiro lugar da cadeia sintagmática. Desta forma, e independentemente da classe gramatical do item analisado, o que se tem é que os elementos favorecem ou inibem a inserção de marcas em virtude de se encontrarem pospostos ou antepostos ao núcleo do SN. Segundo Scherre (1988:47), este comportamento é determinado pelo Princípio da Iconicidade, que advoga existir uma relação directa entre a dimensão linguística ou formal e a dimensão não-linguística ou conceptual (Haiman, 1983:782).

As variáveis sociais observadas por Scherre – *sexo, grau de escolarização, faixa etária e ambiente de origem do falante* –, permitiram concluir que a CPL-var do PB se encontra em fase de variação sociolinguística estável (Scherre, 1988:520), uma vez que a influência da faixa etária na inserção de marcas não é relevante. Por fim, estas mesmas



variáveis levaram ainda a autora a discordar dos resultados de Guy (1981a:301-302) acerca da possibilidade de o PB popular estar sujeito a um processo de descrioulização. Assim, em trabalhos levados a cabo posteriormente sobre a concordância sujeito/verbo no PE e no PB, Naro & Scherre (1993, 2000) mostram-se em desacordo com a hipótese de terem ocorrido processos de transmissão irregular do tipo crioulizante no português popular do Brasil. Em contrapartida, e sem desconsiderarem o contexto sociohistórico do país, Naro & Scherre, (1993:437) assumem uma posição semelhante à de Révah (1959), reafirmando que as derivações direccionadas de fenómenos internos ao português estarão na origem da variação. Os linguistas chegam, posteriormente, a recorrer a textos medievais do português, em particular do período anterior à estabilização das normas gramaticais, para sustentarem o seu ponto de vista, apesar de, no total das estruturas analisadas, a ausência de concordância apresentar apenas 1% de ocorrências (Naro & Scherre, 2000:250). Se, por si só, os dados escritos não reflectem o registo oral espontâneo, ou seja, não constituem o *corpus* ideal para a investigação científica em sociolinguística variacionista e quantitativa, o facto de os fenómenos registados configurarem apenas 1% de ocorrências não torna confiáveis os resultados estatísticos apresentados.

Recorrendo também a dados do PB antigo escrito (cartas e textos de teatro) e do Projecto Norma Urbana Culta de Salvador (NURC), Tarallo (1996) recusa a hipótese da crioulização e posterior descrioulização, uma vez que o que está a acontecer no Brasil é um afastamento em relação ao PE, e não uma aproximação a este. O entendimento de que o fenómeno seria uniforme em todo o território brasileiro, acaba também por descartar a possibilidade de ter ocorrido transmissão linguística irregular na formação sociohistórica do PB, já que esta pressupõe um *continuum* de níveis diferenciados de socialização/nativização de uma L2. Por outro lado, e a exemplo do que sucedeu com o estudo de Naro & Scherre (2000), os textos seleccionados por Tarallo (1996) não se aprestam para identificação dos fenómenos resultantes do contacto entre línguas, tal como o *corpus* da NURC, apesar de ser oral, não será o mais indicado para a observação das características da fala popular (Baxter & Lucchesi, 1997; Lucchesi, 2001).

Mussa (1991a, 1991b), por seu lado, apresenta uma visão superstratista do PB, já que entende que este não pode ser visto como um resultado das influências dos substratos

africanos. Segundo o autor, não há diferenças formais entre os processos de criouliização e de mudança natural, pois são ambos caracterizados por simplificação fonológica e estrutural (Mussa, 1991b:76). Assim, o que aconteceu no PB foi uma competição entre formas diferentes dos dialectos em contacto, optando os falantes pelas mais simples e/ou menos estigmatizadas. Como idêntica estratégia terá sido também seguida pelos falantes do português africano, o facto de este partilhar traços com o PB não passará, pois, de mera coincidência (Mussa, 1991a:237).

Já Lopes (2001:102), estudando atomisticamente a concordância nominal das normas culta e popular usadas na cidade de Salvador, encontra evidências de que o contacto entre línguas poderá estar na base da concordância variável: *“processos variados de aquisição do português ocorreram no decorrer da história desse país, envolvendo aquisição de Português L1 com dados divergentes ou insuficientes oriundos de contactos com falantes que aprenderam essa língua como L2”*. Adoptando a metodologia de Scherre (1988), Lopes (2001) constitui uma nova variável social, a *etnia*, justificada pelo facto de a cidade de Salvador possuir uma população maioritariamente negra. A variável envasa dois factores (*falantes com sobrenomes religiosos e falantes com sobrenomes não-religiosos*) e a constituição dos grupos foi viabilizada pelos estudos da área de genética da Universidade da Bahia (UFBA), que permitiram estabelecer a relação entre o sobrenome e a etnia através de análises do tipo sanguíneo das pessoas pesquisadas. Com recurso a esta variável, a autora pôde confirmar que a aquisição das regras de concordância por parte dos indivíduos negros e afrodescendentes é inferior à dos indivíduos não-afrodescendentes, um aspecto que surge também intrinsecamente relacionado com as oportunidades educacionais.

O quadro teórico que suportou os resultados de Lopes (2001) baseou-se no pressuposto cognitivo dos Quatro Morfemas (*The 4M Model*), de Myers-Scotton (1997, 2001) e Myers-Scotton & Jake (2000a; 2000b), a que nos referiremos mais detalhadamente no ponto 2.4.2.3. do presente trabalho. Paralelamente, o comportamento da CPL-var foi também observado em função dos pressupostos de Lightfoot (1999a), que entende que o organismo linguístico, apesar de ser biologicamente controlado, é desenvolvido a partir dos estímulos ambientais que intervêm no processo de aquisição. Desta forma, as conclusões indicam que a regra da concordância se dá a partir das

formas fonológicas mais salientes, enquanto a inibição se conota aos itens menos salientes. Os resultados evidenciam ainda que os elementos pré-nucleares são mais marcados do que os itens pós-nucleares, enquanto as categorias nucleares são mais pluralizadas quando ocorrem na primeira e na quinta posições. Aliás, a marcação na primeira posição é quase categórica, o que confirma a sequência de aquisição dos morfemas, com os morfemas sistêmicos estruturais (*early system morphems*) sendo primeiramente adquiridos em paralelo com os morfemas de conteúdo (*content morphems*) e, só posteriormente, se dando a aquisição dos morfemas de nível funcional sistêmico (*late system morphems*).<sup>17</sup>

No seu trabalho, Lopes (2001), contrariamente ao avançado por Guy (1981a), dá ainda conta que os numerais, devido ao seu elevado grau de saliência semântica, favorecem a inserção de marcas. Por outro lado, o favorecimento da marcação é também motivado pela pausa final, enquanto a inibição é favorecida pelas consoantes sonoras. Já relativamente às variáveis sociais, Lopes (2001) constata que a escolaridade exerce forte influência na aquisição das regras de concordância, a ponto de fazer desaparecer divergências na marcação registada entre sexos, uma vez que nas classes não-escolarizadas são as mulheres quem mais aplica a pluralização.

Estabelecendo ainda comparações com os estudos levados a cabo para o MRJ (Scherre, 1988), para a região sul (Fernandes, 1996) e para a fala universitária de João Pessoa, Paraíba (Carvalho, 1997a), Lopes (2001) conclui que o uso da concordância nominal de número nestes dialectos apresenta o mesmo padrão da cidade de Salvador. Ainda assim, os elementos nucleares do *corpus* de Salvador em segunda, terceira, ou quarta posições exibem uma marcação menos acentuada, justificando Lopes (2001:377) este aspecto com o facto de os dados linguísticos primários (DLP's) para aquisição do português dos falantes desta região serem mais divergentes do que os do MRJ, região sul e João Pessoa.

Dando conta dos processos de pidginização e crioulização na formação da língua portuguesa no Brasil, Baxter & Lucchesi (1997) chamam a atenção para a pobreza a nível da concordância na fala de comunidades negras rurais que se encontram praticamente isoladas no Estado da Bahia. Como os estudos sobre comunidades rurais (Amaral, 1920; Veado, 1982; Wagner, 2001) e comunidades brancas de Mato Grosso,

interior da Bahia (Isensee, 1964; Callou, 1998), dão conta do mesmo tipo de fenómeno, Scherre (1994:38) entende que a “*variação na concordância de número no português falado do Brasil, longe de ser restrito a uma região ou classe social específica, é característico de toda comunidade de fala brasileira, apresentando diferenças mais de grau do que de princípio, ou seja, as diferenças são mais relativas à quantidade de marcas de plural e não aos contextos lingüísticos nos quais a variação ocorre*”. Deste modo, no Brasil não existirão características estruturais novas introduzidas pelo contacto entre línguas ou pela nativização do português entre os segmentos de falantes de outras línguas e seus descendentes, mas apenas uma ampliação de fenómenos lingüísticos regidos pelos mesmos factores estruturais que regiam o PE arcaico (Naro & Scherre, 2003:295). No entanto, esta visão afigurar-se-á simplista, uma vez que convive com a perspectiva formalista hegemónica e normativa de um sistema lingüístico que estaria na base de uma história lingüística única no Brasil (Lucchesi, 2008a:368), descartando a hipótese de existirem processos distintos e independentes de variação e mudança que concorreram, em paralelo, para a formação das duas grande normas do PB actual (Lucchesi, 2004a, 2007). Efectivamente, a mínima ou reduzida incorporação de morfologia flexional na estrutura morfológica e processos de recomposição da gramática do português popular brasileiro, ou seja, as suas estruturas mais marcadamente criouliantes, têm sido conotados à formação da norma popular brasileira, determinada, ao longo de toda a história do Brasil, pelos processos de transmissão lingüística irregular circunscritos, isto é, às variedades rurais mais refractárias à influência normatizadora dos círculos institucionais urbanos (Lucchesi, 2003:278).

O facto de os traços lingüísticos referidos se verificarem mais a nível do registo oral popular valida que se estude a fala das comunidades rurais isoladas formadas por descendentes de africanos que se fixaram no Brasil. Deste modo, P. Andrade (2003), decidiu analisar, atomística e sintagmaticamente, a concordância nominal de número no dialecto afro-brasileiro da comunidade de Helvécia, extremo sul do Estado da Bahia, a fim de verificar quais os elementos que resguardam as marcas das origens africanas do dialecto. Refira-se que, em comunidades rurais isoladas, como a de Helvécia (HEL-Ba), que fez parte da Colónia Leopoldina que perdurou de 1818 até à abolição da escravatura, é possível procurar determinar se a CPL-var entre os elementos do SN tem a ver com o

modo como se deu a aquisição entre os ancestrais falantes da comunidade. Assim, e partindo do pressuposto que a redução da morfologia flexional registada no PB resulta dos processos de transmissão linguística irregular decorrentes do amplo contacto entre línguas que ocorreu em solo brasileiro, este tipo de aquisição terá também sido regido pela ordem como os morfemas são adquiridos, isto é, pode ser conotado ao Modelo dos Quatro Morfemas (Myers-Scotton, 1997, 2001; Myers-Scotton & Jake, 2000a; 2000b), já avançado por Lopes (2001) no seu estudo sobre as normas culta e popular de Salvador.

A análise da CPL-var do SN de HEL-Ba permite concluir acerca de um quadro inicial de quase inexistência da regra de concordância entre os elementos que constituem a cadeia sintagmática, mas que o dialecto se encontra em processo de mudança em progresso no sentido de adquirir a regra. Este aspecto é confirmado com recurso à observação das variáveis sociais *faixa etária, sexo, escolaridade e estada fora da comunidade*. Quanto aos resultados da análise quantitativa, P. Andrade (2003:140) conclui que, relativamente à posição dos constituintes na cadeia sintagmática, o efeito mais significativo na aplicação da marcação é dado pela tendência geral verificada no PB de se inserir a marcação na primeira posição (Nina, 1980; Guy, 1981a; Scherre, 1988; Lopes, 2001). Este aspecto, segundo a autora, vem também confirmar a aplicação da hipótese funcionalista, já que há uma tendência para se expressar a pluralização na primeira oportunidade (Kiparsky, 1972:195). Tal fenómeno será ocasionado pela aquisição de um modelo imperfeito do português, determinado pelo processo de transmissão linguística irregular. E não havendo registo deste tipo de marcação no PE, a explicação só poderá ser conotada ao fenómeno do contacto entre línguas que aconteceu no Brasil (cf. Guy, 1981a).

P. Andrade (2003:141) confirma também a propensão de os elementos mais salientes serem mais marcados, uma vez que, havendo na composição do SN um elemento morfofonologicamente mais perceptível, ocorre igualmente tendência para se aplicar a regra da pluralização. A ocorrência deste tipo de fenómenos dar-se-á porque os contextos mais salientes concedem ao ouvinte uma margem para perceber os desvios gramaticais do falante, levando-o a evitar o uso de formas não-padrão.

Relativamente às *marcas precedentes ao elemento analisado*, evidenciam que o falante tem noção da pluralização semântica. Desta forma, os numerais apresentam-se como um factor determinante para a inserção de marcas, em virtude de serem morfemas de conteúdo semanticamente transparentes (P. Andrade, 2003:112), possibilitando a aquisição da regra em situações de contacto entre línguas. Outro factor determinante para a aquisição da regra é o número de elementos que compõem o SN, já que, quanto mais reduzidas forem as estruturas deste, mais facilmente o falante processa as regras de marcação. Já no que diz respeito às variáveis sociais, para além de permitirem ajuizar que a comunidade de Helvécia se encontra em processo de mudança (os jovens pluralizam mais os elementos do SN do que os idosos), revelam também que os homens aplicam mais a marcação PL, em virtude de terem um papel mais interventivo na vida social de HEL-Ba (P. Andrade, 2003:144).

Face aos resultados obtidos, P. Andrade (2003:95-96) confirma, na análise distribucional do PL, a correlação entre posição linear e classe gramatical proposta por Guy (1981b:179), dado o dialecto de HEL-Ba apresentar SN's de estrutura mais reduzida e regular do que os dos dados urbanos analisados por Scherre (1988), ou seja, configurando apenas dois elementos: DET+N. Paralelamente, P. Andrade (2003:146) corrobora em parte o pressuposto de Guy (1981a), no que concerne ao facto de a variação resultar do amplo e intenso contacto entre línguas que aconteceu durante a formação sociohistórica do país. Assim sendo, a variação teria como consequência uma criouliização leve, decorrente da mudança de língua e reestruturação determinada pelo forte *input* de PtgL2 irregular (Baxter & Lucchesi, 1997).

Tieppo (2003) realiza também estudos quantitativos, numa perspectiva sintagmática, sobre a variação na concordância de número da fala de uma comunidade afro-brasileira do município de Rio de Contas, localizada na Chapada da Diamantina, com vista a poder determinar o peso que o contacto entre línguas teve na formação histórica da presente realidade linguística do Brasil. A variável linguística *configuração sintagmática do SN* evidenciou que as marcas são preferencialmente inseridas quando o SN é formado apenas por DET+NUM (p.e. *os quatro*). A autora alega que estes casos não configuram, em rigor, verdadeiros casos de concordância, já que apenas é necessário incluir marcação no primeiro elemento. Ora, como a tendência geral é a de inserir marcação

apenas neste elemento, encontra-se justificado o facto de a marcação ser praticamente categórica neste tipo de configuração sintagmática. Aliás, este aspecto é depois confirmado pela estrutura DET+N, que desfavorece altamente a concordância (Tieppo, 2003:5). Outro aspecto confirmado por Tieppo (2003:4-5), e já verificado em estudos anteriores, é a tendência de os elementos modificadores (possessivos, adjectivos ou quantificadores) surgirem bastante mais marcados quando se encontram à esquerda do núcleo, por oposição às situações em que estão à direita deste.

Recorrendo à variável *número de constituintes flexionáveis do SN*, Tieppo (2003:5) corrobora P. Andrade (2003) quanto à tendência de a regra de concordância ser mais aplicada em constituintes com apenas um item flexionável, isto é, quando o processamento linguístico se revelar mais facilitado, por ocorrer em estruturas simplificadas. Refira-se que idêntico comportamento não é registado no PA, uma vez que os dialectos de HEL-Ba e Rio de Contas se encontram em situação de mudança no sentido da aquisição da regra de concordância, enquanto a fala de Almojarife, ao que tudo indica, apresenta estabilização da variação. Estes aspectos encontram-se discutidos noutras partes do presente trabalho, nomeadamente no ponto 4.1.2.6.2 e no capítulo 5, no qual se observa o comportamento das variáveis extralinguísticas. Relativamente às variáveis sociais *faixa etária* e *escolaridade*, Tieppo (2003:7) verifica que tanto as faixas etárias mais jovens como as mais escolarizadas realizam mais a concordância, confirmando uma situação típica de mudança aquisicional no sentido do aumento do uso da regra de concordância nominal de número na comunidade.

Tendo em conta que comunidades afro-brasileiras, como a de HEL-Ba, se constituíram a partir de uma diversidade étnica que originou contacto entre línguas, especialmente das europeias com as do grupo níger-congo atlântico, proporcionou-se uma situação de transmissão linguística irregular, determinante para a aquisição do português com uma simplificação da sua estrutura gramatical, isto é, retirando-lhe os elementos menos relevantes para a comunicação emergente. Considerando ainda que este tipo de contacto se espalhou por todo o país nos primeiros quatro séculos da sua existência, com o português culto sendo influenciado pelo português popular com a chegada de imigrantes estrangeiros a partir do século XIX, é possível inferir que o fenómeno da variação na concordância de número no PB não se restringe a uma

determinada região ou classe social específica, sendo característico de toda a comunidade de fala brasileira. Por outro lado, é possível verificar também que vários factores de ordem linguística e extralinguística interferiram para determinarem as actuais diferenças marcantes entre o PB e o PE. É o que acontece a nível da CPL-var entre os elementos do SN, tendo tais diferenças que ser entendidas como não aleatórias (Tieppo, 2003:7), isto é, como fenómenos sistemáticos de variação previsível, resultantes do quadro alargado de contacto entre línguas (P. Andrade, 2003:146) e que originou uma situação de transmissão linguística irregular de L2 para L1 no Brasil (Baxter & Lucchesi, 1997).

Como se referiu no início do presente capítulo, o comportamento do fenómeno da CPL-var no português popular brasileiro poderá ser conotado ao das línguas nativas do grupo níger-congo atlântico, enquanto substratos dos CP's atlânticos e substratos directos ou ancestrais de variedades africanas de português. Deste modo, passaremos, de seguida, a dar conta de trabalhos pioneiros sobre CPL-var que têm sido recentemente levados a cabo sobre a CPL-var no SN de algumas variedades africanas de português, incluindo o PA. Abordaremos igualmente o modo como se processa a regra da concordância nominal de número em alguns CP's atlânticos (CCV e santomense), bem como em substratos destes, pertencentes ao grupo níger-congo atlântico. Posteriormente, faremos o enquadramento sociohistórico e linguístico da comunidade de Almojarife, visto a sua fala constituir o principal objecto de estudo do nosso trabalho. Portanto, apontaremos também algumas características do PA, estabelecendo paralelismos morfossintácticos compartilhados, a nível da estrutura do SN, entre este dialecto e línguas africanas do grupo níger-congo atlântico, CP's atlânticos, variedades africanas de português e variedades brasileiras de português.

## **1.2. Variedades africanas de português e concordância nominal variável de número**

### **1.2.1. Variedades africanas de português com substrato do grupo níger-congo atlântico**

#### **1.2.1.1. Português dos tongas (Roça Monte Café, São Tomé)**

Os estudos quantitativos acerca da CPL-var no SN de variedades africanas iniciaram-se com a pesquisa atomística de Baxter (2004) sobre o português da comunidade semi-



isolada dos tongas, fixada na Roça Monte Café, São Tomé. Os falantes da comunidade são descendentes de africanos falantes de umbundu e kwa, “contratados” no final do século XIX e durante o século XX. Estes trabalhadores foram transportados para a ilha de São Tomé, a fim de prestarem serviço nas grandes roças de café e cacau, com contratos renováveis por períodos de três a cinco anos. Porém, antes do estudo de Baxter (2004), já Rougé (1992) havia chamado a atenção para a variedade de português falada pelos tongas, que apresenta traços dos seus substratos africanos (muitas vezes do tipo koiné), mas que não teria sido afectada pelo esparso contacto com as línguas de outros residentes permanentes ou temporários da ilha, nomeadamente o santomense,<sup>18</sup> o CCV ou as variedades linguísticas de Moçambique (zona de Quelimane).

Segundo Baxter (2004:104), o complexo mosaico sociolinguístico que determinou a emergência do PT pode ser sumariado da seguinte forma:

- (i) o modelo predominante de português utilizado pela massa trabalhadora seria o PtgL2, o qual terá existido num *continuum* de variedades de interlíngua, face à constante chegada de contratados que eram integrados na comunidade;
- (ii) as primeira e segunda gerações de crianças tiveram como L1 um koiné de base umbundu, que aprendiam antes do PtgL2;
- (iii) as crianças aprenderam português nos campos de trabalho e outros locais da roça, principalmente através do contacto com o PtgL2 dos trabalhadores e, esporadicamente, com os trabalhadores falantes de PtgL1 ou capatazes europeus.

Face a este contexto de transmissão linguística irregular, o PTL1 da geração mais idosa apresenta substancial e consistente reestruturação a nível do SN (representações de género e número), do SV (marcação de tempo e aspecto) e na realização das relações de caso ou construção das negações. Estes traços morfossintácticos são partilhados com os das gramáticas do CCV ou do santomense (Baxter, 2002), enquanto que, no plano do nivelamento e desenvolvimento das regras variáveis de concordância, o PT estabelece semelhanças com as variedades vernáculas do PB, de Angola, de Moçambique, de São Tomé (variedade popular e variedade de Almojarife), de Cabo Verde (Mindelo) e mesmo de Macau.

Para observar o desenvolvimento diacrónico da regra de concordância PL no SN do PT, Baxter (2004) seguiu as metodologias propostas por Scherre (1988) e Lopes (2001), constituindo cinco variáveis linguísticas e quatro extralinguísticas.<sup>19</sup> Tais variáveis permitiram que Baxter (2004) descartasse a possibilidade de derivações direccionadas de fenómenos internos ao PE estarem na origem da variação registada na regra de concordância PL do SN (cf. Naro, 1981; Naro & Scherre, 1993, 2000). Contrariamente, Baxter (2004:121) conota a variação à situação de contacto que envolveu componentes de PtgL2 influenciadas pelo umbundu e, simultaneamente, pela aquisição deste. Deste modo, a marcação revela tendência para ser inserida no elemento imediatamente à esquerda do SN, deixando, depois, de haver inserção de marcas nos itens seguintes. Como os falantes do PT eram confrontados com modelos difusos, terão optado por uma estrutura simplificada, isto é, isenta de marcação. Efectuou-se, desta forma, uma reconfiguração a nível de marcação no SN, a qual estabelece paralelismo com a do grupo níger-congo atlântico (balanta e wolof), substrato do CVV. Daí o autor entender que, no contexto sociohistórico específico de Monte Café, a L1 africana tenha sido, directa ou indirectamente, determinante no processo das transferências registadas na aquisição do PtgL1, via DLP's defectivos do PtgL2. Isto mesmo seria confirmado com uma observação sobre a marcação PL no desempenho dos falantes mais novos da comunidade, conforme a procedência dos pais: se nascidos em África ou na roça Monte Café. De facto, os falantes com pais nascidos na roça favorecem a inserção de pluralização, enquanto os falantes com pais nascidos em África a inibem (Baxter, 2009:283-284). O efeito, segundo Baxter (2009:284), radica no PtgL2 falado pelos pais dos informantes, o que, em última instância, vem corroborar Guy (1981a:300-301), quanto ao facto de o substrato africano ser uma potencial fonte da variação da concordância de número no SN, localizando a marcação na posição à esquerda do núcleo, tal como acontece no PA e PVB.

Os aspectos linguísticos e sociolinguísticos compartilhados pelas falas das comunidades da Roça Monte Café e de HEL-Ba levaram Baxter (2009) a produzir um estudo quantitativo, numa perspectiva comparada, sobre a CPL-var de ambos os dialectos. A metodologia adoptada seguiu uma perspectiva que combina teorias estruturais e aquisicionais, permitindo as primeiras que se observassem quer a

configuração estrutural da concordância quer o lugar da inserção do morfema PL, e as segundas que se percebessem as motivações que estarão na base da variação de número. As análises estruturais tiveram como base de partida as visões de autores que postulam a ocorrência de uma categoria funcional SNUM entre o DET e o SN (Picallo, 1991; Carstens, 1991; Ritter, 1991; Valois, 1991; Bernstein, 1993; Cinque, 1994; Montrul, 2004), entendimento este que determina que as projecções funcionais são motivadas pela semântica. Simultaneamente, foi também adoptado o pressuposto de que a noção de concordância de número é reflexo de uma relação sintáctica local (Chomsky, 2001b), isto é, que a concordância é computada entre os elementos do SDET, por acção de uma relação de C-comando a partir do DET, com vista a acontecer o mapeamento e consequente pareamento (*matching*) de traços. Além destas noções, foram também levados em consideração os pressupostos já avançados por Lopes (2001) e P. Andrade (2003) acerca da teoria cognitiva dos Quatro Morfemas (Myers-Scotton, 1997, 2001; Myers-Scotton & Jake, 2000a; 2000b), bem como a visão compartilhada por Costa & Silva (2006b) acerca dos princípios sobre morfologia distributiva e não distributiva. Estes, tendo como base os trabalhos de Galves (1993), Scherre (1994) e Embick & Noyer (2001), entendem que o PE e o PB manifestam distintas realizações do morfema associado à pluralidade. Como tal, os morfemas podem associar-se aos núcleos, independentemente dos processos de movimento, realizando-se de duas maneiras: (i) como morfemas dissociados, que não figuram na sintaxe, mas são inseridos após o *spell-out*, como um reflexo de configurações sintácticas; (ii) como morfemas *singleton*, que se associam aos elementos funcionais, responsáveis pela informação de número e referência.

Relativamente aos aspectos aquisicionais, a presença de PtgL2 nas duas línguas, determinada por situações de contacto que envolvem estágios intermédios de ASL em simultâneo com aquisição de L1 a partir do *input* de modelos de L2, isto é, fases de bilinguismo, fez com que se levassem em consideração as teorias sociolinguísticas que avaliam o contributo dos processos de ASL para análise da variação. Também foi tido como relevante o facto de a aquisição envolver mudança de alvo, dado que, em fases iniciais, a L2 das massas escravizadas teria várias fontes: (i) a L1 dos administradores; (ii) a L2 de outros escravos, baseada na L1 dos administradores; (iii) a L1 da população

escrava, baseada na L2 de outros escravos (cf. Smith, 2006). Paralelamente, foi ainda tida em conta a situação de desequilíbrio sociodemográfico entre contingentes escravizados (a grande maioria) e não escravizados (uma acentuada minoria), que levou o contacto com a LA a tornar-se esparso. Desta forma, ao configurar-se um cenário propício à mudança de língua, substancial material da L2 pôde ser não só incorporado na nova L1 da comunidade nas fases iniciais da aquisição mas também transmitido geracionalmente (Winford, 2003a:245-247), fossilizando-se algum desse material. Simultaneamente, a presença de bilinguismo na comunidade determinou a ocorrência de transferências de elementos das L1's originais para a nova L1 (cf. Thomason & Kaufman, 1988; Lefebvre, 2003; Montrul, 2006; Lefebvre, White & Jourdan, 2006b; Siegel, 2008).

Atendendo não só ao facto de diversos estudos sobre CPL-var do PB apresentarem a posição linear do item pluralizável como a chave para a inserção da flexão de PL mas tendo também em conta que Guy (1981b:179) advoga existir uma relação intrínseca entre posição linear do item e sua classe gramatical, fixando o eixo da marcação PL nos determinantes e originando uma configuração de marcação que terá respaldo no sistema linguístico bantu e kwa (Guy, 1981a, 1989, 2005), Baxter (2009) observa a forma como estas duas variáveis, em concomitância com a variável social idade, podem intervir na forma como se perfila e desenvolve aquisicionalmente a variação PL nas duas comunidades, uma vez que ambas apresentam configurações sintagmáticas de complexidade reduzida. Para tanto, Baxter (2009) tem ainda como primordial, para as suas análises, determinados factores que podem ter influência na forma da L2 dos adultos e da L1 irregular das crianças. Para a primeira, devem então ser considerados os seguintes aspectos: (i) simplificação, isto é, presença na L2 de formas reduzidas e regularizadas (Klein & Perdue, 1997; Siegel, 2004a; Mather, 2006), orientadas pelo tipo de *input* (p.e. frequência, saliência e transparência semântica ou pragmática) e pela GU; (ii) transferência e relexificação, ou seja, presença de estruturas da L1 do aprendente na sua L2 incipiente (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006); (iii) ASL imperfeita, que ocorre em virtude de o falante adulto não adquirir os traços das novas categorias funcionais (*Failed Functional Features Hypothesis* – FFFH) (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995;

Hawkins & Chan, 1997; Franceschina, 2002, 2003), visto já ter sido ultrapassado o período crítico para a ASL. Quanto à L1 irregular das crianças, não podem ser descartados os recentes achados sobre aquisição da morfologia de PL no PB, nomeadamente do modo como as classes morfológicas fechadas (*closed class morphemes*) intervêm no processamento dos DLP's disponíveis para a aquisição de L1 e orientação do desenvolvimento da pluralização (Ferrari-Neto, 2003; Corrêa, Augusto & Ferrari-Neto, 2005; Ferrari-Neto, Corrêa & Augusto, 2005).

Os resultados de Baxter (2009) revelam que, em ambos os dialectos (PT e HEL-Ba), ocorre uma forte correspondência entre posição linear e classe gramatical, sendo a posição imediatamente pré-nuclear o ponto de inserção da pluralização. Nesta conformidade, Baxter (2009) entende que o perfil inicial da variação da comunidade bilingue de Monte Café deve ser essencialmente atribuído ao substrato africano. Quanto às variáveis extralinguísticas, ajudam a perceber que a variação aponta para mudança, no sentido da aquisição da morfologia padrão e aplicação das regras de concordância. Por seu lado, Figueiredo (2009a) aponta para o PA um perfil idêntico ao de HEL-Ba, com os itens imediatamente pré-nucleares a revelarem tendência para serem os primeiros a receberem a marcação mórfica de PL. Como tal, “*estes elementos têm um papel mais activo na construção estrutural da concordância, relegando os itens pós-nucleares para funções sintácticas de mera adjunção*” (Figueiredo, 2009a:47). Baxter (2009), com recurso ao mecanismo formal misto que propõe, indica que a distribuição da concordância PL no SN segue um padrão uniforme em HEL-Ba e no PT, e que, a avaliar pelos pressupostos de Figueiredo (2009a), será também idêntico ao do PA. Deste modo, para se perceber como se processa a ordem de entrada da marcação PL no SDET de HEL-Ba e no PT, Baxter (2009:293) avança com o seguinte esquema:

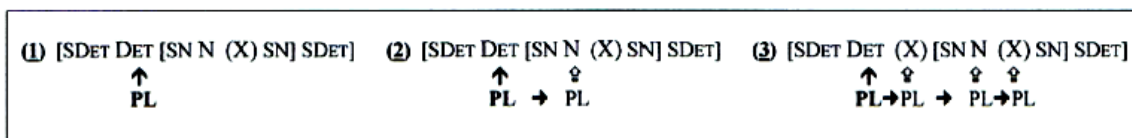


Fig. 1. Dialecto de HEL-Ba e PT: desenvolvimento da atribuição de PL.

A configuração mostra que o desenvolvimento da concordância PL se inicia com a introdução de um PL *singleton* na categoria funcional DET, que serve de âncora para o

controle da pluralização. Quando essa posição não é ocupada por um elemento DET, o SN torna-se alvo de PL, em virtude das suas características lexicais, fazendo “*lembrar o papel fulcral do PL no classificador nominal na tipologia níger-congo, e portanto no substrato níger-congo atlântico (balanta e wolof) do crioulo de Cabo Verde*” (Baxter, 2009:292). Portanto, a concordância torna-se parcial e é atribuída a partir da posição DET, isto é, a partir do elemento pré-nuclear adjacente. Posteriormente, o PL desenvolve-se, ainda com características de concordância parcial, através de morfemas dissociados, que são atribuídos aos elementos mais próximos do *fulcro*. A ordem de atribuição contempla os elementos do SDET, primeiro, e os itens à direita do núcleo, depois. Não obstante, o perfil da âncora pré-nuclear é mantido geracionalmente. De certa forma, é aqui estabelecido um paralelismo com os pressupostos advogados por Lucchesi (2000a) no seu estudo sobre a concordância variável de género (CGEN-var) da fala de HEL-Ba, e que, a propósito da concordância plural no SN, sustentam actuar uma correlação entre Princípio da Simplicidade, que leva a concordância a fixar-se inicialmente nas estruturas mais simples dos SN’s, constituídas por nome e determinante, e Princípio da Integração, que faz com que a marcação ocorra mais nos itens flexionáveis à esquerda do núcleo do SN (com incidência na posição adjacente) e menos nos elementos à sua direita.

Desta forma, se houve realmente um modelo inicial baseado no PE, será lícito concordar com o pressuposto teórico de que os africanos e os seus descendentes exageraram esse modelo (Naro, 1981; Naro & Scherre, 1993, 2000, 2007a). Contudo, a semelhança entre o papel da posição pré-nuclear para o sistema de marcação PL de HEL-Ba e Monte Café com CP’s, como o CCV (Castro & Pratas, 2006), contribui para que se entenda esta estrutura como resultante das características sociohistóricas compartilhadas pelos três dialectos. Portanto, parece bastante lógico assumir que o africano teria influenciado o desenvolvimento do PL, havendo ou não variação no modelo original (Baxter, 2009). Desta forma, tanto a hipótese contactista como a hipótese derivista são capazes de explicar a realização da CPL-var, propondo Baxter (2009:293) que ambas passem a figurar num modelo sociolinguístico único para o PB. Ainda assim, em pesquisas futuras sobre esta língua, deverá ser tido em conta que as

pressões exercidas na variação quer pelo contacto quer pela deriva terão sido determinadas pelas diferentes circunstâncias em que emergiu a variedade em análise.

### **1.2.1.2. Português vernáculo de Angola**

Não temos conhecimento de existirem estudos numa perspectiva sociolinguística quantitativa sobre o português vernáculo de Angola (PVA). Contudo, vários autores têm feito referência à concordância variável neste, estabelecendo paralelos morfossintáticos e morfofonológicos quer com os substratos africanos quer com o PVB (Marques, 1985 [1983]; Gärtner, 1989, 2003; Asher, 1994; Mello, 1997; Baxter, 1998; Chatelain, 2001; Mingas, 2000; Zau, 2001; Veiga, 2002a; Figueiredo, 2003; Inverno, 2004, 2005, 2008, 2009; Cabral, 2005a, 2005b). Ainda que o PVA seja língua de classes sociais específicas em Angola, estabelecidas nos centros urbanos costeiros, numa situação de nítida diglossia, as partilhas estruturais entre a variedade de português e os dialectos do substrato, nomeadamente a nível da CPL-var no SN, resultam de processos paralelos de formação no primeiro, ou seja, de reestruturação parcial, em consequência de situações de transmissão linguística irregular. Após a sua fixação no litoral do reino do Congo, não como colonizadores, mas como parceiros no tráfico de escravos e marfim, os portugueses rapidamente assumiram o controlo de Loando, em 1576, estabelecendo a colónia de Angola (Parreira, 1989; Ogot, 1999:278). O domínio foi assegurado por uma elite de afro-portugueses, resultado de uma política de miscigenação entre europeus e mulheres da nobreza angolana (Chatelain, 2001:14). Esta elite, sendo bilingue em português e línguas bantu, assumiu postos na função pública e servia de interlocutora entre os portugueses e os chefes indígenas, a fim de se garantir a paz com estes e se manterem activas as rotas que possibilitavam o resgate e o comércio de escravos (Venâncio, 1996:51).

Assim, e ao contrário do que sucedeu no Brasil, onde a variedade de português reestruturado acabou por se expandir por todo o território, gerou-se em Angola uma situação de diglossia estável, englobando uma elite afro-portuguesa bilingue, detentora de um código linguístico gramaticalmente elaborado (o PE), por um lado, e uma grande maioria indígena, falante de variedades vernaculares de português, por outro lado. Como o desenraizamento desta população não se realizou em grande escala a nível interno,

também não aconteceu uma situação que impedisse a inteligibilidade linguística entre africanos. Simultaneamente, como a aculturação destes à língua e cultura portuguesas foi bastante incrementada, sobretudo a partir do século XVIII (Carreira, 1979; Venâncio, 1996; Ogot, 1999), não se verificou uma estabilização generalizada das variedades básicas de português faladas no território, que pudesse estar na origem do PVA. Contudo ocorreram interferências mútuas entre o português e línguas bantu, sobretudo nas zonas residenciais periféricas às cidades litorais (os musseques). Deste modo, no século XIX Schuchardt (1888:249) referia-se já ao português “*mestiço*” de Luanda, um composto de “*palavras portuguesas acomodadas ao genio do bundo*”, enquanto Valman (1978:22), posteriormente, fazia também menção a um “*pequeno português*”, falado nos principais centros urbanos angolanos. Estes aspectos levam a crer que os autores davam conta de interferências, configurando, possivelmente, um processo de “*empréstimo estrutural moderado*” (Inverno, 2004:4), situação que requer, por considerável período de tempo, não só um acentuado bilinguismo por parte dos falantes das línguas que concedem os empréstimos mas também a manutenção das L1’s dos falantes envolvidos na situação de empréstimo (Thomason & Kaufman, 1988:37).

Dois outros aspectos concorreram para acentuar a situação de diglossia em Angola: (i) a ocupação do território limitada à orla marítima até meados do século XX (Gann & Duignan, 1981:9; Chatelain, 2001:5), que restringiu o uso do PE aos centros urbanos costeiros; (ii) o atenuar do processo de miscigenação com a posterior chegada de grande quantidade de mulheres provenientes da metrópole e que acompanharam os seus maridos no processo de colonização do interior, a partir da segunda metade do século XX. Nesta fase, também as políticas de aculturação levadas a cabo pelo Estado Novo visaram a implementação do PE, já que as oportunidades de ascensão social, nomeadamente através da colocação em cargos da administração pública, eram apenas concedidas aos que demonstravam “*saber ler, escrever e falar fluentemente português*” (Bender, 1980:216).<sup>20</sup> Contudo, uma vez que a maioria da população não tinha acesso à educação, o português era aprendido pelo contacto prático e funcional, o que fez com que as línguas do substrato fossem afectadas não só por empréstimo lexical massivo mas também pelo sistema fonológico e morfológico do português (Oliveira, 1990:69-89). Tal levou a que, posteriormente, ocorresse uma mudança para o português, já que este se



assumiu como uma das L1's dos falantes bilingues das classes menos escolarizadas que se fixaram nas zonas costeiras. Emergiu, deste modo, uma variedade reestruturada de português marcada por influências dos substratos. Este processo de mudança linguística, apesar de apontar para uma situação de formação comparável à do Brasil, é mais recente e restrito social e geograficamente, não estando, portanto, concluído (Inverno, 2004:5).

No que concerne à CPL-var no SN do PVA, tem sido indicada em diversos estudos a ausência de concordância entre os elementos não-nucleares do SN e o núcleo deste, um fenómeno que alguns autores apontam como sendo um dos traços morfossintáticos mais evidentes da variedade reestruturada de Angola (Cabral, 2005:73). Baxter (1998), a exemplo do que apontaria posteriormente para o PVB (Baxter, 2009), entende que o fenómeno resulta da convergência entre posição linear e classe gramatical, muito por influência do sistema de marcação PL dos substratos bantu (umbundu, kimbundu e kikongo). Assim, na aprendizagem do PtgL2, os nomes terão sido interpretados como elementos invariáveis, enquanto os elementos não nucleares do SN foram entendidos como afixos variáveis de marcação semântica pré-nuclear, a exemplo do que sucede nos dialectos do substrato. Daí que a marcação de número ocorra neste tipo de elementos, colocados à esquerda do núcleo do SN (Inverno, 2005:135), e não no núcleo (Marques, 1985 [1983]:218; Gärtner, 1989:159; Vilela, 1999:56; Inverno, 2004:5, 2005:134). Este aspecto acaba mesmo por determinar fenómenos de apócope sobre o *-s* final que não detém função de marcador de número PL em algumas palavras:

[1] PVA:  
*Quando acabar as férias\_* (Inverno, 2005:133)

### 1.2.1.3. Português de Moçambique

O contexto sociohistórico que levou à formação do português de Moçambique (PM) é em tudo comparável ao de Angola. Assim, estamos perante uma variedade não-nativa de uma língua ex-colonial, adquirida natural ou formalmente como L2 pela maioria das crianças falantes de L1's bantu, com especial destaque para o emakhuwa, ainda que este seja usado apenas por 26,3% dos utilizadores das línguas nativas (Instituto Nacional de Estatística de Moçambique, 2010). Quanto à presença dos portugueses em Moçambique, data do século XV (1498). Contudo, por razões económicas e culturais, a difusão do

português foi bastante incipiente durante os primeiros quatro séculos de colonização da região. De facto, quando os portugueses desembarcaram nesta ex-colónia, já os árabes se haviam aí estabelecido, controlando os principais centros urbanos. Assim, só em finais do século XIX/princípios do século XX é que os europeus conseguiram o controlo do território, nomeadamente através das chamadas campanhas de “pacificação” (1886-1918) (Gonçalves, 2004:230). Por outro lado, como a atenção da metrópole estava essencialmente dirigida para as riquezas do Brasil, o domínio total da região terá sido adiado em virtude de a mesma ser considerada periférica para o contexto do império, como bem o demonstra o facto de a sua administração ter sido feita a partir da Índia até 1752.

A exemplo do que aconteceu em Angola, os portugueses fixaram-se na zona costeira, sendo aí que ocorreu maior contacto entre o PE e as línguas indígenas, enquanto no interior a população continuava a usar as suas L1’s. Este aspecto originou uma situação de diglossia, que se mantém até aos dias de hoje, já que nas zonas rurais o português é apenas utilizado em escolas, onde o seu ensino é compulsivo (Gonçalves, 2004:233). Aliás, a tardia implementação de uma política educacional consistente<sup>21</sup> terá também acentuado a situação periférica da ex-colónia. As consequências foram idênticas às de Angola, criando-se uma população bilingue minoritária e sendo as línguas do substrato, faladas pela maioria da população (98%) (Firmino, 1995:221), afectadas lexical, fonológica e morfologicamente com elementos do português. Este aspecto impediu a emergência de um pidgin e, posteriormente, com a independência do país (1975), a comunidade de falantes do português cresceu consideravelmente, devido a razões socioeconómicas e ao incremento da rede de educação (Gonçalves, 2004:232). Assim, o número de falantes de PtgL2 aumentou cerca de 15% (Firmino, 1995:224; Gonçalves, 2004:232), acentuando a variação linguística que caminha no sentido da mudança, uma vez que o número daqueles que usam o PtgL1 reestruturado marcado por influências dos substratos cresceu mais de 50% (Firmino, 1995:224; Gonçalves, 2004:232). Como se verifica, este processo de variação linguística ainda não estabilizado (Carvalho, 1991:62; Gonçalves, 2004:236), tal como em Angola, é recente e restrito social e geograficamente, mas apontando para uma situação de formação que pode ser comparada à do Brasil.

Em trabalho sobre a concordância variável no SV do PML2, numa perspectiva generativista, Gonçalves (2004) analisa o modo como factores linguísticos e extralinguísticos podem bloquear o restabelecimento de novos valores paramétricos (cf. Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), tornando o processo de mudança da L2 menos gradual que o da L1 e também menos dramático que o da crioulização. O estudo evidencia ainda como o conhecimento prévio da gramática da L1 bantu dificulta o processo de aquisição do PtgL2, necessitando este, para valores de determinados parâmetros da LA, a exposição a DLP's extra. Pelo facto de Moçambique constituir uma sociedade pós-colonial, estes DLP's extra poderão não se encontrar disponibilizados, principalmente devido à preponderância quantitativa de falantes do PtgL2. Em consequência de tal, a instanciação em cadeia de novos/errados parâmetros, que origina a variação, torna-se cada vez mais frequente no *input*, eliminando a possibilidade de um reajustamento paramétrico e fazendo com que os novos/errados parâmetros sejam tomados como modelos para uma nova reinstanciação paramétrica do PML1. Consoma-se, desta maneira, a mudança que acentua a distância tipológica entre a LA e a capacidade gramatical (ou língua-I) atingida pelos falantes do PM (Gonçalves, 2000:255). Em resumo: alguns traços das L1's bantu são retidos no PtgL2 e sobrevivem na gramática do PM, tal como algumas propriedades gramaticais da LA se perdem no *continuum* de aquisição (Gonçalves, 2004:229), originando uma “*moçambicalidade gramatical*” (Gonçalves, 2000:255).

Tendo não só em conta que a sociohistória do PM é bastante recente mas também considerando que a gramática do PE não foi plenamente abandonada por todos os falantes moçambicanos, os núcleos comunitários operam com mais do que uma gramática. Ocorre, desta forma, uma espécie de diglossia não só no plano intercomunitário mas até a nível “interno” dos próprios falantes (cf. Lightfoot, 1999b:2). Tais espécies de diglossia tornam difícil precisar qual das gramáticas gera determinadas produções linguísticas, seja no plano da realização de grupo seja a nível individual. Este aspecto faz com que a frequência de traços não-padrão no discurso dos falantes do PM não seja uniforme para todas as comunidades nem mesmo para todos os membros de uma determinada comunidade, originando diferentes variedades comunitárias e

subvariedades nestas, sendo tais traços disponibilizados através do *continuum* dialectal. Contudo, e de um modo mais geral e uniformizado, poder-se-á dizer que o *continuum* dialectal do PM apresenta uma extensão que vai desde a variedade não-padrão falada pelos indivíduos menos escolarizados até à variedade padrão das classes superiores e com mais elevado nível de escolaridade.

A existência deste *continuum* em Moçambique levanta questões metodológicas, quando se toma determinada subvariedade como objecto empírico para o estudo da gramática do PM. No caso concreto da subvariedade de português falada no Maputo (PMp), é na área da sintaxe (léxico-sintaxe, morfossintaxe e sintaxe propriamente dita) que ela mais se diferencia do PE (Carvalho, 1991; Gonçalves, 1997). No que concerne aos fenómenos sintácticos e morfossintácticos que apresentam desvios mais sistemáticos relativamente ao modelo da LA, podem apontar-se “*as regras de utilização dos artigos e pronomes, os mecanismos de encaixe de orações subordinadas, e as regras de concordância verbal e nominal*” (Gonçalves *et alii*, 1998:36). Ainda no caso concreto da variação a nível da concordância nominal, estes autores referem que o nome se apresenta no singular, em ocorrências que exigem a flexão no PL. Por outro lado, a pluralização apenas aparece inserida morfologicamente nos artigos e pronomes (exemplo [2]), ou é determinada pela presença de um numeral não singular (exemplo [3]) (Gonçalves *et alii*, 1998:107-108):

[2] PMp:  
*Eu não mostro **os meus dente**\_* (Gonçalves *et alii*, 1998:108)

[3] PMp:  
*O lobolo aumentou, eram **quatro conto**\_* (Gonçalves *et alii*, 1998:108)

Em adição, também se verifica a falta de concordância em número do núcleo nominal com adjectivos (exemplo [4]), pronomes (exemplo [5]) e artigos (exemplo [6]) (Gonçalves *et alii*, 1998:108):

[4] PMp:  
*... com **uns preços um bocado baixo**\_* (Gonçalves *et alii*, 1998:111)

[5] PMp:  
***Meus neto**\_ são dezasseis neto\_* (Gonçalves *et alii*, 1998:110)

- [6] PMp:  
*Os vizinho\_ hãode vir.* (Gonçalves *et alii*, 1998:110)

Tudo indica que o fenómeno da CPL-var no SN resulta da convergência entre posição linear e classe gramatical, evidenciando a forma específica como as propriedades gramaticais particulares das L1's dos falantes em aquisição do PtgL2 podem interferir no processamento do *input*, levando à posterior instanciação de novos/errados parâmetros (Gonçalves, 2004:235).

Moreno & Tuzine (1998) elaboram também um estudo sobre o PMp, mas numa perspectiva sociolinguística laboviana que visa analisar a relação existente entre a variação linguística e os factores sociais. Trabalhando unicamente com valores percentuais,<sup>22</sup> os autores observaram a interferência de quatro variáveis sociais (*escolaridade, idade, profissão e local de residência*) no léxico-sintaxe (pronomes pessoais reflexivos) e morfossintaxe (concordância verbal e concordância nominal) do PMp. Apesar da dimensão reduzida do *corpus* analisado, os resultados sugerem que os fenómenos da variação do PMp não podem ser explicados recorrendo-se unicamente a uma ou duas variáveis isoladamente, já que a variação é motivada por vários factores actuando em conjunto. Ainda assim, tudo parece indicar que o nível de escolaridade e o tipo de profissão são as variáveis sociais mais relevantes para a ocorrência da variação registada no discurso dos informantes. Não obstante, Moreno & Tuzine (1998:86) alertam para o facto de os fenómenos de variação terem sido observados a partir de dados “*recenseados numa pequena amostra e, por isso mesmo, só futuros estudos poderão confirmar ou não as presentes conclusões*”.

Assim, seria Jon-And (2008) quem se debruçaria novamente sobre o PMp numa perspectiva quantitativa, observando a CPL-var no SN produzido por 18 informantes de ambos os sexos, distribuídos por três faixas etárias (20-40 anos; 41-60 anos e +60 anos).<sup>23</sup> Curiosamente, a autora levou em conta a variável social *idade*, que os resultados de Moreno & Tuzine (1998:84) haviam considerado não-significativa para a motivação da variação, em virtude de o português dominar cada vez mais o registo oral da vida quotidiana no Maputo, principalmente entre as gerações menos idosas. Como Jon-And (2008) não constituiu as variáveis *escolaridade e profissão* para o seu estudo, ficam por

confirmar os achados de Moreno & Tuzine (1998) no que concerne à maior influência destas variáveis na variação registada no PMp.

Para além das variáveis extralinguísticas *idade* e *sexo*, Jon-And (2008) observou ainda o peso exercido pelas seguintes variáveis linguísticas na CPL-var no SN: *saliência fónica, posição em relação ao núcleo, contexto fonológico seguinte* (excluída por ter pouco efeito), *marcas precedentes* e *classe gramatical* (ambas excluídas por revelarem interferência com outras variáveis). As discussões detalhadas em torno deste estudo serão apresentadas nos capítulos 4 e 5, dedicados à análise e comparação dos nossos resultados com os de outras pesquisas. Não obstante, é possível adiantar que as conclusões da pesquisa apontam para que as variáveis fonológicas detenham efeitos semelhantes aos dos registados para o PT e para o Brasil, enquanto a variável linguística *posição em relação ao núcleo do SN* se apresenta como aquela que mais motiva a variação no PMp, “*possivelmente por causa de influência de línguas bantu, L1s dos informantes maputenses*” (Jon-And, 2008:7).

## 1.2.2. Variedades africanas de português com substrato crioulo

### 1.2.2.1. Português cabo-verdiano e seu substrato (crioulo de Cabo Verde)

O CVV e o santomense têm em comum o facto de possuírem, como substrato, línguas do grupo níger-congo atlântico, estabelecendo paralelismos formais no que concerne ao sistema de marcação PL e que podem ser conotados às L1's dos ancestrais aprendentes da interlíngua. De facto, ambos os crioulos podem inserir a marcação PL apenas no elemento pré-nominal, permanecendo inalterado o radical dos restantes itens do SN:

[7] CCV:  
*Góra, nu ta kume uns banana.* (Pereira, Arim & Carvalho, 2006:12)  
Agora 1PL ASP comer IND banana  
“Agora, nós vamos comer umas bananas.”

[8] FORRO:  
*Ê sêbe tudu inen kwa se.* (Hagemeijer, 2007:122)  
3SG saber tudo 3PL coisa MSPEC  
“Ele sabe todas estas coisas.”

Este sistema de marcação estrutural transitou quer para o PCV (exemplo [9]) quer para o PA (exemplo [10]), permitindo considerar a probabilidade de as distintas variedades de português adquiridas em situação de contacto por falantes de diferentes CP's atlânticos manifestarem idênticos padrões de variação. Por outro lado, o paralelismo deste tipo de marcação com o do PB argumenta a favor do pressuposto da influência das ancestrais línguas africanas no sistema de pluralização deste (cf. Guy, 1981a:301-302):

[9] PCV:  
*As coisa foram subindo.* (Jon-And, 2009:1)  
“As coisas foram subindo”

[10] PA:  
*Vem do pé todas dia paqui* [FRANCM2]  
“Vem a pé todos os dias para aqui”

Formado algumas décadas depois do início da ocupação do arquipélago de Cabo Verde por parte dos portugueses e escravos resgatados na costa africana (actuais regiões do Senegal e Guiné-Bissau) e transplantados pelos primeiros, o CCV tem na sua génese condições sociolinguísticas especiais, já que a comunidade relativamente estável do arquipélago era constituída por falantes de diferentes L1's do grupo níger-congo atlântico. Muitos destes falantes já falariam um pidgin rudimentar de base lexical portuguesa, que deixaria de ser visto como L2 de comunicação emergencial para assumir o estatuto de língua de comunicação interétnica na comunidade. Sendo esta transplantada, apesar de ser geográfica e culturalmente heterogénea, a sua língua interétnica evoluiu para um crioulo, uma vez que os falantes deixaram de usar os seus idiomas de origem, salvo em situações de ambiente familiar (Pereira, Arim & Carvalho, 2006:3). Contudo, das dez ilhas que constituem o arquipélago, apenas nove foram povoadas, mas em épocas diferentes.<sup>24</sup> O tempo que mediou entre a primeira e a última fase do povoamento do arquipélago e o modo como o mesmo aconteceu nas diferentes ilhas foram determinantes para que ocorresse uma variação dialectal acentuada, nomeadamente entre as ilhas de Santiago e de S. Vicente (Pereira, Arim & Carvalho, 2006:3).

Desde a sua formação até aos nossos dias, o CCV tem mantido um contacto quase exclusivo com a língua portuguesa. Privados das relações com as suas terras de origem, os africanos e seus descendentes acabaram por deixar de usar as suas L1's. Desta forma, em Cabo Verde existem apenas duas línguas nacionais: o CCV e o português, que ganhou o estatuto de língua oficial, embora se preveja para breve o estatuto de “língua de ensino e da administração” para o crioulo. Estigmatizado durante séculos, o CCV, apesar de ser a L1 de todos os cabo-verdianos, não era estudado nas escolas nem usado em contextos oficiais. A política de sucessivos governos portugueses, que visava erradicar este dialecto e substituí-lo pelo português, chegou mesmo ao ponto de proibir o seu uso em locais públicos. Em consequência de tal, o português e o CCV passaram a coabitar em estado de afastamento linguístico, com o segundo sendo remetido para o espaço familiar, enquanto o primeiro se tornava a língua elitista do ensino e da vida pública, único idioma de acesso à literatura e aos bens sociais. Estas barreiras artificiais fizeram com que alguns falantes do CCV tentassem torná-lo o mais parecido possível com o português que sabiam. Como muitos deles não eram escolarizados, emergiram duas variedades regionais no CCV: a das ilhas do Barlavento, isto é, do Norte (Santo Antão, São Vicente, São Nicolau, Boavista, Sal, Santa Luzia), mais próxima do português actual; e a das Ilhas do Sotavento, ou seja, do Sul (Santiago, Maio, Fogo, Brava), classificada como acrolecto. O conjunto destes aspectos conduz, pois, a um *continuum* de variedades, umas mais fundas (menos assimiladas) e outras mais leves (mais assimiladas). Assim, mesmo que olhemos apenas para a variedade de Barlavento, verificam-se algumas diferenças nos registos dos falantes do interior, dotados de fraca ou nenhuma escolarização, que falam um crioulo fundo, e dos falantes das zonas urbanas, mais escolarizados, que utilizam um crioulo leve. Por outro lado, o *continuum* resultante da presença generalizada do português, entre variedades basilectais e acrolectais do CCV, tem levado a que todas elas se encontrem em processo de descrioulização (Pereira, Arim & Carvalho, 2006:3).

Müller (1994a:87), analisando a aquisição do alemão e do francês como L1's em crianças bilingues, defende que a mesma sofre um processo de maturação, razão pela qual as gramáticas não apresentam marcação de género ou número no estado inicial da aquisição. Portanto, estas noções gramaticais só são disponibilizadas pela GU num



estádio posterior de desenvolvimento gramatical, pelo que Müller (1994a:62) parte do pressuposto que as gramáticas adultas possuem pelo menos uma projecção funcional acima do SN (Fig. 34), no qual se encontram alojados os traços flexionais dos nominais que não estão disponibilizados na fase inicial da aquisição. Esta será a situação da fase inicial de aquisição dos pidgins que estiveram na base do CCV e do santomense, já que os indivíduos em idade adulta, quando submetidos à aquisição de uma L2 emergente em situação de contacto massivo, não assimilam o sistema de flexões da LA, passando a sua interlíngua a registar ausência ou redução destas. Posteriormente, os falantes responsáveis pela estabilização estrutural, face à continuada ausência de realização do DET nos DLP's do *input* que recebem, e conforme sugerido pela FFFH, poderão não reinstanciar os traços perdidos pelas gerações antepassadas, pelo que continuam a apagar, em muitas das suas realizações, as marcas formais de plural, como sucede no PCV e no PA.

Em português, alguns nomes genéricos (p.e. *O homem é um ser inteligente. / A coisa agora está preta.*) também podem ser usados no singular, com referência PL. Contudo, e ao contrário do que sucede nos crioulos referidos, nos quais os exemplos [7] e [8] não permitem outras alternativas de pluralização, os genéricos portugueses admitem várias realizações: (i) recorrendo ao singular, antecedido de artigo definido (exemplo [11a]); (ii) utilizando o PL, antecedido de artigo definido (exemplo [11b]); (iii) fazendo recurso ao singular, antecedido de artigo indefinido (exemplo [11c]); (iv) aplicando o singular e omitindo o artigo (exemplo [11d]); (v) servindo-se do PL e elidindo o artigo (exemplo [11e]):

- [11] PTG:
- a. *A cobra é peçonhenta.*
  - b. *As cobras são peçonhentas.*
  - c. *Uma cobra é um bicho peçonhento.*
  - d. *Cobra é bicho peçonhento.*
  - e. *Cobras são bichos peçonhentos.*

Para os casos em que a marca formal de PL não é realizada, Scherre (2001:98-99) considera estar-se perante representações abstractas de pluralização, em que a carga semântica de pluralidade é percebida pelos falantes a partir do contexto, uma situação que acontece também nos crioulos acroletais de Cabo Verde. Daí que, no

português, se flexione apenas o adjunto adnominal com determinados etnónimos (p.e. *os bantu*), estabelecendo-se aqui um sistema de pluralização estrutural absolutamente idêntico ao dos exemplos [7], [9] e [10]. Contudo, este tipo de marcação não representa a regra geral do português, ou seja, não constitui sistematização, como acontece no CCV, PCV e PA, nos quais as motivações que a determinam são distintas das do português, abrangendo outras categorias de nomes.

Depreende-se então que os crioulos se apresentam como sistemas linguísticos que apresentam similaridades a nível de traços linguísticos, em virtude de as formas captadas pelos aprendentes de L2 envolvidos no contacto dependerem de aspectos compartilhados por estes: idade da aquisição, frequência no *input* disponível, saliência fônica, compreensão, etc. Tal comprovação pode ser observada recorrendo quer à flexão em número quer à configuração estrutural do SN de outros crioulos. Assim, como característica comum aos CP's da Alta Guiné, observa-se que, no SN, os artigos (exemplo [15]), demonstrativos (exemplos [13] e [16]), possessivos (exemplo [12]) e quantificadores (exemplos [14] e [17]) precedem o nome, funcionando como determinantes tanto no CCV como no crioulo da Guiné-Bissau (CGB).

[12] CCV:  
*Nos kàza e grànd* (Cardoso, 1989:30)  
 1PL casa é grande  
 “A nossa casa é grande.”

[13] CCV:  
*kel minina* (Cardoso, 1989:21)  
 DEM menina  
 “Aquela menina.”

[14] CCV:  
 El deve ten *algun dñer.* (Cardoso, 1989:35)  
 3SG deve tem Q dinheiro  
 “Ele deve ter algum dinheiro.”

[15] CGB:  
*N tené un libru.* (Peck Jr., 1988:33)  
 1SG tem INDF livro  
 “Eu tenho um livro.”

[16] CGB: *es gazya li* (Kihm, 1994:136)  
 DEM rapariga aqui  
 “Esta rapariga aqui.”

[17] CCGB: *Parsi-n kuma no ka ten ja mafe* (Kihm, 1994:237)  
 Parece-1SG COMP 1PL NEG tem Q *mafe*  
 “Parece-me que já não temos mais *mafe*.”

Por seu lado, o artigo indefinido deriva do português “um”, mas é invariável, tanto no CCV (exemplo [18]) como no CGB (exemplo [19]):<sup>25</sup>

[18] CCV: *un mnina* (Cardoso, 1989:21)  
 INDF menina  
 “Uma menina/umas meninas.”

[19] CGB: *um karta* (Kihm, 1994:54)  
 INDF rapaz  
 “Uma carta/umas cartas.”

Por conseguinte, nestes crioulos, a indicação de número do referente é dada pelo contexto, pelo que o determinante possui apenas traço semântico de singular (exemplos [18] e [19]). Contudo, os acrolectos, variedades mais próximas da LA, apresentam já algumas flexões, seguindo regras muito próprias para codificar o género e o número. Deste modo, o número PL só é obrigatoriamente marcado num dos elementos da expressão nominal, em geral o primeiro (o referente). Assim, caso o nome seja antecedido de um determinante artigo (exemplo [7]), um demonstrativo (exemplo [20]) ou um possessivo (exemplo [21b]), é sobre este elemento que vai incidir a pluralização, estabelecendo-se aqui, um paralelismo evidente com a forma de marcar a pluralização não só do PA mas também das outras variedades de português reestruturado de África e do Brasil, que temos vindo a observar neste trabalho:

[20] CCV – VARIEDADE DE SANTIAGO: *Kes mininu fémia* (Pereira, 2006:35)  
 DEM meninofêmea  
 “Aqueles meninas.”

- [21] CCV – VARIEDADE DE S. NICOLAU:
- a. **ña** *kàza* (Cardoso, 1989:30)  
 1SG casa  
 “A minha casa.”
- b. **ñas** *kàza* (Cardoso, 1989:30)  
 1PL casa  
 “As minhas casas.”

Se o nome não for antecedido de nenhum determinante PL ou que indique uma quantidade PL, será ele a receber a marca de número (exemplo [22]):

- [22] CCV – VARIEDADE DE SANTIAGO:
- Fidjus** *di Maria sta duenti* (Pereira, 2006:35)  
 “Os filhos da Maria estão doentes.”

Todavia, a marcação de número também pode ser feita com recurso ao morfema *-s*, depois de palavras acabadas em vogal (exemplo [22]), *-is*, depois de palavras terminadas em consoante (exemplo [23a]), ou *-sis*, depois de palavras finalizadas em vogal nasal (exemplo [23b]):

- [23] CCV – VARIEDADE DE SANTIAGO:
- a. *kudjer* > *kudjeris*  
 (Pereira, Arim & Carvalho, 2006:12)  
 “a colher” > “as colheres”
- b. *armun* > *armuns/armunsis*  
 (Pereira, Arim & Carvalho, 2006:12)  
 “o irmão” > “os irmãos”

Note-se, ainda, que os falantes das variedades acrolectais usam também a forma singular para construir frases como as do exemplo [24], em que os nomes não carecem de marca morfológica de PL, visto fazerem parte de um conhecimento geral compartilhado pelos falantes, ou seja, actuarem como indicadores de uma dada identidade PL:

- [24] CCV – VARIEDADE DE SANTIAGO:
- Es tene odju reagladu na kabésa.* (Pereira, 2006:47)  
 Eles tem olho arregalado na cabeça.  
 “Eles estão espantados.”

Formas singulares essas que podem também ser usadas, quando respeito a entidades plurais já referidas anteriormente no discurso, isto é, previamente identificadas como tal:

- [25] CVV – VARIEDADE DE SANTIAGO:  
*Mininu dja txiga.* (Pereira, 2006:93)  
 “As crianças/os meninos/as meninas [já referidos antes] chegaram”

Por fim, importa referir que algumas variedades descrioulizantes do CGB podem apresentar também flexões de número, marcadas no núcleo do SN (exemplo [26]):

- [26] CGB:  
*N tené libru(s).* (Baptista, Mello & Suzuki, 2007:73)  
 1SG tem livro(s)  
 “Eu tenho livros.”

As diferentes formas de marcar o PL nos CP’s atlânticos evidenciam que ocorre um trajecto de inserção de marcas que começa por apresentar paralelismos com o dos substratos, mas que se vai reestruturando, posteriormente, em direcção ao sistema de pluralização da LA. Por outro lado, quer as diferentes maneiras de marcar o número quer a forma como se configura o SN poderão levar-nos a entender que a simplicidade da estrutura destas línguas é, afinal, aparente. Isto mesmo pode ser confirmado por construções praticadas pelos falantes nativos do crioulo de Santiago ou do santomense, que, a exemplo do que sucede com os falantes nativos do português, podem mudar a ordem básica dos constituintes da sua língua (SVO), bastando que recorram, para tal, à alteração de traços prosódicos a fim de apresentar uma frase deslocada à esquerda (exemplos [27] e [28]):

- [27] CCV – VARIEDADE DE SANTIAGO:  
*Es sok, bo traz-êl pra li*  
 (Pereira, Arim & Carvalho, 2006:22)  
 DEM saco 2SG traz-3SG para cá  
 “Esse saco, trá-lo para cá”

- [28] FORRO:  
*Inen poto se, mosu fis’ e.* (Hagemeijer, 2007:19)  
 3PL porta MSPECrapaz fechar 3SG  
 “As portas, os rapazes fecharam-nas.”<sup>26</sup>

Analisando a concordância de número internamente ao SDET do CCV, Castro & Pratas (2006) indicam que a mesma pode ocorrer de acordo com as seguintes possibilidades:

- (i) Se existir um DET no SDET, somente este elemento apresenta o morfema *-s*, marca de PL, uma situação que encontra paralelo evidente no PA:

[29] CCV:  
*Uns / alguns / kes / nhas livru bonito* (Castro & Pratas, 2006:12)  
 um-PL / algum-PL / este-PL / meu-PL livro bonito  
 “uns / alguns / estes / os meus livros bonitos”

[30] PA:  
 a. *Quero vi passia, fica uns dia aqui* [OSVALH1]  
 “Quero vir passear, ficar uns dias aqui”  
 b. *fico lá alguns tempo* [MMDEUSM3]  
 ?“Fico lá alguns tempos”  
 c. *mataram esses home* [MANOH3]  
 “mataram esses homens”  
 d. *fazê meus serviço* [SALVH3]  
 “fazer os meus serviços”

- (ii) Caso o DET seja representado por uma categoria vazia, a marcação de PL é inserida no primeiro elemento do SDET, seja ele um nome (exemplo [31a]) seja ele um adjetivo (exemplo [31b]). Estas construções também encontram paralelismos nas estruturas do PA (exemplos [32a] e [32b]):

[31] CCV:  
 a. *livrus bonito* (Castro & Pratas, 2006:12)  
 livro-PL bonito  
 “livros bonitos”  
 b. *purmerus livru* (Castro & Pratas, 2006:12)  
 primeiro-PL livro  
 “primeiros livros”

[32] PA:  
 a. *Eu faço trabalhos privado.* [CASTEH1]  
 “Eu faço trabalhos privados”  
 b. *Temo que utilizá arma, e com boas barbatana* [OSVALH1]  
 “Temos que utilizar arma, e com boas barbatanas”

(iii) As ocorrências com quantificadores universais apresentam, por seu lado, três possibilidades. Caso exista um DET (demonstrativo ou possessivo), a pluralização é realizada neste (exemplo [33a]); se o quantificador universal estiver localizado depois do nome, a marca de PL é inserida neste (exemplo [33b]); por fim, quando o quantificador universal é colocado inicialmente e sem DET's, não é inserida a marca de PL –s em nenhum dos elementos do SDET (exemplo [33c]). Construções similares são também detectadas no PA, uma vez mais (exemplos [34a], [34b] e [34c]):

- [33] CCV:
- a. *tudu kes / nhas mininu*  
(Castro & Pratas, 2006:13)  
todo este-PL / meu-PL menino  
“todos estes meninos / todos os meus meninos”
  - b. *mininus tudu* (Castro & Pratas, 2006:13)  
menino-PL Q  
“os meninos todos”
  - c. *tudu mininu* (Castro & Pratas, 2006:13)  
Q menino  
“todos os meninos”

- [34] PA:
- a. *fez tudo **essas** coesa* [MANEVEM2]  
“fez todas essas coisas”
  - b. *já cozeu **folhas** toda na panela* [ANTOM1]  
“já cozeu as folhas todas na panela”
  - c. *nós aqui, tudo criança daqui pa brincadera pode falá diolecto*  
[MINISM3]  
“nós aqui, todas as crianças daqui podem falar dialecto nas brincadeiras”

A propósito da construção em que o quantificador universal é colocado inicialmente sem DET's e que leva a que nenhuma marca de PL –s seja inserida nos outros elementos do SDET (exemplo [34c]), veja-se como o informante [CASTEH1], do estrato mais novo e mais escolarizado de Almojarife (8º ano de escolaridade, Tabela 8.1), começa por realizar uma estrutura idêntica à do crioulo, mas tem percepção da variação e tenta corrigi-la, de imediato, de acordo com a regra da concordância do português:

- [35] PA:  
*isso e todo documento... os documentos que eu tenho* [CASTEH1]

Soares & Grosjean (1984:385), abordando as questões inerentes ao acesso ao léxico por parte dos falantes bilíngues, defendem que, que estes têm de aceder a duas redes linguísticas distintas, nas quais se encontram os armazenamentos lexicais de cada uma das suas línguas, ou seja, a forma (fonologia e propriedades sintáticas) e o significado das palavras, mas que se encontram interconectadas no que concerne às representações conceptuais. Assim, estes falantes são capazes de falar numa só língua, mas buscam primeiro as palavras no léxico menos activado, apresentando interferências latentes deste na língua em que estão a comunicar, ou seja, o falante usa a língua com códigos alternados e com recurso a léxico emprestado (Paradis, 1981, 1985, 2001 [1989]).

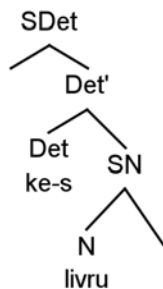
Por seu lado, Castro & Pratas (2006), seguindo as propostas avançadas em termos de MD (Halle & Marantz, 1993, 1994; Noyer, 1997; Embick, 1997; Embick & Noyer, 2001), que abordaremos no ponto 2.6.2.1 do presente trabalho, entendem que não existe Léxico,<sup>27</sup> pelo que as funções que lhe são atribuídas são distribuídas entre outras componentes da gramática.<sup>28</sup> Deste modo, e conforme esquemas das Figs. 16 e 17, a Sintaxe manipula apenas as categorias abstractas dispostas nas estruturas não lineares. Depois da Sintaxe, a Inserção Vocabular (*Vocabulary Insertion*) insere as formas fonológicas – itens de vocabulário – nas estruturas sintáticas abstractas. Por seu lado, a Ordem Linear (*Linear Order*) é uma propriedade unicamente da representação fonológica, enquanto a Linearização (*Linearization*) ocorre concomitantemente com a Inserção Vocabular. Contudo, previamente a esta, a propriedade da Ordem Linear (adjacência e relações precedentes) não pode ser definida numa estrutura sintáctica.

Tal como procederia Baxter (2009) para o PT e o dialecto de HEL-Ba, Castro & Pratas (2006) levaram igualmente em conta as assunções de Costa & Silva (2006b) acerca das diferenças na concordância de número internamente ao SDET no PE (concordância plena) e no PB (número marcado quer nos artigos definidos ou indefinidos e demonstrativos, núcleos DET, quer em todos os elementos pré-nominais). Partindo do pressuposto que o PL é um morfema *singleton* (Costa & Silva, 2006b), Castro & Pratas (2006) assumem que somente o DET carrega a marca de PL, uma vez que é neste elemento que ancora a informação de número (Enç, 1991; Ferrari-Neto, Corrêa & Augusto, 2005; Castro & Ferrari-Neto; 2007), sendo redundante a marca morfofonológica de PL no nome, uma vez que o que conta é a sua expressão em DET

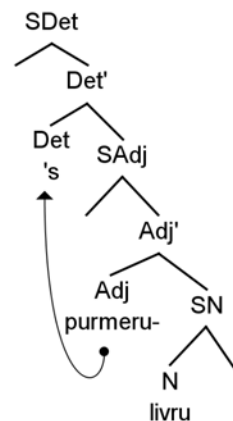


(Castro & Ferrari-Neto, 2007:75). Ora, como vimos nos exemplos que fornecemos, esta é uma situação igualmente válida para o PA. Nesta conformidade, seja no CCV seja no PA, o DET é uma extensão da projecção da expressão nominal (Abney, 1987), sendo aí realizado o morfema de PL, em virtude de assumir a posição de núcleo do SDET. Por seu lado, os itens como *purmeru* (exemplo [31b]), no CVV, e “boa” (exemplo [32b]), no PA, têm que ser observados como complementos pré-nominais (AP’s), e não como adjuntos ou especificadores (Cinque, 1994) que carregam complementos do SN (Abney, 1987; Castro, 2004).

Segundo Costa & Silva (2006b), em línguas do tipo do CCV, o marcador de PL emerge pós-sintacticamente como um morfema do tipo *singleton*, visível no núcleo DET através de um sufixo. Caso haja um elemento visível em DET (Fig. 2), o sufixo é atribuído aos elementos patentes nos exemplos [29] e [33a], para o CCV, e exemplos [30 (todos)] e [34a], para o PA. Se não existir elemento visível em DET (Fig. 3), o sufixo “baixa” e é atribuído ao núcleo complemento (exemplos [31a], [31b] e [33b], para o CCV, e exemplos [32a], [32b] e [34b], para o PA. Internamente ao SDET, a marca PL “baixa” da posição vazia DET e é atribuída ao nome nu.



**Fig. 2.** Marcação plural no CCV  
(Marcador visível em DET).



**Fig. 3.** Marcação plural no CCV  
(Ausência de marcador em DET).

Contudo, este tipo de concordância morfológica reduzida não é válida para o exemplo [33c], do CCV, nem para o exemplo [34c], do PA, uma vez que nenhum dos elementos do SDET recebe marcação. Tal deve-se ao facto de em estruturas do tipo Q+N, o quantificador “*tudu*” deter a função de modificador do SDET e, conseqüentemente, a

sequência ser interpretada como massiva, não necessitando nenhum dos itens de receber marca morfológica de PL (Maling, 1976:716; Vicente, 2006). Contrariamente, nos exemplos [33a] e [33b], do CCV, e exemplos [34a] e [34b], do PA, a marca de PL é inserida porque, em ambos os casos, as sequências são interpretadas como definidas. Assim, na primeira sequência (Q+DET+N), Q é núcleo do SQ, enquanto o SDET é o seu complemento, sendo válida a premissa de que o marcador de PL é atribuído como sufixo do tipo *singleton* ao núcleo DET (Costa & Silva, 2006b). Quanto à segunda situação, verifica-se uma estrutura do tipo N+Q que requer algum tipo de especificação pós-quantificador, apesar de a interpretação ser definida. Neste caso, Q é o núcleo do SQ, um adjunto à esquerda do SX.

Partindo do princípio que a GU actua na aquisição de uma L2 através da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006) e que durante o processo de aquisição o falante vai testando sucessivas possibilidades gramaticais de aproximação à LA, criando ele próprio uma interlíngua que se vai reformulando à medida que adquire novos dados (Corder, 1967; Leiria, 1996), Cardoso (2005) aponta algumas possíveis interferências do crioulo de Santiago na aquisição do PtgL2 falado em Cabo Verde. Tendo em conta que, neste arquipélago, a aquisição do PtgL2 ocorre essencialmente a nível formal, isto é, na escola, configura-se uma situação que propicia a ocorrência de interferência forte ou moderada na interlíngua, muitas vezes responsável por “*ruptura tipológica*” (Mota, 1996:519), já que “*as regras de concordância no crioulo distanciam-se do português como o céu da terra, tanto em número como em género*” Silva (1998:114). Assim, a nível morfossintático, o PtgL2 que tem o CCV como substrato apresenta interferências caracterizadas pela substituição das categorias gramaticais funcionais ou de determinadas regras gramaticais, por outras semelhantes da L1 (Cardoso, 2005:13). No caso da pluralização, estes aspectos são evidentes nas realizações linguísticas do PCV que evidenciam ausência de realização do artigo definido (exemplos 36a. e 36b.), isto é nomes nus, ou de marcação de número (exemplos 37a. e 37b.), apresentando transposições estruturais do CCV, em tudo semelhantes às realizadas no PA, e que abordaremos já no próximo ponto deste trabalho:

[36] PCV:  
 a. *O sobrinho não deu corda* (Cardoso, 2005:14)

↓  
 ∅

CCV:  
 b. *Xibinhu ka da kórda* (Cardoso, 2005:14)  
 sobrinho ASP dar corda  
 “o sobrinho não deu a corda”

[37] PCV:  
 a. *O sobrinho disse lobo para amarar o macaco pé.* (Cardoso, 2005:14)

↓  
 ∅

CCV:  
 b. *Xibinhu fla lobo pa mara makaku pé* (Cardoso, 2005:14)  
 Sobrinho falar lobo para amarrar macaco pé  
 “O sobrinho disse ao lobo para amarrar os pés do macaco”

Aplicando a mesma metodologia que utilizara para analisar a CPL-var no SN do PMp, Jon-And (2009) observa a influência de cinco variáveis linguísticas – *saliência fônica, posição em relação ao núcleo, posição linear, categoria gramatical e animacidade* – e duas variáveis extralinguísticas – *escolaridade e idade de início de aquisição (age of onset) do português* – para determinar as motivações que subjazem à CPL-var nos SN’s produzidos por falantes do PtgL2 do Mindelo, que têm o crioulo acrolectal de São Vicente/Santo Antão (Barlavento) como L1. Ambas as variáveis sociais foram rejeitadas como não significantes para a marcação, o que levanta a possibilidade de o facto de a LA ser apenas aprendida em contexto formal não ser suficiente para eliminar o fosso tipológico entre o PE e o PCV (Cardoso, 2005:11).

Relativamente às variáveis linguísticas *categoria gramatical e posição em relação ao núcleo* cruzada com a *posição linear*, ambas apresentam um efeito forte para a marcação PL, possivelmente por interferência do substrato (Jon-And, 2009:7), confirmando-se as substituições de categorias ou regras gramaticais da LA por outras semelhantes da L1 (Cardoso, 2005:13). Estes efeitos são semelhantes no PCV e outras variedades africanas e variedades brasileiras de português (Jon-And, 2009:7), fazendo pressupor, uma vez mais, que a instanciação de novos/errados parâmetros e sua frequência no *input*, que apaga a possibilidade de ocorrer reajustamento paramétrico (Gonçalves, 2004:229), leva a que o PCVL1 apresente uma nova reinstanciação paramétrica, com a variação

caminhando no sentido da mudança. Assim, o fenómeno que resulta da convergência entre posição linear e classe gramatical poderá, também aqui, ser associado ao sistema de marcação PL dos substratos africanos do CCV (cf. Baxter, 1998, 2009), com alguns traços das L1's sendo retidos neste e sobrevivendo na gramática da variedade reestruturada (cf. Winford, 2003a:245-247), ao mesmo tempo que algumas propriedades gramaticais da LA se esvanecem no *continuum* de aquisição (Gonçalves, 2004:229). Das similaridades registadas entre as gramáticas do PCV, adquirido por falantes de um CP, e as gramáticas do PVB, PT, PVA e PM, adquiridos por falantes cujos substratos não são CP's, parece resultar evidente que não ocorrem divergências substanciais no que concerne aos padrões de variação (Figueiredo, 2008:37), já que todos estes dialectos têm na sua génese uma situação sociohistórica específica, isto é, de transmissão linguística irregular.

#### **1.2.2.2. Português de Almojarife (São Tomé) e seu substrato (santomense)**

O PA reestruturado, língua falada por uma comunidade semi-isolada de São Tomé, tem, no bojo da sua formação e evolução, uma extensa situação de contacto entre línguas, que se iniciou logo com a colonização do Golfo da Guiné, em 1485. Nesta fase, ocorreram as primeiras e esparsas tentativas de fixação de colonos na ilha, que aí estabeleceram rudimentares engenhos do açúcar (Negreiros, 1895:33), recorrendo à mão-de-obra de escravos resgatados no Benim e no reino do Congo (Brásio, 1952-1:54). Muitos destes escravos seriam já falantes de um proto-crioulo (Hagemeijer, 2009), que ascendeu a língua de comunicação interétnica na ilha. Esta evoluiria para crioulo, uma vez que as comunidades geográfica e culturalmente deslocadas pelos colonos constituíram uma sociedade nova de falantes de distintas L1's do grupo níger-congo atlântico, que abandonaram as mesmas (excepto em situações de ambiente familiar), visto necessitarem de comunicar entre si e com os seus amos.<sup>29</sup>

No início do século XVI, São Tomé era já um importante entreposto no tráfico negreiro (Ferronha, 1989), iniciando-se o período da colonização massiva da ilha (1493 a 1515). Nesta altura, o número de colonos fora já ultrapassado pelo dos escravos, originando uma situação de parco contacto linguístico entre ambos, e que fazia com que os segundos tivessem cada vez menos acesso às formas da LA. Todavia, o número de

mestiços e falantes bilingues exponenciou-se quando D. João II de Portugal, por decreto régio, atribuiu aos colonos o direito de tomarem uma mulher escrava por companheira (Brásio, 1952-1:331-332). Ao conceder liberdade às mulheres africanas tomadas pelos convictos, bem como aos seus descendentes (Brásio, 1952-1:331-332), primeiro (1515), e às concubinas (Brásio, 1952-1:376), depois (1517), o rei D. Manuel I permitiu que a franja da população liberta assimilasse ainda mais a cultura e traços linguísticos dos colonos, constituindo a comunidade de forros (escravos que recebiam carta de alforria). Quanto aos mestiços nascidos destas relações, bilingues em português e crioulo, ganharam projecção na vida política, social e económica da colónia, sendo decisivos para o processo de complexificação e distensão léxico-gramatical do santomense (Ferraz, 1979:17). Após conflitos vários com escravos revoltosos e a tomada do arquipélago pelos holandeses (1641), os colonos abandonaram temporariamente este e o santomense sistematizou-se na forma próxima do actual basilecto (Ferraz, 1979:19). Presentemente, este crioulo é uma das línguas de comunicação generalizada em São Tomé, sendo falado por 72,4% da população (Instituto Nacional de Estatística de São Tomé e Príncipe, 2003/2005). Não obstante deter um léxico em grande parte de origem portuguesa, as suas características morfossintáticas estabelecem paralelismos com as L1's africanas faladas pelos ancestrais escravos transplantados (Hagemeijer, 1999, 2009). O seu substrato linguístico será maioritariamente kwa (falantes oriundos do Benin) e bantu (falantes provenientes da Guiné, Gabão e Angola) (Ferraz, 1975, 1979). Recentes pesquisas têm evidenciado que o proto-crioulo do Golfo da Guiné, que representa a raiz dos quatro crioulos desta região (Hagemeijer, 1999, 2009), evidencia traços que o relacionam quer com o edóide quer com os idiomas bantóides da área H, especialmente com as variedades kongo (Lorenzino, 1998; Hagemeijer, 1999, 2007, 2009; Schang, 2000; Tomás *et alii*, 2002; Güldemann & Hagemeijer, 2006).

No caso concreto da estrutura sintagmática do santomense, uma das formas de marcar o PL recorre ao determinante definido *inen* (exemplo [38b]), homófono do pronome pessoal, 3ª pessoa PL, possivelmente proveniente do pronome do edo,<sup>30</sup> língua do grupo níger-congo atlântico, falada na Nigéria, e colocado em posição pré-nominal para funcionar como marcador de flexão em número, isto é, como elemento que determina a

inserção da marcação PL em nomes de traço [+humano]. Quanto aos restantes itens do SN, mantêm-se invariáveis, sendo iguais para o singular e para o PL:

- [38] FORRO: (Hagemeijer, 2000:118)
- a. *ome plejidu*  
 homem preguiçoso  
 “o homem preguiçoso”
- b. ***inen*** *ome plejidu*  
 3PL homempreguiçoso  
 “os homens preguiçosos”

Estruturalmente, o sistema de marcação PL no SN do santomense estabelece paralelismos quer com línguas do grupo níger-congo (exemplo [39]) quer com construções do PA (exemplo [40]):

- [39] IORUBÁ:  
***àwon*** *omo*<sup>31</sup>  
 3PL criança  
 “as crianças”

- [40] PA:  
*já trabalhê esses lugar tudo* [LUISH2]  
 “já trabalhei nesses lugares todos”

A transposição da estrutura frásica do grupo níger-congo atlântico por falantes nativos pós-adolescentes, em situação de aquisição do PtgL2, determinou, no santomense, o uso de construções sintagmáticas em que a pluralização é inserida apenas no primeiro elemento pré-nominal. Aliás, são várias as similaridades entre o santomense e o seu substrato níger-congo, a nível do sistema morfossintático (Ferraz, 1975; 1979). Uma delas manifesta-se na presença de pronomes disjuntivos em frases preposicionadas, tanto nas construções interrogativas como nas relativas (exemplos [41] e [42]):

- [41] FORRO: (Ferraz, 1975:36)
- Nãji su bi de?*  
 Onde você vem de?  
 “De onde vem você?”

- [42] ZULU: (Ferraz, 1975:36)
- Umuzi engavela kuwo*  
 Aldeia eu~vim de  
 “A aldeia de onde eu vim”

No processo de aquisição do português por falantes do santomense L1 ocorreu transferência do sistema estrutural de pluralização deste para a L2. Posteriormente, o mesmo sistema transitaria geracionalmente (cf. Winford, 2003a:245-247) no processo de nativização do PtgL2, com base nos DPL's defectivos transmitidos em cadeia, e que determinaram uma nova reinstanciação paramétrica no PtgL1 (PA). Deste modo, na situação de transmissão linguística irregular originada, os ancestrais falantes do PA marcaram a pluralização com recurso a morfemas do *singleton* na categoria funcional DET, que passou a servir de âncora para o controlo da pluralização. Este tipo de marcação está também presente noutras variedades africanas e brasileiras de português, tenham elas CP's atlânticos ou línguas do grupo níger-congo atlântico como substrato directo ou ancestral:

- [43] PLd:  
*vigia as criança* (Mingas, 2000:67)
- [44] PMp:  
*Há muitas dificuldade nas escola.* (Gonçalves & Stroud, 1998:110)
- [45] PT:  
*issos coesa tudo* (Baxter, 2004:98)  
“essas coisas todas”
- [46] PCV:  
*as coisa foram subindo* (Jon-And, 2009:1)
- [47] MRJ:  
*meus filho casado* (Scherre, 1988:356)

Ainda a propósito da marcação PL no PA, refira-se que detectámos também possíveis resquícios do recurso ao uso do pronome pessoal, 3ª pessoa PL, da LA (exemplo [48]), um fenómeno que abordaremos mais em pormenor no ponto 2.11.3.1.3 do presente trabalho:

- [48] PA:  
*outra pessoa foi tirô eles rôpa, viu arma aqui [ZECAH1]*  
“outra pessoa foi buscar a(s) roupa(s) e viu a arma lá”  
(?) “outra pessoa foi buscar a(s) roupa(s) deles e viu a arma lá”

O tipo de marcação PL estrutural do santomense determina, igualmente, que os nomes e os adjectivos permaneçam invariáveis, baseando-se nas formas masculinas singular do português (terminadas em *-o*), quando esta língua as marca em género (exemplo [49]):

- [49] FORRO:  
*gatu* (Ferraz, 1979:60)  
 “o gato/a gata”

Este fenómeno demarca-se do sistema da LA, visto que o PE implica concordância entre o género do núcleo do SN e os seus determinantes e qualificadores (adjectivos):

- [50] PE:  
*eu tenho um irmão simpático* > *eu tenho uma irmã simpática*

No kimbundu, por exemplo, quase todos os nomes que referem a ideia de sexo são epicenos, salvo raras excepções, como acontece com os termos *tata* (pai) > *mama* (mãe), *diala* (homem) > *muhetu* (mulher) ou *dikolombulu* (galo) > *sanji* (galinha).<sup>32</sup> Verifica-se então que, para as diversas classes de nomes representadas por afixos que fazem a diferenciação entre singular e PL, “as distinções sexuais não são importantes” (Welmers, 1973:159). Deste modo, caso não se realize um dos nomes atrás mencionados, a distinção masculino/feminino é fornecida quer pelos termos *diala* e *muhetu*, que acompanham o nome no caso de ele deter o traço [+humano] (exemplos [51a] e [51b]) ou [+animado] (exemplos [52a] e [52b]). Por vezes, com substantivos do tipo [+animado] e [+humano] referentes a entidades que ainda não atingiram a idade adulta podem ocorrer também os vocábulos *ndumbi* (macho) e *mukaji* (fêmea). Contudo, importa reter que o nome núcleo do SN se mantém invariável:

- [51] KIMBUNDU:  
 a. *mona adiala* (Zau, 2001:24)  
     filho PREP+homem  
     filho de homem  
     “filho”  
 b. *mona amuhetu* (Zau, 2001:24)  
     filho PREP+mulher  
     filho de mulher  
     “filha”



- [52] KIMBUNDU:
- a. *mbua adiala ami* (Zau, 2001:23)  
 cão PREP+homem 1SG  
 cão de homem meu  
 “o meu cão”
- b. *mbua amuhetu ami* (Zau, 2001:23)  
 cão PREP+mulher 1SG  
 cão de mulher meu  
 “a minha cadela”

Como se verifica, a marcação de género restringe-se aos elementos nucleares, já que os outros constituintes do SN se mantêm inalterados. Quanto aos CP’s da Alta Guiné e do Golfo da Guiné, é também a partir do núcleo do SN que se recompõe a morfologia do género. Deste modo, configura-se um claro paralelo no plano da morfossintaxe de género entre estes CP’s e as línguas bantu, com as construções do santomense a revelarem que o valor referencial do sexo ocorre, por norma, com os nomes adjectivados ‘*ome* (homem) e ‘*muala* (mulher), os quais acompanham o nome, mantendo-se este invariável nas formas masculina e singular:<sup>33</sup>. Formam-se, assim, nomes compostos que, no PE, têm forma simples:

- [53] FORRO:
- a. *bué ‘ome ame* (Ferraz, 1979:60)  
 boi homem 1SG  
 “o meu boi”
- b. *bué ‘muala ame* (Ferraz, 1979:60)  
 boi mulher 1SG  
 “a minha vaca”

Lucchesi (2000a), para a variedade de HEL-Ba, e Figueiredo (2009a), para o PA, concluíram que a marcação de género se dá a partir do núcleo nominal em ambos os dialectos, o qual surge como pólo irradiador da marcação mórfica de género. Não obstante, ambas as variedades de português apresentam não-realização de matéria gramatical da LA, com erosão da morfologia flexional do nome afectando mais o mecanismo morfossintáctico da concordância do que a flexão mórfica do género dos núcleos nominais. Desta forma, ao desaparecer o mecanismo morfossintáctico de função meramente gramatical para a indicação do género, conservou-se a flexão deste nos substantivos dos seres animados recorrendo-se ao valor referencial do sexo através de nomes adjectivados do tipo *macho/fêmea* ou *homem/mulher*, a exemplo do que sucede

nos substratos, sejam eles bantu sejam eles CP's, como o santomense ou o CCV (exemplo [54]). Assim, é possível encontrar no PA (exemplos [55a] e [55b]) realizações cujas marcações de género são idênticas não só às do seu substrato (o santomense) e do CCV mas também às de algumas línguas do grupo níger-congo atlântico:

[54] CCV:  
*Mininu fêmia é spértu* (Pereira, 2006:34)  
“As meninas são espertas”

[55] PA:  
a. *mia mãe nõ pariu homem macho.* [MANEVEM2]  
“A minha mãe não deu à luz/teve rapazes”  
b. *Dois filho mulher qu'ê tem* [FRANCM2]  
“As duas filhas que eu tenho”

O mesmo tipo de variação na marcação de género é ainda detectável tanto no PVA (exemplo [56]), reflectindo transposições literais do substrato kimbundu, como nas falas rurais de Piracicaba, São Paulo, resultantes do contacto da língua portuguesa com dialectos do grupo ameríndio tupi-guarani e africanos do grupo níger-congo atlântico. Em Piracicaba (exemplo [57]), tal como sucede em Almojarife (exemplo [55a]), o uso dos substantivos adjectivados chega mesmo a marcar redundantemente a distinção de sexo nos nomes com traço [+animado]:

[56] PVA:  
*O meu avô de mulher come banana.* (Zau, 2001:24)  
“A minha avó come banana”.

[57] PIRACICABA:  
*Tenho mais fê em menina mulher do que em menino homem.*  
(Rodrigues, 1974:52)

Apesar de este tipo de marcação de género não estar presente no PE, não deve ser desconsiderada a possibilidade de representar uma estratégia natural aplicada em situações de dificuldade comunicativa intercultural, visto que as línguas kwa, também pertencentes ao grupo níger-congo, recorrem, frequentemente, ao género gramatical. Esta é, aliás, uma situação que também acontece no santomense:

- [58] FORRO:  
 a. *soglu* (Ferraz, 1979:60)  
     “o sogro”  
 b. *sogla* (Ferraz, 1979:60)  
     “a sogra”

Os dois tipos de marcação de género existentes no santomense (recurso à marcação com morfema sufixal e às construções que apelam ao valor referencial de sexo) poderão, também, estar na origem da hesitação na aplicação da regra de concordância em género que se verifica no PA:

- [59] PA:  
*faço **trabalhos caseira** até à tarde* [CARMOM1]

A exemplo do que sucede no PA, outras variedades de português, sejam africanas sejam brasileiras, também apresentam CGE-var, um fenómeno já apontado em alguns estudos (p.e. Baxter, 1998):

- [60] PT:  
*agora que tem **muitos capoeira** tá chamare cobra* [Isabel Simba – M2]<sup>34</sup>
- [61] PVA:  
***no mesmo barriga*** (Mendes, 1985:149)
- [62] PMp:  
*Eu tenho **minhas netos*** (Gonçalves & Stroud, 1998:107)
- [63] HEL-Ba:  
*ôpero **no perna e tudo*** (Lucchesi, 2000a:207)

Estes tipos de marcação poderão ter implicação directa no sistema de pluralização de todas estas variedades de português. No caso da marcação do tipo sufixal, e como debateremos no ponto 3.6.7, dedicado à constituição da variável independente *marcação de género*, constituída para constatar se existe uma correlação entre a marcação de género e a marcação PL em cada constituinte do SN, o morfema de género, por ser mais directamente conectado à estrutura mórfica do item em análise, visto estar estreitamente ligado ao radical deste, é aplicado anteriormente ao da marcação de PL (Müller, 1994a; Franseschina, 2002, 2005; Corrêa, Name & Netto, 2004; Montrul, 2004; White *et alii*,

2004;). Deste modo, é de esperar que a marcação de número PL, por ser posterior à do género, possa ser condicionada por esta, quando ela ocorre. Por seu lado, a marcação de género com recurso ao valor referencial de sexo terá determinado que a variável independente *marcação de género* não fosse considerada relevante pela ferramenta computacional VARBRUL para a questão da CPL-var no SN.

Continuando a observar aspectos relacionados com a estrutura do SN do santomense, importa referir que, em trabalho que aborda os nomes nus e o domínio nominal no santomense, Alexandre & Hagemeyer (2007:42) defendem que este idioma apresenta um elevado grau de sensibilidade em relação à identificação dos nomes, em virtude de a relevância destes se limitar ao domínio estrutural do SN. Nos casos da marcação de género e número, a sensibilidade aos princípios semânticos determina que a correcta interpretação das frases requeira não apenas a interpretação do significado lógico-semântico mas também a interpretação do significado informal (Vallduví, 1993:12-13). Nesta conformidade, os traços da animacidade do nome jogam um papel importante na marcação de número e de género, com a inserção de marcas revelando apetência pelos nomes com traço [+humano], visto que estes são interpretados como elementos individuais, enquanto os nomes com traço [-humano] são tidos como mais genéricos (Alexandre & Hagemeyer, 2007:41). Com a marcação de género dependendo da animacidade, os nomes humanos (exemplos [58a] e [58b]) ou extensões do humano, como as profissões (p.e. *kuzinhêru* > *kuzinhêra*), apresentam distinção lexical de género, ao passo que os nomes não humanos necessitam de informação adicional de género (exemplo [53a] e [53b]) (Alexandre & Hagemeyer, 2007:42, nota 7).

Relativamente à estrutura do SN, constata-se, a exemplo do que sucede nos outros CP's da Alta Guiné, a ocorrência do determinante indefinido invariável *ũa*<sup>35</sup> antes do nome núcleo (exemplo [64]) para marcar o número singular, sendo a ausência de flexão atribuída à influência dos substratos africanos (Ferraz, 1975:36):

[64] FORRO:  
*ũa ja* (Alexandre & Hagemeyer, 2007:46)  
“um dia”

Tendo em conta as questões da animacidade em relação à marcação de género e número, o determinante lexical *ũa* transporta especificações de número e definitude

(Alexandre & Hagemeyer, 2007:40), introduzindo informação nova nos contextos sintáticos (Lucchesi, 1993) e tendo como oposto o quantificador *ũa dôsu* (“alguns”):

[65] FORRO:  
*ũa dôsu ome* (Alexandre & Hagemeyer, 2007:40)  
 um dois homem  
 “alguns homens”

Quanto ao sistema de pluralização (exemplo [38b]), Alexandre & Hagemeyer (2007:42) entendem estar-se na presença de um determinante definido marcador de PL, que se terá desenvolvido a partir do âmbito do domínio dos demonstrativos, já que a literatura não é consensual acerca da origem etimológica do mesmo. Considerando-se ainda a relação entre animacidade e marcação de número, verifica-se que a sequência *inen+N* apenas se realiza com nomes de traço [+humano], expressando um significado definido que diz respeito a um grupo não-individualizado familiar ao ouvinte e ao falante. Assim e a exemplo do que sucede com *ũa*, para o singular, o determinante definido lexical *inen* carrega, para o PL, as especificações de número [+plural] e [+definitude] (Alexandre & Hagemeyer, 2007:42). Este aspecto aponta para uma conexão entre o SN definido singular e ausência de artigo definido, estabelecendo semelhanças quer com o CCV (Baptista, 2007) quer com HEL-Ba (Baxter & Lopes, 2009). Contudo, como o santomense, ao contrário dos Crioulos da Alta Guiné, de HEL-Ba e do português, não exhibe marcadores de PL em forma de sufixo do N, isto é, não possui marca de número visível, o seu oposto singular é o SN nu (SNN) (exemplo [66]), que abrange um vasto campo de funções semânticas distribuídas por todas as posições sintáticas, sejam elas predicativas sejam argumentativas (Alexandre & Hagemeyer, 2007:37). Como tal, os determinantes nulos devem receber o mesmo tratamento dos determinantes lexicais, podendo o SNN ser totalmente integrado no mesmo tipo de estrutura nominal que incorpora os determinantes lexicais (ponto 1.3.2.2.2, dedicado à estrutura do SN no forro).

[66] FORRO:  
*Maji vimpema ka<sup>36</sup> fe ome mali* (Alexandre & Hagemeyer, 2007:50)  
 Mas vinho de palma ASP faz homem mal  
 “Mas o vinho de palma é mau para os homens/um homem.”

Outra forma de pluralizar os nomes faz-se com recurso à reduplicação (Ferraz, 1979:58-61), que é também usada para marcar a intensificação:

[67] FORRO:  
*Ne anka se godo-godo* (Ferraz, 1979:58)  
 3PL caranguejo MSPEC gordo gordo  
 “Aqueles caranguejos muito gordos”

Os itens utilizados como pluralizadores na reduplicação podem: (i) intensificar o PL (exemplo [68]); (ii) fornecer o significado de “*cada um dos dois*” (exemplo [69]):

[68] FORRO:  
*E pega ka se kota mō-mō-mō-mō-mō d~e* (Ferraz, 1979:59)  
 3SG pegou caranguejo MSPEC cortou mão-mão-mão de-3SG  
 “Ele pegou naquele caranguejo e cortou-lhe todas as patas”

[69] FORRO:  
*ome-ome o?* (Ferraz, 1979:59)  
 Homem-homem INT  
 “São ambos rapazes?”

Um dos tipos de reduplicação do santomense elimina a terminação do participio passado no primeiro elemento do composto:

[70] FORRO:  
*Ine sa zunta-zuntadu* (Ferraz, 1979:60)  
 3PL está junto-junto  
 “Eles estão todos juntos (misturados)”

Lorenzino (1998:119) considera a estratégia da reduplicação uma fonte produtiva deste tipo de construções nos crioulos. Ao que tudo indica, estar-se-á, uma vez mais, perante uma estratégia natural que visa facilitar situações de comunicação. No PA são também comuns as reduplicações (exemplo [71]), mas, tal como no português (exemplo [72]), representam estratégia de marcação do grau superlativo dos adjectivos. Daí que, na caracterização dos nossos dados (ponto 3.1.4), tenhamos optado por considerar apenas o primeiro item da reduplicação para efeitos de codificação, a fim de não levar à análise um número desproporcionado de instâncias com o mesmo tipo de intenção semântica, que acabariam por viciar os resultados finais (cf. Tagliamonte, 2006:94):

[71] PA:  
*como esses cacau **alto alto*** [CLOTIM2]  
“como esses cacaeiros muito altos”

[72] PE:  
*Este vestido é **lindo, lindo**.*

No PA, há ainda outras estruturas que evidenciam paralelismo com as do seu substrato crioulo, e que têm influência na CPL-var do SN. No caso do uso do determinante artigo definido, por exemplo, ocorre uma variação do tipo ternário, já que o mesmo tanto pode surgir sob a forma do determinante do PE (*o, a, os, as*, exemplo [75]) como ser omitido (nomes nus, exemplo [73] e [74]), ou ainda aparecer representado pelo pronome pessoal terceira pessoa do PL *eles*, (exemplo [48]):

[73] PA:  
*faço trabalhos caseira até tarde* [CARMOM1]  
“faço os trabalhos caseiros até tarde”

[74] PA:  
*E amigo meu tinha uma mulher* [OSVALH1]  
“E o meu amigo tinha uma mulher”

[75] PA:  
*É os meus pai que me deu essa ideia.* [OSVALH1]  
“Foram os meus pais que me deram essa ideia”

Estas questões podem ter influência directa na codificação dos nossos dados, já que estruturas simplificadas do tipo DET+N, do PE, podem surgir no PA sem o determinante, implicando que o nome surja isolado na cadeia sintagmática. Deste modo, dados que seriam codificados e levados à análise, foram retirados destas porque configuram realizações sintagmáticas com um único item gramatical.

As realizações com nomes nus do PA (exemplos [73] e [74]) têm respaldo no seu substrato, o santomense. Neste, o uso deste tipo de nomes apresenta uma distribuição e interpretação livre a nível gramatical, mas que é restringida pela componente extra-gramatical. Assim, os nomes nus podem ser licenciados somente na *interface* entre o nível gramatical e extra-gramatical, que relaciona o discurso à pragmática (Alexandre & Hagemeyer, 2007:49). A interpretação de definitude atribuída aos nomes nus deve-se ao

facto de os mesmos terem sido previamente licenciados no discurso, sendo já referenciados pelo ouvinte. Se, pelo contrário, os nomes nus constituem informação nova para o ouvinte, então são entendidos como plurais indefinidos. Como as línguas que têm nomes nus para realizarem funções específicas não apresentam determinantes visíveis para preencherem essas mesmas funções (Crisma, 1997:236), o pressuposto adequa-se ao santomense no que concerne ao indefinido singular *ũa*, uma vez que este dialecto não exhibe SNN's com leitura singular indefinida. Por outro lado, o mesmo pressuposto é também válido para a questão do definido PL *inen*, já que este apenas selecciona nomes humanos, que são preferencialmente interpretados como singulares.

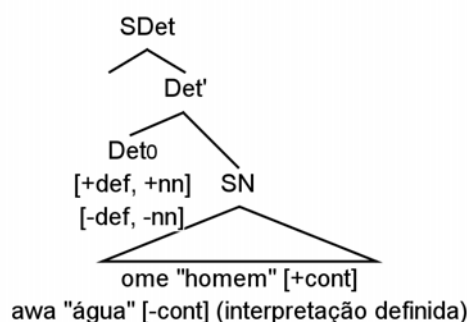
Dado que o sistema de determinantes do santomense usa recursos mínimos para duas funções específicas (*ũa/inen*), as restantes são preenchidas pelos determinantes nulos. Assim, os definidos nulos contemplam, normalmente, o singular (a interpretação preferencial dos SNN's sujeito, que encerram informação previamente licenciada). Quanto aos indefinidos nulos, por norma acompanham o PL (SNN's lexicalmente governados e que contêm informação nova). Contudo, como a informação de número ancora na definitude, a *interface* entre os níveis gramatical e extra-gramatical é crucial para interpretar correctamente as funções dos determinantes nulos. Portanto, os SNN's em posições lexicalmente governadas (V, S) podem ser entendidos como singulares/plurais definidos e plurais indefinidos, mas nunca como singulares indefinidos, que têm de ser precedidos por *ũa* em qualquer posição sintáctica. Por seu lado, os SNN's sujeito não carregam informação nova, pelo que não constituem um passo novo na troca entre definitude e indefinitude, contendo apenas traços do singular, tal como sucede no PA (exemplo [74]), em que o informante já havia feito uso de um ancorador, a fim de licenciar a informação (*ela vivia perto de mô amigo Adento* [OSVALH1] > “ela vivia perto do meu amigo Adérito”). A Tabela 1.1 fornece o panorama sobre a informação contida na estrutura dos SNN's do santomense:



**Tabela 1.1.** Informação encerrada nos SNN's do santomense (Alexandre & Hagemeijer, 2007:56).

Informação nova (indefinido)		Informação licenciada (definido)	
Sujeito	Objecto	Ancorador	Ancorado
SGL	<i>ũa ome/ũa kabla</i> “um homem /uma cabra”	<i>Ome se/Kabla se</i> <i>Ome/Kabla</i>	<i>Ome/Kabla</i> <i>Ome/Kabla</i>
PL	-	<i>Inen ome se/Inen kabla se</i> <i>Inen ome/Kabla</i>	<i>Ome/Kabla</i> <i>Ome/Kabla</i>

Dado que os SNN's exibem um leque variado de leituras que podem ocorrer em posições argumentativas ou predicativas, deverão ser integrados na estrutura que requer um nível de interpretação extra-agramatical proposta na Fig. 5. Esta integração é motivada pelo suporte empírico de que este tipo de SN's projecta sempre um SDET, nível onde ocorre a *interface* entre gramática e discurso pragmático, fundamental para a ancoragem da interpretação (Fig. 4). Note-se que, no santomense, os nomes são sempre interpretados como massivos, requerendo um sistema classificador para a derivação das interpretações contáveis (Alexandre & Hagemeijer, 2007:51). Esta particularidade estabelece parencas tipológicas com a língua chinesa, que Chierchia (1998) classifica como [-predicativa, +argumentativa].



**Fig. 4.** Estrutura sintático-discursiva dos sintagmas nominais nus no santomense (Alexandre & Hagemeijer, 2007:51).

Já no que concerne ao dialecto de HEL-Ba, as similaridades estendem-se muito para além da mera interpretação dos massivos e contáveis, abarcando questões como o desfavorecimento do uso do artigo quer com nomes de referência única ou específica quer com SN's de número singular, ou, em alternativa, o favorecimento da aplicação do artigo definido em SN's nos quais há co-ocorrência deste e modificadores pós-nominais, como o adjectivo, o possessivo ou a oração relativa (Baxter & Lopes, 2009).

Paralelamente, Baxter & Lopes (2009:328-329) notam ainda que o uso do artigo definido pleno em HEL-Ba só é consistentemente favorecido em SN's com função sintáctica de adjuntos adnominais, ao passo que é desfavorecido caso os SN's desempenhem qualquer outra função sintáctica. Estes resultados estabelecem um paralelismo evidente com idênticos achados dos autores para a fala de Salvador (Baxter & Lopes, 2004b) e, no que concerne especificamente aos SN's com funções de sujeito e OD, ocorrem também similaridades entre HEL-Ba e o santomense, já que este não apresenta assimetria na distribuição dos SN's sem determinante de referência definida que figuram como argumentos internos do verbo (Alexandre & Hagemeyer, 2007:57).

As realizações que configuram o uso ternário do determinante no PA têm respaldo, todas elas, em formas do santomense, pelo que importa fazer uma abordagem à estrutura do SN, neste crioulo. Observando a colocação do adjetivo cadeia sintagmática (exemplos [38a] e [38b]), verifica-se que o mesmo só pode ocorrer em posição pós-nominal, tal como sucede nas línguas do ramo bantu (exemplos [76a] e [76b]). No português, o adjetivo não detém esta rigidez formal (exemplos [77a] e [77b]):

- [76] KIMBUNDU:
- |    |  |  |   |
|----|--|--|---|
| a. | <i>diala (di~ala)</i><br>AFXSG+homem<br>“o homem branco”             | <i>dia (di~a)</i><br>AFXSG+PREP(de)        | <i>mundele (mu~nde)</i> <sup>37</sup><br>AFXSG+branco |
| b. | <i>mala (ma~ala &gt; mala)</i><br>AFXPL+homem<br>“os homens brancos” | <i>ma (ma~a &gt; ma)</i><br>AFXPL+PREP(de) | <i>mindele(mi~nde)</i><br>AFXPL+branco                |
- [77] PTG:
- |    |                    |
|----|--------------------|
| a. | <i>casa bonita</i> |
| b. | <i>bonita casa</i> |

A concordância entre os elementos que constituem o SN é controlada nas línguas bantu pelo afixo pré-nominal e concordantes aglutinados em posição prefixal aos restantes itens. A similaridade com o sistema de marcação PL do santomense e do PA, nas situações em que há recurso ao item *inen*, no primeiro, e de marcação PL no elemento pré-nominal no PA, com recurso a morfema do tipo *singleton*, determina que o radical das palavras se mantenha inalterado nos restantes elementos que compõem a cadeia sintagmática. No PA a esmagadora maioria dos adjetivos surge também preferencialmente na posição pós-nominal, conforme se verá no ponto 4.1.2.1.1 do

presente trabalho. Em SN's de estrutura reduzida (2 ou 3 elementos), que configuram a grande maioria das produções sintagmáticas dos almozarifanos, estabelece-se, então, um claro paralelo entre determinantes e primeira posição, substantivos e segunda posição e adjetivos e terceira posição (cf. Guy, 1981b).

A não-marcação de PL sufixal nos elementos do SN do santomense implica que determinados itens gramaticais, como os possessivos, também se mantenham inalterados. Esta categoria gramatical ocorre também em posição pós-nominal (exemplos [53a] e [53b]), um aspecto que encontra paralelo nos substratos níger-congo atlânticos (exemplo [78]) e noutros CP's atlânticos (exemplo [79]):

[78] IORUBÁ:  
*omo mi*<sup>38</sup>  
 filho 1SG  
 “o meu filho”

[79] CCV:  
*na ja bráb di mi* (Holm, 1988/1989:274)  
 na ilha Brava de 1SG  
 “na minha ilha Brava”

A construção é também detectada no PA (exemplo [74]) e outras variedades africanas de português (exemplo [80]), originando frases agramaticais no que concerne à posição linear do possessivo:

[80] PVA:  
*eu tenho horário meu* (Inverno, 2005:140)  
 “eu tenho o meu horário”

Idêntica situação acontece com a pluralização e colocação dos determinantes demonstrativos, que apenas possui uma forma no santomense (*se*) e se situa sempre em posição pós-nominal (exemplos [67] e [68]). Lucchesi (1993) atribui a *se* a função de definido, servindo para enfatizar o nome. Como tal, não será um elemento nuclear no santomense. Todavia, esta não é a visão de Alexandre & Hagemeyer (2007:44), que entendem que *se* não detém estas características no santomense. De facto, não só o traço definido é preenchido pelos SNN's como *se* pode ainda co-ocorrer com os marcadores definido *ũa* e indefinido *inen* (Alexandre & Hagemeyer, 2007:44), especialmente com o segundo (exemplo [81]). Assim sendo, os autores propõem que o elemento gramatical *se*

seja observado como um especificador de marcação (MSPEC), que é núcleo do SSPEC (Fig. 5).

- [81] FORRO:  
*So inen ja se so sun Fenanji na ska*  
 Então 3PL *dia* MSPECFOC senhor Fernandes NEG ASP  
*texila vin fa.* (Alexandre & Hagemeijer, 2007:41)  
 Tira vinho NEG  
 “Então, estes dias o senhor Fernandes não está a tirar vinho de palma.”

Uma das características estruturais rígidas de *se* é a de apenas poder ocorrer de modo adjectival, isto é, requerer obrigatoriamente a presença do núcleo nominal (exemplo [82]), o que não sucede com o sistema do português, reforçando o facto de estarmos na presença de um MSPEC.

- [82] FORRO:  
*Ka se sa bwa* (Ferraz, 1979:74)  
 Coisa MSPECestá boa  
 “isto/aquilo é bom”

A relação entre núcleo e MSPEC leva Alexandre & Hagemeijer (2007:45) a considerarem *se* um clítico com forte adjacência nominal visível, enquanto o marcador de PL *inen* e o quantificador *ũa* provêm do léxico com um traço de número inerente, comportando-se nominalmente e permitindo lacunas (nomes nus). De facto, a definitude, enquanto instanciação formal que permite a identificação, conecta a frase ao nível do discurso e pode ser desenvolvida antes do número (Wood, 2003:127), enquanto este transita directamente a partir do seu próprio traço, quando os determinantes são visíveis. A ligação entre definitude e número no santomense activa a leitura do número singular, sendo necessário recorrer ao marcador de PL *inen* para transformar esse singular em PL. Por sua vez, o MSPEC tem a função de reforçar o traço de definitude.

Por outro lado, o alto grau de referencialidade que resulta da co-ocorrência entre o marcador de definitude e o MSPEC é bastante comum com nomes de traço [-humano] (exemplo [81]), estabelecendo o contraste com as ocorrências em que apenas intervêm o marcador de definitude e o nome, aplicadas aos nomes de traço [+humano] (exemplo [38b]). A análise das questões inerentes ao SDET incidindo na sintaxe e semântica dos modificadores que expressam definitude, número, especificidade e interacção entre

todos terá, então, que pressupor que o traço de número está directamente associado à definitude, embora esta possa cair fora do nível gramatical quando se trata de SNN's. Considerando que o traço de número pode ser mapeado a nível do DET0 (Longobardi, 1994:621) e que não ocorre marcação PL por intermédio de afixos no santomense, Alexandre & Hagemeijer (2007:43) entendem que não é requerida uma categoria funcional separada para os SNN's, propondo uma estrutura sintáctica que dispensa uma projecção independente de número e integra ambos os DET's, lexicalizados e ausentes, num único sistema. Desta forma, para os exemplos [83a] e [83b], ter-se-á a estrutura do SDET representada na Fig. 5, a qual, em virtude da sua especificidade, difere das propostas por Castro & Pratas (2006) para os dois tipos de SDET existentes no CCV (Figs. 2 e 3):

- [83] FORRO:
- a. *ũa mina se* (Alexandre & Hagemeijer, 2007:47)  
 IND menina MSPEC  
 “uma menina em questão”
- b. *inen mina se* (Alexandre & Hagemeijer, 2007:47)  
 3PL menina MSPEC  
 “as meninas em questão”

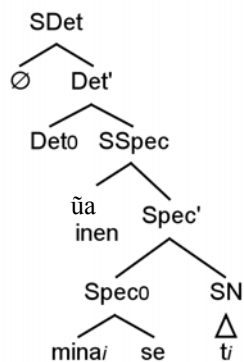


Fig. 5. Estrutura sintáctica do SDET no santomense (marcação de número)  
 (Alexandre & Hagemeijer, 2007:47).

De acordo com a configuração da Fig. 5, o MSPEC *se* é núcleo do SSPEC, um nó que apenas projecta quando ocorre evidência lexical. O nome é adjunto à esquerda do clítico *se*, que é invariável e detém leitura de singular associada a si. Bernstein (1993:217) entende que os demonstrativos e os determinantes definidos são núcleo do SDET, daí excluïrem-se mutuamente (p.e. *\*o este carro / \*este o carro*). Este pressuposto torna

feliz o entendimento de que *ũa* e *inen* são núcleos do SDET no santomense. Enquanto determinantes, ambos C-comandam o seu complemento, SN e SSPEC, se projectados. Como mais nenhum movimento do nome é necessário, a árvore sintáctica proposta configura a generalidade da estrutura dos SDET's no santomense.

Relativamente ao uso do possessivo, obedece ao mesmo tipo de colocação verificada para o MSPEC (pós-nominal), recorrendo os falantes ao pronome pessoal objecto para realizarem tal categoria (exemplo [84]), uma situação que, ao que tudo indica, se aplica ao núcleo do SDET *inen*, homónimo do pronome pessoal, 3ª pessoa PL:

[84] FORRO:  
*Livlu mu* (Ferraz, 1979:81)  
 livro PP/POS  
 “o meu livro”

Outra questão que poderá ter interferência na observação da CPL-var do SN do PA relaciona-se com o uso das partículas locativas que marcam obliquamente o nome e, por norma, não marcam presença no santomense. Tal aspecto reflecte o reduzido sistema preposicional dos crioulos:

[85] FORRO:  
*Amaña nga ba vende* (Ferraz, 1979:74)  
 Amanhã 1SG-FUT ir loja  
 “Amanhã eu vou (irei) à loja.”

Se atentarmos às línguas do grupo níger-congo atlântico, verificamos que o seu sistema preposicional é mais reduzido que o do português. No kimbundu, por exemplo, verifica-se que os complementos circunstanciais de lugar são introduzidos por apenas três preposições:

- (i) *Ku*, locativo referente a direcção, lugar de destino distante, interioridade e proveniência. Estabelece correspondência com as preposições *a*, *em*, *para* e *de*.
- (ii) *Mu*, com o sentido de interioridade. Corresponde à preposição *em* e às locuções prepositivas *dentro de*, *no interior de*, etc.
- (iii) *Bu*, que indica sobreposição, à superfície de. Condiz com a preposição *sobre* e as locuções prepositivas *em cima de*, *acima de*, etc.

As preposições podem ser pré-afixadas aos nomes ou pronomes, formando então SN's com valor espacial:

- [86] KIMBUNDU:  
*mwene wamuya kubata* (Mingas, 2000:75)  
 “mwene u~amuya ku~(di)bata”  
 3SG 3SG+PROG+ir LOC+casa  
 “ele vai para casa”
- [87] KIMBUNDU:  
*mwene wala monzo* (Mingas, 2000:76)  
 “mwene u~ala mu~inzu”  
 3SG 3SG+IMP+estar LOC+quarto  
 “ele está no quarto”
- [88] KIMBUNDU:  
*Mwene wambata anyi bu mutwe?* (Mingas, 2000:76)  
 “mwene u~ambata anyi bu mutwe?”  
 3SG 3SG+IMP+levar o que é LOC cabeça  
 “o que é que ele leva na cabeça?”

O PE, por seu lado, possui um sistema de preposições mais extenso para expressar as funções espaciais (*a, ante, após, até, contra, de, desde, em, entre, para, perante, por, sob, sobre, trás*), podendo algumas destas preposições surgir contraídas com os definidos, os indefinidos (artigos e pronomes), os pronomes pessoais e os pronomes demonstrativos. Deste modo, os falantes de L1's bantu em situação de aquisição do PtgL2 não conseguem estabelecer as funções fornecidas pelas diversas preposições do PE e acabam por produzir ocorrências que se revelam agramaticais neste (Mingas, 2000:76). A este propósito, refira-se que a preposição “em”, correspondente à forma *ku*, é a mais produtiva e, conseqüentemente, aquela que mais desvios revela:

- [89] PANG:  
*Pessoas que estavam lá na trás...* (Cardoso, 1980:56)  
 “As pessoas que estavam lá atrás”
- [90] PANG:  
*Passavam e olhavam nele...* Cardoso, 1980:35)  
 “Passavam e olhavam para ele...”
- [91] PANG:  
*farristas vieram na procura dos bens deixados* (Cardoso, 1980:55)  
 “os farristas vieram à procura dos bens deixados”

[92] PANG:  
*Pra quê queres saber? É **na** tua conta?* (Cardoso, 1980:58)  
“Para que é que queres saber? É da tua conta?”

[93] PANG:  
*se queres gozar, vai gozar **na** tua mãe.* (Cardoso, 1980:60)  
“se quiseres gozar, vai gozar com a tua mãe.”

No PA, a não identificação das funções das preposições do PE é também uma evidência, a ponto de elas poderem ser adoptadas para exprimir situações que se demarcam da referência espacial (exemplos [95], [96] e [97]) ou serem aplicadas de forma redundante (exemplo [97]):

[94] PA:  
*daqui é fui **para até** hospital* [OSVALH1]  
“daqui, eu fui ao (para o) hospital”

[95] PA:  
*o quê que ocê faz de manhã até **na** hora que ocê vai dormir **de** noite?*  
[OSVALH1]  
“o que é que você faz desde a manhã até à hora em que você vai dormir à noite”

[96] PA:  
*o irmão também que stava **na** xofer na Praia Conchas* [CELESH3]  
“o irmão que também era chofer/motorista na Praia das Conchas”

[97] PA:  
*agora quase toda criança naci já **com a** falá português* [CELESH3]  
“agora quase todas as crianças nascem já a falar português”

A indeterminação das funções e uso das preposições do PE é também comum nas outras variedades africanas de português:

[98] PM:  
*Os alunos, aqueles que ser mandado **naquelas** escolas profissionais.*  
(Gonçalves & Stroud, 1998:127)  
“Os alunos, aqueles que são mandados para aquelas escolas profissionais.”

[99] PM:  
*... para os portugueses ir também **no** país deles.*  
(Gonçalves & Stroud, 1998:127)  
“... para os portugueses irem também para o país deles.”



- [100] PM:  
*Tem que viver com comida que ser oferecido **com** outros países.*  
 (Gonçalves & Stroud, 1998:127)  
 “Têm que viver com a comida que é oferecida pelos outros países.”
- [101] PM:  
*Eu é que é mais velho em relação **dos** outros.*  
 (Gonçalves & Stroud, 1998:127)  
 “Eu é que sou o mais velho em relação aos outros.”
- [102] PT:  
*põe dentro de garrafa, vai jogar **no** mar* [Amélia do Nascimento – M1]  
 “põe-se dentro da garrafa e vai jogar-se (atirar-se) ao mar”
- [103] PT:  
*Pa entregá **nas** crianças* [Joana Nunes – M1]  
 “Para entregar às crianças”
- [104] PT:  
*porque avó falô com ele só **com** kimbundu* [Amélia do Nascimento – M1]  
 “porque os avós falaram com ele só em kimbundu”

Abordando a concordância variável de género no SN de HEL-Ba, Lucchesi (2000a:98) chama a atenção para as estruturas desviantes deste, nomeadamente no que concerne à “*redução/eliminação do emprego de nexos gramaticais*” que levam este dialecto a apresentar não só “*o uso de em por a e de de por para/a*” (exemplos [105] e [106]) mas também ausência de preposições em determinadas realizações que requerem o seu uso (exemplos [107] e [108]):

- [105] HEL-BA:  
*fui lá **em** procura de um remédio* (Lucchesi, 2000a:98)  
 “fui lá à procura de um remédio”
- [106] HEL-BA:  
*fica **de** sempre* (Lucchesi, 2000a:98)  
 “fica para sempre”
- [107] HEL-BA:  
*porque eu nunca gostei **nenhuma*** (Lucchesi, 2000a:98)  
 “porque eu nunca gostei de nenhuma”
- [108] HEL-BA:  
*mas nós nunca liguemo festa não* (Lucchesi, 2000a:98)  
 “mas nós nunca ligámos à festa, não”

Estas estruturas apresentam “*um paralelo com crioulos de base portuguesa da África (a exemplo do crioulo de S. Tomé)*” (Lucchesi, 2000a:97). Efectivamente, o santomense caracteriza-se por possuir um reduzido sistema de preposições, ocorrendo frequentemente estruturas que apresentam ausência das mesmas (exemplo [85]). Sendo a redução no uso das preposições um dos fenómenos de não-realização gramatical que caracteriza também os processos de mudança resultantes da transmissão linguística irregular, a ausência destes itens gramaticais na estrutura das frases do santomense (exemplo [85]) e do PVB (exemplos [107] e [108]) estabelece também paralelismos não só com o PA mas também com outras variedades africanas de português:

[109] PA:  
*É, bêra strada.* [CLOTIM2]  
 “É, à beira da estrada.”

[110] PANG:  
*Sim, trabalhar aí sês casa* (Inverno, 2005:140)  
 “Sim, trabalho aí na vossa casa”

[111] PM:  
*Pega cadeira!* (Gonçalves & Stroud, 1998:41)  
 “Pega na/numa cadeira”

[112] PT:  
*a missora começô transmitir* [Pedro Fuíca – H3]  
 “a emissora começou a transmitir”

Ainda no que concerne ao emprego das preposições, Lucchesi (2000a:97) menciona igualmente o uso da forma “ni” por “em” (exemplos [113] e [114]), que se revela quase categórico na fala de Mato Grosso, Bahia. Segundo o autor, dado tratar-se de um item vocabular isolado, “*torna-se ainda mais aceitável a hipótese de um empréstimo lexical, com origem em outros dialectos rurais brasileiros que desenvolveram a forma a partir do contacto entre línguas*” (Lucchesi, 2000a:97-98).

[113] HEL-BA:  
*nunca trabalhei **ni** roça* (Lucchesi, 2000a:97)

[114] HEL-BA:  
*eu já fui **ni** Rio de Contas* (Lucchesi, 2000a:97)

O mesmo locativo regista-se também no santomense, que pode, opcionalmente, recorrer a três tipos de prefixos nominais, com função locativa:

(i) O morfema *ni*, ou sua variante *n-*, antes de vogais:

- [115] FORRO:  
 a. ***ni*** *ke* (Ferraz, 1979:74)  
 em casa  
 “em/na casa”  
 b. ***n-*** *ome* (Ferraz, 1979:74)  
 em meio  
 “no meio”

(ii) A consoante hormogânica nasal *m*<sup>39</sup>:

- [116] FORRO:  
***m-*** *panela* (Ferraz, 1979:75)  
 em panela  
 “na panela”

(iii) A vogal morfofonémica *î*<sup>40</sup>:

- [117] FORRO:  
***î-*** *liba* (Ferraz, 1979:75)  
 em cima  
 “em cima”

Idêntico tipo de realização ocorre em línguas do grupo níger-congo atlântico, como o iorubá (exemplo [118]) ou o kimbundu (exemplo [119]), substituindo o advérbio de lugar que marca obliquamente o nome:

- [118] IORUBÁ:  
***Ni*** *ile* > ***nile*** (Ferraz, 1979:75)  
 PREP casa  
 “em/na casa”

- [119] KIMBUNDU:  
*Ngala* ***ninzala*** (Zau, 2001:28)  
 1SG+estar PREP+fome  
 “eu estou com fome/eu tenho fome”

O mesmo item gramatical marca também presença tanto no PA, que tem um CP atlântico como substrato, como no PT, que tem línguas do grupo níger-congo como substratos:

- [120] PA:  
*porque eu stive lá ni empresa trabalhar* [MANOH3]  
 “porque eu estive lá na empresa a trabalhar”
- [121] PA:  
*com chicote ni costas pa vi pa o pé trabalho* [MANOH3]  
 “com chicote nas costas para vir a pé para o trabalho”
- [122] PT:  
*quando levantei encontrei arroz ni dezoito dobra.*<sup>41</sup>  
 [Amélia do Nascimento – M1]  
 “quando me levantei encontrei o arroz a dezoito dobras”
- [123] PT:  
*eu tenho um tio lá ni sede também* [Amélia do Nascimento – M1]  
 “eu também tenho um tio lá na sede”

Baxter (2004:109) refere que o locativo invariável “ni” substitui, no PT, componentes flexionados do português como <preposição+artigo+número>, representados pelas contracções prepositivas espaciais “nos/nas”, com sentido locativo ou direccional. A forma “ni” ocorre com SN’s sem artigo definido do português, pelo que a dupla função locativo-direccional atribuída pelos falantes do PT a este locativo se ficará a dever a fenómenos de transferência resultantes de, nas línguas do substrato africano, as preposições compartilharem as referidas funções locativo-direccionais. Entretanto, não temos conhecimento da existência deste tipo de realização nem no PANG nem no PM, podendo a mesma ter origem no santomense, por transferência, daí transitando geracionalmente, via DLP’s, para o português de São Tomé e suas variantes, como o PA. Posteriormente, devido ao contacto linguístico entre os falantes destes e os utilizadores do PT, a mesma ter-se-á generalizado na ilha, comprovando-se “*a enorme complexidade da questão do contacto entre línguas em geral*” (Lucchesi, 2000a:98). Baxter (2004:100) advoga também que os valores de “ni” no PT, embora sendo entendidos como conectados ao crioulo de S. Tomé, não terão sido adquiridos pelos tongas via santomense nos estádios primordiais da actividade laboral nas roças, mas sim posteriormente, via português de São Tomé (e PA). Apesar de a construção locativa “ni” não ocorrer no PANG e no PM, note-se a sua aproximação fonológica às forma “no/na”, bastante generalizadas nestas variedades de português (especialmente a forma “na”). Ainda segundo Baxter (2004:123), tanto a dupla função locativo-direccional da

preposição portuguesa “em” como o facto de a consoante /d/ inicial ocorrer como /nd/ em línguas bantu, como o umbundu, terão igualmente concorrido para que a preposição “de” [di] passasse a ser pronunciada como /ni/.

Em HEL-Ba há evidências de que o locativo actua como demonstrativo, que pode funcionar como artigo (Baxter, Lucchesi & Guimarães, 1997:13-14). Segundo Cerqueira (1993), os locativos funcionam como demonstrativos na fase inicial da aquisição do PBL1, pelo que os dados do PA poderão reflectir uma fase transitória em que o artigo se desenvolveu “*em estruturas que previamente dependiam do locativo pós-nominal como marcador das propriedades referenciais do substantivo. Nessa fase transicional, esse papel pretérito do locativo começa a recuar enquanto se desenvolve a estrutura pré-nominal*” (Baxter & Lopes, 2009:326-327). Estes aspectos têm repercussão directa na estrutura do SN do PA, já que as formas “ni” e “di” impedem que se incluam, nos dados a serem quantificados e codificados na presente análise, SN’s que no PE configuram a cadeia sintagmática PREP+DET+N, mas que na fala de Almojarife interpretámos com a sequência PREP+N, isto é, SNN’s. Como tal, este tipo de SN’s não foi codificado e quantificado por nós, em virtude de representar estruturas em que acontece o apagamento da categoria DET (exemplos [120] e [121]). Opostamente, o uso desviado de algumas preposições do PE, com especial incidência para a forma contraída “na”, não só pode determinar concordância variável de género entre o núcleo do SN e o DET como originar também SN’s simples no PA (PREP+DET+N), codificados e quantificados na presente análise. Note-se que este tipo de SN’s poderá ser inexistente no PE, caso configure estruturas do tipo PREP+N ou com nome simples (exemplo [96]).

Resta referir que, embora o aparecimento da forma “ni” no dialecto de Mato Grosso possa também ser equacionado como um empréstimo lexical proveniente de outras variedades rurais brasileiras, a verdade é que não pode ser descartada a possibilidade de o seu surgimento nestas ter a ver com uma transplantação da responsabilidade dos escravos provenientes de S. Tomé, após terem aqui permanecido durante algum tempo, prestando trabalho nos engenhos de açúcar (Holm, 1988/1989, 1992; Baxter, 1992, 1997, 2003).

### **1.2.2.3. A comunidade de Almojarife**

Após a reconquista de São Tomé aos holandeses, em 1644, os portugueses deparam-se com uma realidade diferente da anterior, reflectida na “africanização” do arquipélago. A ilha encontrava-se praticamente dividida em duas partes (Henriques, 1989:276), criando-se uma situação de parco contacto linguístico entre os membros estabelecidos nas regiões interiores (escravos africanos e seus descendentes, a maioria falante do santomense e alguns bilingues em forro e PtgL2 ou PtgL1 nativizado) e litorais (colonizadores, falantes do PE, e seus descendentes libertos, bilingues em santomense e PtgL1). Ganhando autonomia ao refugiarem-se no mato, os africanos deixavam de estar inseridos no espaço linguístico-cultural dos colonos. Paralelamente, a conjuntura económica não favorecia a exploração açucareira (Henriques, 1989:278), evitando nova instalação massiva de colonos no arquipélago e acentuando a situação de afastamento linguístico entre estes e a população sediada nas zonas rurais.

Face a este panorama, a melhoria económica só seria retomada no século XIX, com a introdução das culturas do café (1800) e do cacau (1822) nas ilhas. No entanto, vivendo em situação de auto-suficiência no interior das ilhas, a maioria dos libertos ou descendentes de africanos sedeados em São Tomé não se mostrava predisposta a trabalhar para os colonos, fazendo com que os proprietários das roças tivessem de recorrer à mão-de-obra “contratada”, provinda inicialmente de Angola e Moçambique (finais do século XIX), mas também de Cabo Verde, em período posterior (meados do século XX). Apesar de a abolição da escravatura ter acontecido no império português a 25 de Fevereiro de 1869, os trabalhadores das roças eram forçados a trabalhar em condições desumanas. Esta situação levou a que a maioria dos libertos continuasse a não procurar emprego nas roças (Ribeiro, 1875:12), pelo que, nestas, apenas permaneciam os “contratados”, constituindo comunidades cada vez mais isoladas. Por seu lado, os libertos, tal como sucedeu no Brasil, buscavam outras alternativas de sobrevivência, estabelecendo-se a sua grande maioria na periferia dos centros urbanos. Outros, porém, permaneceram em pequenas comunidades rurais semi-isoladas, compostas por núcleos familiares, como terá acontecido com os forros que se instalaram em Almojarife.

Ainda que a falta de documentação sobre esta pequena comunidade não permita determinar com exactidão a sua origem, o facto de os seus falantes serem bilingues em

santomense e português reestruturado indicia que ela não terá emergido aquando da desactivação dos engenhos do açúcar e retirada dos colonos após tomada do arquipélago pelos holandeses, no século XVII, e que determinou também a fixação de vários grupos comunitários no interior da ilha, recorrendo, essencialmente, à agricultura de sobrevivência. Assim, o grupo de forros que se estabeleceu na ampla Baía de Almoхарife seria constituído por libertos santomenses que, em movimentos deambulatórios iniciados em finais do século XIX/princípios do século XX, buscavam pontos de fixação afastados das roças, a fim de não serem compulsivamente integrados na força braçal destas. Uma vez instalada na baía, a comunidade procurou o seu sustento na agricultura de pequena escala e na pesca artesanal, uma actividade até então exercida quase exclusivamente pelos angolares (Esquadrão de Polícia Militar N° 2222, 1962:1-2).

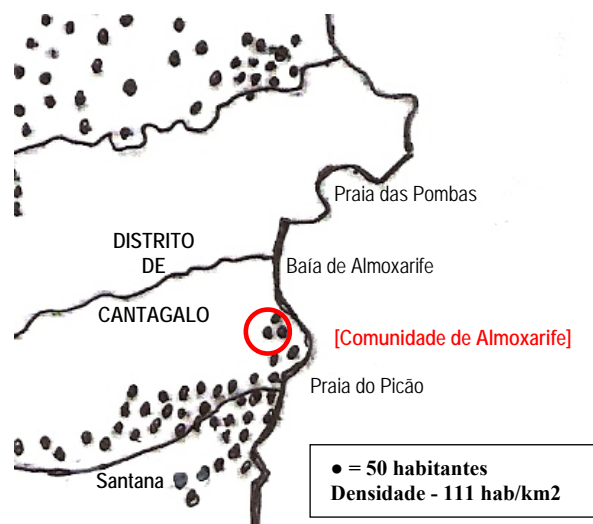


Fig. 6. Índice demográfico de Santana e Almoхарife (Tenreiro, 1961:164).

Situada na zona Este da ilha de São Tomé, caracterizada como litoral baixo, a Baía de Almoхарife é confinada a norte pela Praia das Pombas e a sul pela Praia do Picão. Pertence ao Distrito de Cantagalo, que tem por capital a povoação de Santana. De acordo com o Censo de 1960 (Tenreiro, 1961:163), o distrito albergava cerca de 13258 habitantes, espalhados por uma área de 119 km<sup>2</sup>. A densidade populacional rondava os 111 habitantes por km<sup>2</sup> (Fig. 6). Quanto à capital, possuía, à data, 7702 habitantes (4353 do sexo masculino e 3349 do sexo feminino), enquanto a pequena comunidade de

Almoxarife era constituída por cerca de 150 forros, que se concentraram na parte sul da baía (Tenreiro, 1961:164).

A situação de afastamento linguístico relativamente ao PE acentuou-se para os almoxarifanos, ajudando a fixar as características linguísticas defectivas na língua nativizada, em virtude de os DLP's transmitidos geracionalmente preservarem a gramática dos falantes das gerações mais velhas, marcada por acentuada não-aquisição de morfologia relativamente à LA. O isolamento da comunidade e conseqüente pouco contacto com os falantes do PE prolongaram-se ainda por bastante tempo, fosse pelo receio dos membros da comunidade em serem obrigados a regressar às roças fosse pela política do Estado português, que somente em 1926 aprovaria, por decreto emitido pelo Estado Novo, o *Estatuto Orgânico das Missões Católicas Portuguesas de África e Timor*. Este, por seu lado, atribuía à Igreja Católica o papel exclusivo na evangelização e “civilização” das populações indígenas não assimiladas dos territórios coloniais portugueses. Quanto à rádio, que poderia permitir às comunidades isoladas um contacto mais próximo com o PE, seria introduzido em São Tomé em 1925, mas somente em 1949 é que foi fundada a Rádio Clube de São Tomé, com uma maior capacidade de alcance e que se encontra ainda em funcionamento.<sup>42</sup> Relativamente à televisão, só seria introduzida na ilha alguns anos depois da independência, já em 1982. Contudo, privados de rede eléctrica em Almoxarife, os membros da comunidade apenas tinham acesso a ela caso se deslocassem a Santana, uma situação que se mantém até aos dias de hoje.

No final da década de 60 do século XX (apenas sete anos antes da independência) surgem as primeiras referências sobre a subida do nível de educação da população em relatórios oficiais, que não é acompanhado pelas oportunidades de emprego (CTI de São Tomé, 1968a:1) e faz com que alguns forros das novas gerações voltem a procurar trabalho nas roças (CTI de São Tomé, 1968b:1-2). Como se verifica, apesar de o ensino e o exército português, que integra elementos nativos, permitirem às gerações mais novas um contacto estreito com o PE, facultando ao PtgL1 nativizado vectores de expansão estrutural em direcção à LA e minorando os processos de transferência, o contacto com o português dos “contratados”, marcado por características das línguas bantu, não cessa para os falantes forros.

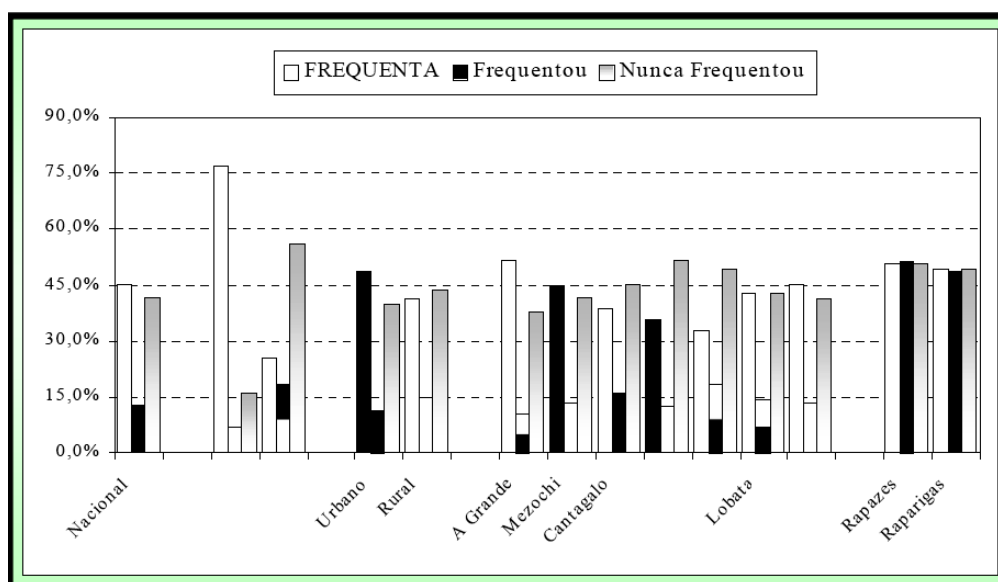


Nos anos 70 do século XX, o preço do cacau cai nos mercados internacionais e a economia de plantação da ilha entra em recessão, acontecendo o êxodo de muitos cabo-verdianos, que regressam a casa ou partem para a metrópole (CTI de São Tomé, 1970:1, 1972a:1, 1972b:1). Actualmente, e no que diz respeito aos estrangeiros residentes no Distrito de Cantagalo, representam cerca de 10% da população deste, sendo também o grupo cabo-verdiano aquele que continua a deter o maior efectivo de residentes estrangeiros no arquipélago (57%) (Instituto Nacional de Estatística de São Tomé, 2003a:24-26). Na década de 70, a ilha possui já uma eficiente cobertura escolar a nível do território, coadjuvada por “... *uma boa rede de estradas em que a pavimentação de asfalto ultrapassou já os 220 Km...*” (Unidade da Companhia de Artilharia 3376, 1973a:3), bem como a existência de “... *boas rodovias que normalmente servem as sedes das propriedades agrícolas*” (Unidade da Companhia de Artilharia 3376, 1973a:3), aspecto que terá contribuído para que, de certa forma, o português reestruturado falado pelos núcleos urbanos e pelas comunidades rurais mais isoladas se nivelasse, eliminando-se bipolarizações entre a fala de ambas, sobretudo a nível das gerações mais novas e mais escolarizadas da ilha. Nesta fase, o contacto mais estreito com o PE por parte dos falantes bilingues de São Tomé, em geral, e da comunidade de Almojarife, em particular, vai permitir a incrementação de empréstimos de traços lexicais e estruturais da LA (cf. Thomason & Kaufman, 1988; Lefebvre, 2003; Montrul, 2006; Lefebvre, White & Jourdan, 2006b; Siegel, 2008), colocando gramáticas em competição e originando a CPL-var no SN produzido, actualmente, pelos almojarifanos.

Em 12 de Julho de 1975, com a independência de São Tomé e Príncipe, fechava-se um capítulo de meio século de vida do arquipélago como colónia, marcado por períodos de auge e decadência económica, com consequências profundas na construção da identificação sociolinguística de um povo descendente de antigos escravos, transplantados e desenraizados geográfica e culturalmente das suas origens. Com a independência, foi instaurado um regime monopartidário, que procedeu à nacionalização das roças, levando cerca de 4000 portugueses a abandonarem o arquipélago. O período que se segue à independência é marcado por grande conflitualidade interna, sucedendo-se as tentativas de golpes de estado. Com as questões socioeconómicas jogadas ao esquecimento, as estruturas do país foram-se degradando, sendo as comunidades rurais

votadas ao abandono e tornando-se cada vez mais isoladas, face à desarticulação dos sistemas educacional, rodoviário e de comunicação.

No que concerne à questão da educação no Distrito de Cantagalo, presentemente cerca de  $\frac{1}{4}$  da sua população jovem encontra-se em fase de idade escolar obrigatória (6-12 anos),<sup>43</sup> não existindo praticamente assimetria entre ambos os sexos. Contudo, os valores que se referem à frequência escolar das crianças com idade compreendida entre os 5 e os 17 anos evidenciam que o número de analfabetos do Distrito de Cantagalo (46%), nesta faixa etária, é superior ao dos que frequentaram (15%) ou frequentam (39%) a escola (Fig. 7). Este aspecto, não só revela as dificuldades que as crianças dos meios rurais isolados ou semi-isolados, nos quais se inclui a comunidade de Almojarife, sentem para conseguir frequentar presentemente a escola como tem também evidentes reflexos na questão da estabilização da variação registada a nível da concordância do SN, visto o elevado número de analfabetos não ter contacto com as normas do PE, que poderiam ajudar na aquisição/expansão da regra de concordância.

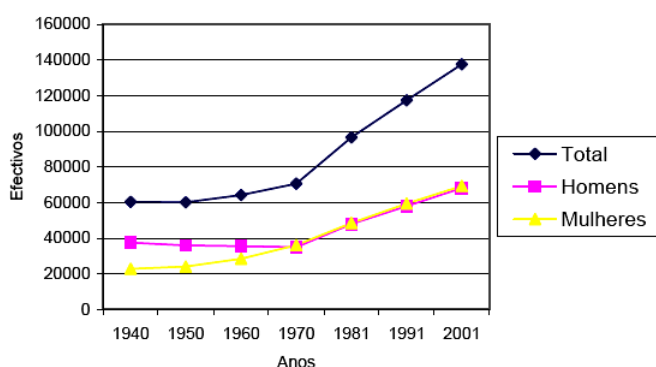


**Fig. 7.** Percentagem das crianças com 5-17 anos que têm acesso à escolaridade. (Fonte: Instituto Nacional de Estatística de São Tomé e Príncipe, 2003b:23).

A desarticulação do sistema educacional, traduzido no corte do contacto entre o português popular e o PE, tem como consequência, segundo o Censo de 2001, o facto de 98,9 % da população actual do arquipélago entender e usar o português, mas apenas a classe política e a alta sociedade, ou seja, aquelas que fazem grande parte dos seus

estudos em Portugal, utilizarem a variedade padrão (Instituto Nacional de Estatística de São Tomé e Príncipe, 2003/2005).

Dados do Censo de 1970 (Unidade da Companhia de Artilharia 3376, 1973b) mostravam que a população de Santana, contrariando a tendência geral de crescimento populacional verificada no país, decrescera ligeiramente (7496 indivíduos), levando a um ajustamento na distribuição entre sexos (3794 homens e 3702 mulheres), que se mantém até à actualidade. Este aspecto indicia que os membros mais jovens abandonam a região e não regressam, fazendo com que a fala almoxarifana continue a conservar as suas características mais crioulizantes. Quanto ao esbatimento da diferença que existia no efectivo de homens e mulheres, estará relacionado, sobretudo, com o facto de, na região, serem os primeiros quem mais migra para os centros urbanos<sup>44</sup> ou emigra para o estrangeiro. No entanto, a nível geral do país, é possível constatar que, entre 1940 e 1970, a população masculina foi sempre superior à feminina, tendo o desequilíbrio origem na população estrangeira levada para São Tomé sob contrato forçado para trabalhar nas roças, nas quais predominava a mão-de-obra masculina. Contudo, finalizada a entrada de “contratados” e aumentando a emigração dos homens em busca de melhores condições de trabalho e vida no estrangeiro, inverteu-se a situação, com o número de habitantes do sexo feminino<sup>45</sup> ultrapassando o do sexo masculino (Fig. 8):



**Fig. 8.** Evolução da população de São Tomé e Príncipe – 1940 a 2001.  
(Fonte: Instituto Nacional de Estatística de São Tomé e Príncipe, 2003a:9).

### 1.2.2.3.1. Características da fala de Almoxarife

A situação que determinou o aparecimento da comunidade de Almoxarife tem, obviamente, implicações directas no português usado por esta. A economia da comunidade nunca teve qualquer relevância para o país, já que os seus membros se

debatem com as dificuldades resultantes do semi-isolamento e dos seus poucos recursos de subsistência, provenientes da pesca e da agricultura (milho, batata-doce, mandioca ou quiabo) de mera subsistência. Ainda assim, a comunidade, principalmente por intermédio das mulheres, transacciona uma escassa parte destes recursos com os poucos comerciantes locais ou em mercados populares de reduzida escala. Portanto, a fala de Almojarife, ainda que de forma pouco significativa, não deixa de estar submetida a um *continuum* linguístico, aspecto que, por exemplo, não sucede tão frequentemente com as comunidades mais isoladas de algumas roças.

O acesso entre Almojarife e Santana é feito através de uma estrada de terra batida, mas nunca existiu uma rede de transportes que permitisse aos membros da comunidade utilizarem-na para se deslocarem até à capital do distrito. Como o percurso tinha que ser feito a pé, na época colonial eram os próprios comerciantes que, utilizando as suas viaturas, se deslocavam à baía para comerciarem com os almojarifanos. Este aspecto terá contribuído bastante para o semi-isolamento da comunidade antes da independência do país. Não obstante, vencido o percurso entre Almojarife e Santana, era possível aos membros da comunidade deslocarem-se com facilidade para outros pontos da ilha, já que, estando incluída no leque das vilas rurais mais importantes da colónia, com a categoria de povoação administrativa, Santana era servida por uma estrada com pavimento em asfalto e que a ligava a São Tomé, Angolares e Caué (Unidade da Companhia de Artilharia 3376, 1873b:6). Contudo, após a independência, o isolamento da comunidade acentuou-se consideravelmente, visto que os colonos abandonaram a ilha, as rodovias se degradaram, a ponto de se tornarem intransitáveis, e se extinguiu a rede de transportes públicos que permitia aos seus usuários deslocarem-se regularmente para outros locais. Disso mesmo nos dá conta o excerto do diálogo entre o entrevistador e o informante [OSVALH1], que deixa transparecer a inexistência de contactos entre os membros da comunidade e os habitantes de Santana:

DOC: *Você custuma i a Santana?*

INF: *Vô umas vezes devido quê zê essa prigiça de i a pé e volta a pé.*

DOC: *Hum, mas você vai assim em Santana mais pra fazê o quê?*

INF: *Quê zê é vô mais a Santana assisti programa.*

DOC: *Hum.*

INF: *Só!*

Como se verifica, o *continuum* linguístico da comunidade almoxarifana tinha como consequência, até um passado bastante recente, a manutenção da situação de afastamento linguístico relativamente ao PE, contribuindo para conservar as características defectivas do PtgL2. Os falantes da faixa etária mais idosa (+65 anos) afirmaram todos que tiveram o santomense como L1, embora alguns deles tenham também aprendido o PtgL1 através do contacto com o PtgL2/PtgL1 dos pais, impregnado de características crioulizantes.

No que concerne à geração intermédia de Almojarife (41-65 anos), apenas a informante [MANEVEM2] (Tabela 8.1) aprendeu o português e o santomense simultaneamente, já que todos os outros falantes tiveram o segundo como L1. Alguns dos falantes desta faixa etária trabalharam sazonalmente nas roças, fazendo com que o PA entrasse em contacto com o português adstrato dos “contratados”, marcado por características das línguas bantu. Este aspecto terá também contribuído para que a situação de pouco contacto com o PE se mantivesse, já que a comunicação entre os capatazes e trabalhadores era feita, muitas vezes, com os primeiros recorrendo a formas precárias do português.

O *continuum* linguístico da comunidade almoxarifana teve como consequência, até um passado bastante recente, a manutenção de características defectivas em relação à LA. Estas constituíram, posteriormente, os modelos disponíveis para o processo de nativização do PAL1, adquirido em simultâneo com o santomense pela geração mais nova da comunidade (20-40 anos), mas com o primeiro revelando-se como a língua dominante e o segundo sendo estigmatizado no contexto de aquisição bilingue:

DOC: *Qual qual foi a língua que você aprendeu primeiro?*

INF: *Português.*

DOC: *Não aprendeu dialecto em casa não?*

INF: *Ê també aprendi. [Põe lá pra mim, Sã!]*

DOC: *Sua vó falava em casa com ocê como?*

INF: *Falava dialecto também.*

DOC: *Falava mais dialecto ou português em casa, sua avó?*

INF: *Hum, mas assim quan gen fala dialecto assim senhora não gostava.*

DOC: *Ahã tinha que falar português?*

INF: *Hum.*

DOC: *A... a criança que fala dialecto com mais velhos é falta de respeito?*

INF: *Não.*

DOC: *Hum, pode falar?*

INF: *Pode falá.*

DOC: *Mas por que que sua avó não gostava?*

INF: *Minha avó diz gen era muita criança ainda quê pa tá usa dialecto e assim quem gen pa um meio de pissoa vai falar assim mesmo.*

DOC: *Hum.*

INF: *Calha feio.* [ANTOM1]

No ponto 1.1 do presente abordou-se a questão dos fenómenos induzidos por contacto: a transferência e o empréstimo (Thomason & Kaufman, 1988:73-74). O santomense e o português dos “contratados” compartilhavam algumas afinidades relativamente ao sistema de marcação PL, já que ambas as línguas recorriam a um marcador pré-nominal, ou seja, um marcador inserido apenas antes do núcleo do SN, mantendo-se invariáveis todos os restantes elementos da cadeia sintagmática. Este sistema de marcação ancora no dos substratos do grupo níger-congo atlântico e resulta de transferência, já que os falantes responsáveis por esta eram monolíngues e, à medida que elaboravam a interlíngua, transferiam estruturas da língua nativa para a língua na qual procuravam comunicar (Thomason & Kaufman, 1988:73-74). Posteriormente, os falantes bilingues do português nativizado levaram a cabo empréstimos a partir da LA (Thomason & Kaufman, 1988:73-74), incorporando os determinantes do português no seu sistema de marcação e diminuindo a frequência de utilização de formas que revelavam apego às estruturas do santomense, como acontece com o uso do pronome pessoal terceira pessoa do PL (exemplo [48]) ou de nomes nus (exemplos [73] e [74]). No entanto, apesar de a faixa etária mais nova de Almoxarife ter tido um contacto mais próximo com formas da LA, sobretudo por intermédio do ensino, os factores linguísticos e extralinguísticos que intervieram na instanciação de novos/errados parâmetros (Gonçalves, 2004:235) não possibilitaram que ocorresse o restabelecimento de novos valores paramétricos (cf. Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashes, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), suficientes para diluir completamente os traços já adquiridos. Desta forma, o PA reestruturado revela algumas diferenças em relação ao PE, seja a nível fonológico (p.e. fenómenos de lambdacismo ou de abaixamento na realização de vogais altas do PE) seja no plano da morfossintaxe (p.e. não realização de determinadas preposições ou variação na concordância nominal de número e na concordância verbal), ocorrências típicas das situações de transmissão linguística irregular e que, no caso concreto da fala de Almoxarife, respaldam nas duas

gramáticas em competição: crioula do substrato santomense e portuguesa, da LA (PE elitista e do ensino).

Os fenómenos que afectam o PA apresentam uma incidência mais significativa entre os informantes mais idosos, fazendo prever que a frequência da variação seria mais elevada na fase em que a comunidade foi constituída, o que confirmaremos empiricamente nos capítulos 4 e 5, dedicados à análise das variáveis constituídas. Ainda assim, é possível detectar no PA estruturas variáveis decorrentes de processos de não-aquisição de matéria gramatical que acontecem na sequência do contacto entre línguas. Listamos, seguidamente, alguns dos traços que caracterizam o dialecto de Almoxarife:

#### ► Estrutura do SN

- (i) Uso variável do artigo definido (exemplos [240] e [243]), que pode determinar a realização de nomes nus:

[124] PA:  
*depois eu ia tê com meu amigo* [OSVALH1]  
“depois eu ia ter com o meu amigo”

[125] PA:  
*faço trabalhos caseira até tarde* [CARMOM1]  
“faço os trabalhos caseiros até tarde”

- (ii) Variação na colocação do possessivo (exemplo ([68]);

[126] PA:  
*E amigo meu tinha uma mulher* [OSVALH1]  
“E o meu amigo tinha uma mulher”

- (iii) Concordância variável de número entre os elementos do SN, com tendência para a marcação ser feita com recurso a um PL *singleton*, visível na categoria DET (Castro & Pratas, 2006), dispensando, depois, a marcação no nome núcleo (Castro & Ferrari-Neto, 2007:75) e em outros elementos do SN (Kiparsky, 1972:195), por se revelar redundante:

[127] PA:  
*já trabalhê esses lugar tudo* [LUISH2]  
“já trabalhei nesses lugares todos”

- (iv) Variação na flexão de género dos nomes referentes a seres animados, com recurso ao valor referencial do sexo usando nomes adjetivados:

[128] PA:  
*Dois filho mulher qu'ê tem* [FRANCM2]  
“As duas filhas que eu tenho”

- (v) Concordância variável de género entre os elementos do SN:

[129] PA:  
*faço trabalhos caseira até tarde* [CARMOM1]  
“faço os trabalhos caseiros até tarde”

### ► Morfologia verbal

- (vi) Uso de formas do presente para expressar acções e estados passados:

[130] PA:  
*Ê os meus pai que me deu essa ideia.* [OSVALH1]  
“Foram os meus pais que me deram essa ideia”

- (vii) Uso do modo indicativo para expressar acções e estados que requerem modo conjuntivo (exemplo [131]) e vice-versa (exemplo [132]):

[131] PA:  
*eu ñ viu nenhum governo que tá judá nós* [FRANCM2]  
“eu não vi nenhum governo que nos esteja a ajudar”

[132] PA:  
*às vez ponha quinze* [MMDEUSM3]  
“às vezes põe quinze”

- (viii) Uso de formas aparentemente finitas em contextos que requerem formas do infinitivo (exemplo [133]), e vice-versa (exemplo [134]):

[133] PA:  
*pode compra livro depois mais tarde* [CELESH3]  
“pode comprar o livro depois, mais tarde”

[134] PA:  
*Viver mal, viver num país tá muito mal* [MMDEUSM33]  
“Vivemos mal, vivemos num país que está muito mal”

- (ix) Não distinção entre posse e existência, com recurso ao verbo “ter” para expressar a segunda:



[135] PA:  
*aqui ni praia Mocharife ñõ tem nada ñõ tem luz, ñõ tem nada*  
[OSVALH1]  
“aqui na praia de Almocharife não há nada, não há luz, não há nada”

(x) Redução de concordância verbal, que atinge todas as pessoas do verbo:

[136] PA:  
*eu ñõ viu nenhum governo que tá judá nós* [FRANCM2]  
“eu não vi nenhum governo que nos esteja a ajudar”

[137] PA:  
*nós compra coesa bucado barato* [FRANCM2]  
“nós comprávamos as coisas um bocado baratas”

[138] PA:  
*É os meus pai que me deu essa ideia.* [OSVALH1]  
“Foram os meus pais que me deram essa ideia”

(xi) Variação relativamente à presença de verbo copulativo (exemplo [139]), que pode também ser substituído por verbo paracopulativo (exemplo [140]):

[139] PA:  
*depois dôtör disse aqui só pa fazer operação* [SALVH3]  
“depois o doutor disse-me que ali era só para fazer operações”

[140] PA:  
*Não, tá sorteido.* [FRANCM2]  
“Não, sou solteira.”

### ► Relações sintagmáticas e oracionais

(xii) Variação no emprego das preposições (exemplo [141]), que pode passar pela omissão destas (exemplo [142]), pela aplicação de formas redundantes (exemplo [143]) ou pelo uso do item “ni”, em substituição de “em” (exemplo [144]):

[141] PA:  
*o quê você faz de manhã até na hora que ocê vai dormir de noite?*  
[OSVALH1]  
“o que é que você faz desde manhã até à hora em que você vai dormir à noite”

[142] PA:  
*É, bêra strada* [CLOTIM2]  
“É, à beira da estrada”

[143] PA:  
*daqui ê fui para até hospital* [OSVALH1]  
“daqui, eu fui ao (para o) hospital”

[144] PA:  
*porque eu estive lá ni empresa trabalhar* [MANOH3]  
“porque eu estive lá na empresa a trabalhar”

(xiii) Variação no emprego de complementadores:

[145] PA:  
*há momento pessoa levanta matabicha, nõ almoça até jantá* [ANTOM1]  
“há momentos em que a pessoa se levanta, matabicha,<sup>46</sup> não almoça até ao jantar”

[146] PA:  
*depois dôtor disse aqui só pa fazer operação* [SALVH3]  
“depois o doutor disse-me que aqui era só para fazer operações”

(xiv) Não distinção entre formas rectas e átonas do pronome pessoal, com tendência para uso das primeiras em detrimento das segundas:

[147] PA:  
*eu tem entregá el coesa del* [OSVALH1]  
“eu tinha de lhe entregar as coisas dele”

[148] PA:  
*cercaram ele* [CARMOM1]  
“cercaram-no”

(xv) Não distinção entre as formas rectas e tónicas do pronome pessoal, usando-se as segundas pelas primeiras:

[149] PA:  
*mim depois saída de outro comício* [CELESH3]  
“eu depois saí de outro comício”

(xvi) Não distinção entre as formas do OD e OInd pronominais, substituindo-se as primeiras pelas segundas:

[150] PA:  
*nõ lhe vi* [ABILH2]  
“não o vi”

(xvii) Variação na colocação do pronome pessoal:

[151] PA:  
*o seôro se dá bem co esse seu filho* [ABILH2]  
“o senhor dava-se bem com esse seu filho”

A variação e/ou não-realização de partículas gramaticais (artigos, morfemas flexionais, preposições e complementadores) são marcas distintivas do PA reestruturado, enquanto dialecto que emergiu por transmissão linguística irregular. A sociohistória da comunidade de Almojarife, por seu lado, concede evidências que permitem relacionar, empiricamente, tais fenómenos quer com o processo massivo, defectivo e não-normatizado de aquisição do português, marcado por acentuada alteração nos seus parâmetros sintácticos e levado a cabo por escravos africanos geográfica e culturalmente transplantados para o entreposto de São Tomé (p.e. não realização de preposições), quer com o processo defectivo e não-normatizado de nativização do PtgL2, marcado por fenómenos de reestruturação morfossintáctica mais leves (p.e. CPL-var no SN, com a marcação a incidir apenas na categoria DET). Inicialmente, devido à situação de contacto restrito relativamente ao PE, o PA não teve substancial acesso aos modelos deste, o que determinou que o seu processo de reestruturação ocorresse sem referências normatizadoras. Com campo aberto para a fixação de estruturas desviantes, o PA incorporou variação de diferentes tipos, que foi transmitida de geração em geração. Mais tardiamente, com o nascimento de um sistema de rodovias e de uma rede de transportes públicos que interligava meios rurais e urbanos, a criação e expansão de um sistema de educação pública abrangente, a implantação da rádio na ilha e a incorporação de nativos nos quadros do exército colonial e da administração pública local, alguns falantes do PA passaram a ter contacto mais estreito com a LA. Este facto possibilitou que a variação se diluísse, mas o processo em direcção à aquisição das regras da norma foi interrompido após a independência do país, já que o sistema socioeconómico deste se desestruturou, lançando as comunidades do interior para um isolamento que condiciona o *continuum* linguístico. Como tal, a variação registada na comunidade encontra-se presentemente estabilizada (capítulo 5, análise das variáveis extralinguísticas), pelo que os actuais traços do dialecto de Almojarife terão que ser atribuídos não só ao processo de transmissão linguística irregular que levou ao seu aparecimento mas também ao extenso período de semi-isolamento em que a comunidade se encontrou (e encontra) mergulhada.

Os fenómenos de variação que se detectam no PA não encontram similaridades significativas no PE, mas estabelecem paralelos claros com outras variedades africanas e brasileiras de português que emergiram por contacto, nomeadamente no que concerne à variação que atinge os sistemas de concordância verbal e de concordância nominal de número. Esta particularidade permite-nos, desde já, questionar a hipótese de estes fenómenos ancorarem em mudanças fonéticas inerentes ao próprio português, como tem sido defendido pela corrente derivista a propósito das motivações históricas das características que diferenciam o PB do PE (Naro, 1981; Naro & Scherre, 1993, 2000, 2007a, 2007b). Por outro lado, os paralelismos referidos permitem assumir que o contacto entre línguas terá sido crucial para a emergência e fixação dos fenómenos em questão quer no PA quer no PB. Assim, uma análise variacionista e comparativa sobre a CPL-var sistemática no SN do PA, enquanto dialecto falado por descendentes dos antigos escravos transplantados para o entreposto de São Tomé na rota do tráfico afro-negreiro para o Brasil, permitirá visualizar até que ponto se pode estabelecer uma corrente que conecte tal fenómeno às línguas africanas (cf. Guy, 1981a, 1989, 2005).

### **1.3. Conclusões parciais**

No presente capítulo lançámos um olhar sobre alguns trabalhos acerca da CPL-var no SN em variedades brasileiras e africanas de português, emergentes em situação de transmissão linguística irregular. A CPL-var no SN de tais variedades reflecte aspectos da própria sociohistória que determinou o aparecimento das comunidades que comunicam com recurso a elas. No caso do Brasil, este tipo de variação tem sido conotado, por um lado, a mudanças fonéticas naturais internas ao português, que foram determinadas pela deriva diacrónica (Naro & Lemle, 1976; Naro, 1981; Naro & Scherre, 1993, 2000, 2007a, 2007b), e, por outro lado, a fenómenos morfossintácticos de reestruturações colaterais, resultantes do contacto massivo entre línguas ocorrido no país (Baxter & Lucchesi, 1997, Lucchesi, 1998a, 2001; Baxter, 2009), mas que radicam nos ancestrais dialectos africanos falados pelos escravos africanos transplantados e seus descendentes (Guy, 1981a, 1989, 2005; Baxter & Lucchesi, 1997, 1998; Lucchesi, 2001; P. Andrade, 2003; Baxter, 2009). No caso específico da comunidade de HEL-Ba, Baxter (2009:293) observou a CPL-var no SN combinando teorias estruturais e aquisicionais,

tendo concluído que: (i) a variação é motivada por questões como o desequilíbrio demográfico entre contingentes escravizados e não-escravizados, a mudança de LA na comunidade, a presença de bilinguismo nesta, ou as situações de pouco acesso às formas linguísticas da LA; (ii) o desenvolvimento da concordância PL inicia-se com a introdução de um PL *singleton* visível na categoria DET, para, posteriormente, ganhar características de concordância parcial, através de morfemas dissociados, que são atribuídos aos elementos do SDET mais próximos do núcleo, primeiro, e aos itens à direita deste, depois. Para Baxter (2009:292), a variação da concordância deverá ser conectada à situação de contacto que envolveu componentes de PtgL2 influenciado pelas línguas bantu, cujas estruturas de marcação PL se assemelham às das línguas atlânticas da tipologia níger-congo atlântica (balanta e wolof), substratos do CVV. Nestas, a pluralização é controlada por um classificador nominal que estabelece paralelismos estruturais com a categoria DET das variedades africanas e brasileiras de português, reestruturadas por transmissão linguística irregular, e que tenham, como substratos directos, quer as próprias línguas africanas do grupo níger-congo atlântico (PM, PVA e PT) quer os CP's atlânticos dos quais estas são também substratos (PA e PCV). Para reforçar este pressuposto, apresentámos exemplos de paralelismos estabelecidos a nível do SN entre línguas africanas atlântico-ocidentais, CP's atlânticos, variedades africanas de português com substrato do grupo níger-congo atlântico, variedades africanas com substrato crioulo e variedades brasileiras de português. Quanto à manutenção geracional da CPL-var no SN nas referidas variedades reestruturadas, é determinada pela preponderância quantitativa de falantes de PtgL2, que esbate a possibilidade de um reajustamento paramétrico e leva a que os novos/errados parâmetros sejam tomados como modelos para uma nova reinstanciação paramétrica no PtgL1 nativizado (Gonçalves, 2004:235).

A elaboração do capítulo abordou ainda a descrição sociohistórica da comunidade de Almojarife e levantou as marcas pertinentes do português reestruturado falado pelos almojarifanos. As marcas apontam para a possibilidade de o seu entendimento ter que ser verificado a partir de constatações que ancoram em modelos teóricos distintos, sejam eles da aquisição, sociolinguísticos ou psicolinguísticos. Portanto, o suporte para análise dos nossos dados terá que ser o mais abrangente possível, pelo que, no próximo capítulo,

daremos conta dos quadros teóricos que, no nosso entender, poderão ajudar a determinar se a aquisição da pluralização nos constituintes do SN, e seu posterior desenvolvimento estrutural, ancoram em princípios regidos pelos universais linguísticos ou, se, contrariamente, configuram quer casos de transferência que se fossilizam quer empréstimos provenientes da LA.



## CAPÍTULO 2

*“Ab uno disce omnes.”*

(In, Virgílio, *Eneida*, II, p.65)

### Fundamentação teórica

Algumas das variedades de português que emergiram em situação de transmissão linguística irregular mais radical no Brasil exibem traços de crioulização leve (Baxter, 1992, 1997, 2009; Baxter & Lucchesi, 1997; Lucchesi, 2000a; Lucchesi & Baxter, 2009). No presente capítulo abordaremos as perspectivas teóricas que permitirão verificar se as estruturas crioulizantes de tais variedades apresentam idêntico perfil ao de outras variedades de português, especialmente as que possuem um crioulo como substrato, como acontece com o PA. Portanto, o estudo da CPL-var no seu SN é fundamental para se perceber a forma como os factores que regulam a variação e a mudança actuam em conjunto com o processo de aquisição da L1 para determinarem o modo como o conhecimento linguístico se orienta na mente quer dos falantes almorarifanos quer dos falantes das referidas variedades de PB. Assim, e independentemente de estarmos verificando se a aquisição dos constituintes do SN e seu consequente desenvolvimento estrutural ancoram em princípios regidos pelos universais linguísticos ou, em contrapartida, configuram casos de transferência, mais ou menos fossilizada, estaremos também assumindo que a variação e mudança decorrentes no PA, a existirem, não serão unicamente reguladas pelas suas estruturas gramaticais mas também por estruturas inatas comuns ao ser humano e formalizadas pelo conceito da existência de uma GU (Chomsky, 1957, 1965, 1981a, 1981b, 1986a, 1996 [1995]).

Para Labov (1982:18), a Sociolinguística Variacionista deve centrar atenção na forma como os factores sociais determinam os padrões linguísticos usados colectivamente numa certa comunidade, elaborando os estágios do desenvolvimento histórico da sua fala. Desta forma, a mudança linguística pode ser predita estabelecendo-se a relação entre a variação linguística analisada num momento específico do devir da fala dessa mesma comunidade e o sentido em que a mudança em curso vem operando. Portanto, as variações reflectem estágios não-aleatórios do processo de mudança em progresso, tendo na sua origem a conjunção de fenómenos linguísticos e sociais que vão edificando, progressiva e estruturalmente, o todo linguístico-comunitário.



Assim, caberá à Sociolinguística o papel de observar os fenómenos linguísticos numa perspectiva sociohistórica, relacionando subjectivamente, à estrutura social e ideológica da comunidade, os dados empíricos que reflectem variação e indiciam uma mudança em curso. Não obstante, a não elaboração de uma teoria da estrutura em termos sociolinguísticos tem criado obstáculos ao encaixamento dos processos de variação e mudança na estrutura linguística, uma vez que apenas um reduzido número de estudos (p.e. Tarallo, 1987, 1991; Kato & Tarallo, 1989; Kato, 1999; Ramos, 1999; Duarte, 1999; Kroch, 1989b, 2001, 2004; Kroch & Taylor, 1997; Roberts, 2007) tem procurado estabelecer uma relação entre o comportamento linguístico dos falantes que integram uma dada comunidade de fala (língua-E) e a sua competência linguística enquanto usuários nativos de uma determinada língua natural (língua-I).

No presente capítulo procuraremos dar conta dos princípios teóricos que orientarão a metodologia do nosso estudo, a qual adoptará como pedra basilar de fundamentação a conjugação dos dois modelos referidos, que serão complementados por outros quando o entendermos necessário. Ainda que os modelos concernentes à língua-I e língua-E apresentem divergências na sua forma de observar as línguas, estas divergências não deixam também de se conjugarem “*para o entendimento da questão da mudança, tanto no plano dos princípios teóricos fundamentais, quanto no plano da metodologia e dos critérios da verificação empírica*” (Lucchesi & Ribeiro, 2009:126). Assim sendo, no presente estudo procuraremos, de alguma forma, integrar a dimensão psico-biológica do fenómeno da CPL-var registada no SN do PA na dimensão sociohistórica da comunidade almoxarifana. Com tal metodologia, buscamos entender até que ponto os mecanismos da faculdade humana da linguagem (cf. Chomsky, 1986a, 1996 [1995]) intervêm também nas relações sociais da comunidade, conjugando-se com estas no processo que determina o fenómeno da variação registado a nível da concordância de número entre os elementos do SN do português falado em Almojarife.

Ao aplicarmos os princípios teóricos e a metodologia mencionados ao estudo da CPL-var no SN do PA não temos a pretensão de fornecer uma solução para colmatar a problemática que emerge quando se pretende conjugar dois modelos tão distintos. No entanto, acreditamos que é possível recorrer também a elementos da teoria da gramática para uma análise dos padrões colectivos de comportamento linguístico da comunidade

almoxarifana, contribuindo, com tal, para se observar como se processa a aquisição e regularização de fenómenos variáveis em línguas emergentes em contexto massivo de mono e bilinguismo, sejam elas L2's sejam elas L1's. De facto, analisando os processos em que as gramáticas concorrentes na comunidade de Almoarifife afectam o desempenho linguístico dos seus utilizadores, isto é, o modo como as distintas gramáticas geracionais determinam alternâncias de códigos que se encaixam no conhecimento linguístico dos almoarififanos para elaborarem estruturas gramaticais coesas, é possível não só compreender o desenvolvimento da CPL-var do PA, em geral, mas também fornecer mais alguns dados para complementar o entendimento acerca do *puzzle* que é a realidade actual dos falares rurais e populares do Brasil.

Face ao exposto, iremos focar aspectos em torno das abordagens sobre a teoria da gramática e o estudo dos padrões colectivos de comportamento linguístico, nomeadamente acerca de alguns trabalhos que buscam conjugar os modelos discutidos sob o ponto de vista da Sociolinguística (cf. Weinreich, Labov & Herzog, 2006 [1968]) e da Gramática Generativa, a fim de explicarem, com recurso a modelos matemáticos para a reprodução do efeito global dos dados empíricos, como se dão as mudanças linguísticas a nível comunitário.

Paralelamente, e visto o nosso estudo ter como objectivo principal observar o comportamento da concordância de número a nível do SN, abordaremos a questão da flexão em número nos constituintes que integram o SN do PE, verificando-se que é aparente o conceito de que este idioma insere marcas de número PL em todos os elementos do SN. Para tanto, importará também caracterizar o funcionamento do SN em português, referindo as dificuldades que o sistema de flexão e concordância em número podem colocar na aquisição do PtgL2. Para esta caracterização, iremos descrever, sucintamente, o quadro teórico que orientará as nossas observações, sobretudo as propostas em torno da aquisição de L1, aquisição bilingue e ASL em situações sociohistóricas específicas.

Trataremos igualmente a questão particular da aquisição da categoria funcional SDET, na qual se alojam os traços de género (Koehn, 1994; Müller, 1994a; Ferrari-Neto, Corrêa & Augusto, 2005) e de número (Enç, 1991; Koehn, 1994; Müller, 1994a; Longobardi, 1994; Ferrari-Neto, Corrêa & Augusto, 2005; Castro & Pratas, 2006; Costa & Silva,

2006b; Alexandre & Hagemeyer, 2007; Castro & Ferrari-Neto, 2007; Baxter, 2009). Neste aspecto, será também fundamental perceber como a concorrência entre gramáticas pode afectar a morfologia de número a nível do SDET em situações de transmissão linguística irregular. De facto, esta configura a realidade sociolinguística que determinou a emergência do PA e de determinadas variedades reestruturadas do PB, caracterizadas por uma não-realização de morfologia flexional, que pode ir de grau mais leve a grau mais acentuado. Dado que a manutenção geracional da não-realização poderá estar directamente conotada à não-reinstanciação de determinados parâmetros, que determina fossilização, importará também fazer referência ao possível papel desta no processo de reestruturação do português falado em Almocharife.

## 2.1. Os pressupostos teórico-metodológicos sobre aquisições de L1, bilingue e L2

As variações da fase precoce da aquisição de uma L1 em situação que não configura transmissão linguística irregular são geralmente regularizadas bastante cedo. Contrariamente, tanto a L2, adquirida sobretudo após o período crítico, como a L1 adquirida em contexto de transmissão irregular são sistemas linguísticos que mantêm as variações por mais tempo, podendo algumas fossilizar-se. No PA convivem diferentes gramáticas geracionais e até individuais, com determinadas realizações do mesmo falante revelando apego à reduzida morfologia do santomense (exemplo [152a]), enquanto outras apresentam características da elaborada morfologia do PE (exemplo [152b]):

[152] PA:

- a. *Foi o tempo do Frente, os homem de Frente.*<sup>47</sup> [MAURIH2]  
“Foi no tempo da Frente, dos homens da Frente”
- b. *resolvê poblema dos homens que tão cá* [MAURIH2]  
“resolver o problema dos homens que estão cá”

A variação na marcação de número tem sido conotada quer a mudanças internas ao próprio português, no caso, a deriva românica (Naro, 1981; Scherre, 1988; Naro & Scherre, 1993, 2000, 2007a), quer ao sistema de pluralização dos substratos africanos (Guy, 1981a:301-302) do grupo níger-congo atlântico (Baxter, 2009:292), que faz incidir o papel fulcral PL no classificador pré-nominal. Contudo, uma explicação da variação linguística com base exclusiva em comparações entre a L1 e a L2 é bastante

reduzida (Baxter, 2004; Godinho, 2005), já que existem paralelismos entre a aquisição da L1 e da L2, com os aprendentes a atravessarem diversos estádios na aplicação das regras da língua em elaboração, até atingirem um patamar final de aquisição (Cerqueira, 1994; Lléo, 1997, 1998, 2001; López-Ornat, 1997; Faria, Freitas & Miguel, 2001; Name & Corrêa, 2003; Plag, 2008a, 2008b).

Na ASL, o aprendente elabora uma interlíngua (Selinker, 1972:216), marcada por etapas que se vão aproximando da LA (*approximative system*) (Nemser, 1971:119). Este sistema apresenta interferências da L1<sup>48</sup> a nível morfossintático, fonológico, lexical e semântico (Thomason & Kaufman, 1988:73-74), originando, por vezes, estruturas com marcas da L1 e da LA, e que demonstram que o falante detém conhecimentos da segunda, mas resiste às suas regras gramaticais. Contudo, o facto de detentores de diferentes L1's revelarem idênticos estágios de desenvolvimento na aquisição de uma L2 levanta a hipótese de ambas as aquisições serem também determinadas por mecanismos cognitivos internos e independentes do substrato (Romaine, 2003:411; Godinho, 2005:55). Para Plag (2008a:115), o processo de interlíngua ocorre unicamente a nível individual, enquanto a emergência de uma nova língua em contexto de contacto massivo pressupõe a aquisição e uso colectivo de uma língua-E, por um lado, e de uma língua-I, por outro lado. Daí que a noção de interlíngua e da sua sistematicidade constituam, hoje em dia, pontos fulcrais dos estudos em ASL (p.e. Bayley & Preston, 1996; Siegel, 2004a, 2004b, 2006, 2008; Franseschina, 2005; Pienemann, 2000, 2005a, 2005b; Bruhn de Garavito, & Atoche, 2006; Sprouse, 2006).

Alguns autores têm apontado paralelismos entre a aquisição da L1 e da L2, que afectam quer as crianças quer os adultos detentores de diferentes L1's (Bruhn de Garavito & White, 2002; Montrul, 2004; White *et alii*, 2004). As etapas de variação no processo de aquisição de línguas são influenciadas por fenómenos vários, que vão desde a acção dos mecanismos cognitivos internos relacionados com a L1 e a L2 (Montrul, 2004), até às regras não conectadas a esta (Pienemann, 1998, 2000, 2005a, 2005b; Pienemann & Håkansson, 2007; Plag, 2008a, 2008b, 2008c, 2008d), sem deixar ainda de passar pela própria competência linguística subjacente a cada falante. No caso concreto da ASL massiva, são determinantes, para a maior ou menor variação registada na interlíngua, os contextos sociohistóricos em que esta se dá (Winford, 2003a:355),

fundamentais para a motivação dos falantes em relação à aprendizagem da LA. Por conseguinte, uma correcta análise da variação e da sistematicidade da interlíngua e das línguas nativizadas terá que levar em conta pressupostos relacionados com as perspectivas generativista, sociolinguística e psicolinguística (Ellis, 1994; Godinho, 2005; Baxter, 2009; Silva, 2010).

### **2.1.1. Variação e sistematicidade na *interlíngua*: perspectiva generativista**

O estudo dos fenómenos da *interlíngua* devem distinguir entre o conhecimento da língua que o falante detém, realmente, e aquilo que ele produz em situação de comunicação (Yip, 1995:6, *apud* Godinho, 2005:56, nota 66). Em termos generativistas (Chomsky, 1965), a competência não será a variação detectada na *performance*, mas antes, o conhecimento interiorizado que o falante detém da L2. Desta forma, o patamar avançado de aprendizagem é a condição de excelência para os estudos em gramática generativa, não podendo a discussão em torno da sintaxe complexa demarcar-se das especificidades inerentes à intervenção da GU. Esta activa quer os princípios universais das línguas quer os parâmetros específicos de cada idioma, permitindo que o aprendente construa a gramática da LA com base em *input* restrito. Este aspecto determina que a competência subjacente à *performance* apresente paralelos entre L1 e L2, interessando apurar o conhecimento que o falante detém de um determinado sistema linguístico e não o modo como o utiliza (Lightbown & Spada, 2003 [1993]). Ainda assim, alguns generativistas (p.e. Bruhn de Garavito & White, 2002) não se têm limitado a comparar juízos de gramaticalidade entre falantes de L1's e L2's, uma vez que se debruçam também sobre dados da *performance* (Godinho, 2005:57).

Assim, e no que concerne concretamente às abordagens que procuram estabelecer uma relação entre a mudança linguística e a teoria da gramática generativa, é assumido que a gramática da criança é construída com base na conjugação estabelecida entre os dados auditivos (os DLP's) e os universais linguísticos biologicamente determinados para aquisição das línguas (a GU). A GU demarca o que pode variar entre línguas (os parâmetros) e aquilo que é constante (os princípios). Os primeiros constituem um conjunto de opções definido pela própria GU, que estabelecem as diferenças entre dois sistemas gramaticais. Por seu lado, os segundos configuram os mecanismos mentais

predeterminados dos falantes que, ao interagirem com os parâmetros, permitem analisar as estruturas que poderão estar a violar um dado sistema linguístico.

Apesar de os princípios serem imutáveis, os DLP's podem ser afectados por mudanças culturais, sociais e pragmáticas, o que os leva a variarem de geração em geração. Deste modo, o *input* linguístico que determina a aquisição da linguagem conduzirá, inevitavelmente, a mudanças, vistas como o resultado de um processo abduativo (lei + resultado => caso) (Lightfoot, 1979; Lucchesi & Ribeiro, 2009).

Trabalhos diversos sobre as diferenças de parâmetros entre línguas têm também tentado descrever o parâmetro e suas consequências, bem como as implicações tipológicas e experiências decorrentes das distintas selecções de valores paramétricos. Especificamente no que diz respeito às diferenças paramétricas registadas em diferentes estágios de uma mesma língua, é fundamental explicar não apenas qual o parâmetro que regista diferença e os efeitos originados por esta mas também o porquê de um novo valor se sobrepor geracionalmente a outro (Adams, 1987; Vance, 1989; Roberts, 2007).

Em alguns estudos sobre variação e mudança linguística tem sido assumido que a mudança resulta, essencialmente, da competição entre formas conservadoras e outras que se vão inovando ao serem influenciadas por factores externos de diversos tipos. Portanto, trata-se de um fenómeno que evolui difusa e gradualmente no tempo, em virtude de as gramáticas serem geracionalmente construídas, uma vez que as crianças não se podem abstrair das mudanças que afectam os DLP's a que estão expostas (Lightfoot, 1979:391).

No entanto, tem também sido assumido que a mudança é algo que acontece de modo abrupto, envolvendo apenas a transição de uma geração para outra (Kroch, 1989b). Este aspecto resultará do facto de nenhum sistema linguístico se poder encontrar em estágio intermediário entre dois valores paramétricos. Com efeito, e de acordo com os princípios da Lei do Meio Excluído (*Law of the Excluded Middle* – Roberts, 2007:296), os valores paramétricos respeitantes a entidades como verbos, nomes, tempos ou fonemas têm de ser tratados como entidades formais de natureza discreta, visto os seus traços possuírem apenas um de dois valores possíveis.

Estes pressupostos configuram uma visão generativista sobre parâmetros, aquisição e mudança, sendo ainda coadjuvados por mais dois princípios fundamentais: (i) a

mudança catastrófica ocorrida em determinado momento histórico altera, irreversivelmente, determinados valores paramétricos; (ii) a mudança é um fenómeno que diz respeito apenas à língua-I, não sendo determinada por questões conectadas à língua-E (p.e. o ambiente social, cultural ou histórico em que o aprendente se encontra inserido). À primeira vista, parece que os três pressupostos não encontram pontos de conciliação com a visão da Sociolinguística Variacionista, mas alguns autores entendem que tal contradição é meramente aparente (Kroch, 2001; Roberts, 2007), conquanto se conciliem *“as ideias de sistemas algorítmicos homogêneos, discretos, que são propriedades dos indivíduos, com a mudança gradual da linguagem num dado momento histórico em uma dada comunidade de fala, formalizada analiticamente, pelos sociolinguistas, por algoritmos heterogêneos, graduais e probabilísticos”* (Lucchesi & Ribeiro, 2009:133).

Segundo Roberts (2007), os efeitos das mudanças discretas e instantâneas a nível estrutural encontram-se inseridas nos dados históricos, pelo que é possível observá-las conciliando e analisando a frequência de realizações entre formas inovadoras e conservadoras com o trajecto temporal da mudança linguística. Por outras palavras: estudando-se, em determinado sistema linguístico, o efeito gradual exercido por factores independentes uns dos outros, ou seja, factores sociais (idade, escolaridade, variação idiolectal, etc.) e factores de natureza gramatical (difusão lexical, microparâmetros, macroparâmetros, etc.), obtém-se uma visão acerca do modo como ocorrem as mudanças no mesmo. No que concerne à difusão lexical, tem sido confirmado que mudanças fonológicas (Labov, 1994) e mudanças de traços formais do léxico substantivo (Lightfoot, 1991a) se difundem gradualmente, afectando o sistema lexical de determinados idiomas. Contudo, no segundo caso, os processos de reanálise podem afectar somente estruturas específicas, incidindo unicamente sobre determinados parâmetros do léxico funcional (mudanças paramétricas), e não sobre a totalidade da gramática (Borer, 1984). Nestas situações, e dado que as categorias funcionais são compostas por traços formais, estamos perante casos de mudança gradual, que se opõem aos casos de mudança abrupta, em que é afectado todo o léxico substantivo (Roberts, 1993, 2007).

Quando uma série de mudanças discretas atinge, gradualmente, o conjunto de traços de categorias funcionais de um determinado dialecto contemporâneo, é necessário identificar, a partir da noção de microparâmetros, quais as unidades mínimas da língua histórica que registaram variação paramétrica (Kayne, 2000; Ladgeway, 2000). Efectivamente, variações sincrónicas mínimas registadas em traços que alteram os seus valores no conjunto poderão apontar para uma mudança diacrónica gradual nas comunidades (Roberts, 2007:300-305). Contudo, as mudanças graduais assentes em opções de natureza sociolinguística, que levam determinada forma a tornar-se predominante, poderão não constituir mudança paramétrica. Este aspecto torna-se elucidativo em trabalhos com recurso ao modelo minimalista da gramática generativa, os quais demonstram que a opcionalidade formal numa mesma gramática pode constituir apenas uma possibilidade técnica, sem consequências semântico-pragmáticas (Biberauer & Roberts, 2005; Biberauer & Richards, 2006; Roberts, 2007).

Face à complexidade que a variação nos usos de uma determinada comunidade pode encerrar, Kroch (1989b:200) alertou para a hipótese do Efeito de Taxa Constante (*Constant Rate Effect*), em que a mudança gradual é consequência de gramáticas em competição, que põem em contraste formas conservadoras e inovadoras vernaculares, levando uma delas a impor-se diacronicamente em determinados contextos linguísticos, por melhor se adequar a eles. Tal imposição não deve, então, “*ser considerada um fato da gramática em si mesma, e sim um fato sobre a forma como os indivíduos com competência em mais de uma gramática acessam as opções disponíveis e sobre como os fatores sociais atuam nessas escolhas*” (Lucchesi & Ribeiro, 2009:136).

Todavia, as abordagens sobre mudanças paramétricas e mudanças no léxico substantivo não fornecem resposta para se analisar a forma como a mudança afecta massivamente determinadas comunidades, já que a observam unicamente a nível do sujeito. Ainda assim, a proposta generativa de que a mudança linguística é direccionada pela aquisição da linguagem permite entender o problema da implementação (*actuation problem*) (Weinreich, Labov & Herzog, 2006 [1968]), isto é, observar os estágios em que as duas formas de uma mesma língua têm a sua maior intervenção. Por seu lado, a noção de gramáticas em competição e do Efeito de Taxa Constante (Kroch, 1989b) ajuda a explicar questões relacionadas com a problemática da transição (*transition*



*problem*) (Weinreich, Labov & Herzog. 2006 [1968]), ou seja, como a nova gramática se difunde entre os indivíduos de uma comunidade de fala bilingue. De facto, forças externas actuando sobre as gramáticas em competição produzem variação a nível social, atestada na diglossia sintáctica (p.e. nível formal e nível coloquial), enquanto o bilinguismo origina registos marcados por alternância de códigos (*code-switching*) (ponto 2.3.1 do presente trabalho) ou mistura de códigos (*code-mixing*).

Note-se que a competição entre gramáticas, por si só, não fornece resposta cabal para se entender a mudança linguística na sua totalidade, já que não explica as questões conectadas ao início da mudança (*constraints problems*) nem à problemática da sua conclusão (*actuation problem*), vistas na perspectiva da Teoria da Variação (Weinreich, Labov & Herzog. 2006 [1968]). Assim, o conceito de diferenciação ordenada (*orderly differentiation*), fundamental para se entender a questão da mudança linguística, (Weinreich, Labov & Herzog. 2006 [1968]:151), sobretudo a problemática da transição (*transition problem*) que determina mudanças sintácticas, é explicado por Roberts (2007:320) com base em mudanças a nível paramétrico e que ocorrem em três estágios distintos: no primeiro acontece a remarcação do valor de um determinado parâmetro, que gera nova opção formal e muda a proporção de indivíduos bilingues na comunidade; no segundo, o valor paramétrico difunde-se na comunidade, colocando gramáticas em competição, tornando opaca a opcionalidade formal respeitante a um único sistema e alterando, gradualmente, a proporção de falantes em que as duas formas coexistem; por fim, a opção formal conservadora extingue-se no terceiro estágio.

Para explicarem a difusão gradual da mudança, alguns generativistas entendem-na como uma consequência de “erros” no processo de aquisição da L1 ou L2 e recorrem à noção de parâmetros discretos, procurando, com tal, definir onde ocorre a variação nos sistemas linguísticos em observação (Kroch, 2004). Se os referidos “erros” forem compatíveis com a GU, dá-se a mudança no valor do parâmetro e, seguidamente, a alteração na frequência de uso da forma, que a levam a propagar-se na comunidade. Visão contrária têm outros autores, que admitem um processo inverso na alteração da frequência de uso de uma certa forma linguística (Adams, 1987; Vance, 1989). Assim, esta será determinada por estratégias discursivas que alteram frequências de uso e tornam opaco um determinado parâmetro, conduzido à alteração no seu valor.

Seja de uma forma seja de outra, os estudos necessitam de recorrer à quantificação para poderem analisar as tendências gerais do fenómeno sintáctico em observação. Por outro lado, como as línguas variam parametricamente, o recurso à sintaxe comparada fornece indicações precisas sobre as diferenças registadas no desenvolvimento histórico de um determinado idioma, que podem ser vistas como o resultado da remarcação diacrónica de parâmetros.

### **2.1.2. Variação e sistematicidade na *interlíngua*: perspectiva psicolinguística**

A perspectiva psicolinguística debruça-se sobre os mecanismos internos que conduzem à variação detectável na *performance* (Godinho, 2005:57). Para justificar a variação, Selinker (1972:216) propõe a existência de uma estrutura psico-cerebral que é accionada quando o aprendiz começa a ASL, iniciando-se também a construção da *interlíngua*. Segundo esta perspectiva, a *interlíngua* e a L1 são sistemas unos e independentes, pelo que as produções de ambas representam o uso natural de uma língua (Adjemian, 1976:307). Assim, a realização linguística tem categorias distintas, visto assentar em diferentes fases de produção de uma língua (Levelt, 1989:147), que pode ser consciente, ou não. Para os psicolinguistas, a variação reside nas realizações de superfície das entradas marcadamente abstractas do léxico mental, ou seja, nos morfemas – os lemas – que encerram em si parte da informação semântica, sintáctica e até fonológica do item seleccionado para a produção (Biase & Kawaguchi, 2002; Håkansson, Pienemann & Sayehli, 2002; Pienemann, 2000, 2005a, 2005b; Plag, 2008a, 2008b). A realização de alguns dos morfemas é accionada de forma indirecta, visto deterem uma funcionalidade dependente do elemento directamente seleccionado, como acontece com as preposições que regem determinados verbos. Quanto à informação encerrada nos lemas, diz respeito aos três subsistemas da estrutura lexical (Ellis, 1994:130-131), listados por Godinho (2005:58):

- “ - *A estrutura lexical/conceptual que combina a informação semântica e pragmática disponível universalmente;*
- *A estrutura “predicado-argumentos”, que especifica as propriedades dos verbos em diferentes sub-categorias e a forma como são codificados*

*gramaticalmente os argumentos, isto é, quantos argumentos pode ter um verbo e qual o papel temático que cada um deles recebe;*  
*- Padrões de realização morfológica que explicam os mecanismos de superfície para ordem das palavras, caso, concordância, marcas de tempo e aspecto, etc.”.*

Estes subsistemas interagem, de modo progressivo, em etapas que se assemelham nas L1's e L2's (Myers-Scotton, 2002:23), indo desde as intenções de produção linguística a nível conceptual (tomada de decisões, por vezes inconsciente, e que levam em conta quer aspectos quer consequências sociais e psicolinguísticas), até à efectiva realização da mesma (Levelt, 1989:1-28, *apud* Godinho, 2005:58-59).

### **2.1.3. Variação e sistematicidade na *interlíngua*: perspectiva sociolinguística**

A perspectiva sociolinguística de análise da variação e da sistematicidade da *interlíngua* relaciona a variação linguística e a sociohistória dos seus falantes, ou seja, estabelece a relação entre a competência linguística e a *performance* (Dickerson, 1975; Beebe, 1980, 1988; Bayley, 1994, 1996; Preston, 1996a, 1996b; Chambers, Trudgill & Schilling-Estes, 2003 [2002]). Embora a nossa análise incida sobre dados da produção real, obtidos com recurso a entrevistas, é de considerar que os informantes poderão, de modo mais ou menos espontâneo, ter recorrido também à introspecção. Assim, e para termos em conta aqueles que serão os padrões colectivos de comportamento linguístico da comunidade de Almojarife, entendemos que os actos de fala dos seus membros, tal como sucede com os de qualquer outra comunidade, “*são regidos por factores relativos:*

- (i) *ao conhecimento intuitivo, que faz do indivíduo um membro da comunidade de fala (ou seja, um falante);*
- (ii) *a certas características sociais desse indivíduo (tais como sexo, idade, nível de escolaridade, etc.);*
- (iii) *aos juízos de valor desse indivíduo sobre as escolhas disponíveis no repertório linguístico da comunidade;*
- (iv) *à dimensão ideológica do ato de interação verbal, que remete às relações de poder, estratégias de convencimento, nível de interesse na conversação, etc.;*
- (v) *às circunstâncias específicas nas quais se realiza o ato, relativamente ao conhecimento compartilhado entre os participantes, à intervenção eventual de participantes adventícios ao ato em questão, etc.”* (Lucchesi & Ribeiro, 2009:129).

Para se entender a CPL-var no SN do PA, é fundamental estabelecer, com recurso a métodos quantitativos, a diferença entre realizações pontuais e sistemáticas. Como tal, integraremos no nosso estudo factores que possam reflectir, probabilisticamente, com recurso a variáveis que arrolaremos, as tendências de uso de formas linguísticas e sua proporção de aplicação nos distintos contextos linguísticos. Esta metodologia permitirá elaborar predições comunitárias, avaliando até que ponto as frequências de uso da CPL-var no SN estão relacionadas com uma situação de mudança linguística em curso.

Contudo, o modelo apenas dá conta dos aspectos relacionados com a língua-E, não facultando uma compreensão sobre o conhecimento internalizado na mente dos seus membros, a nível individual, ou seja, da sua língua-I. Nestas circunstâncias, se não observarmos também os pressupostos teóricos acerca da forma como o conhecimento linguístico internalizado determina o comportamento linguístico dos membros da comunidade de Almocharife, não nos é possível entender, abrangentemente, os aspectos que determinam a CPL-var no interior SN do PA nem quais os elementos estruturais que são efectivamente afectados por esta. Portanto, o recurso à teoria da gramática, visto na perspectiva da problemática das restrições (*constraint problems*) à luz da Teoria da Variação (Weinreich, Labov & Herzog, 2006 [1968]), será fundamental para estabelecermos os limites da variação na estrutura do SN do PA, observando não só onde estará situado o ponto de partida da implementação de um possível processo de mudança em curso (cf. Tarallo, 1991) mas definindo também os potenciais processos que determinam esta.

Paralelamente, a estrutura gramatical, ao intervir no processo de difusão da mudança, permite entender a problemática do encaixamento (*embedding problem*) (Weinreich, Labov & Herzog, 2006 [1968]), uma vez que determina quais os contextos específicos que favorecem certa variante e se a sua frequência de uso tem implicação directa no desenvolvimento diacrónico de uma outra variável linguística. Do exposto, verifica-se que a resposta para os fenómenos da variação e da sistematicidade não pode assentar apenas num dos modelos referidos (Zuelenger, 1989:66, *apud* Godinho, 2005:60). Nesta conformidade, Ellis (1994:120) propõe uma conjugação dos quadros teóricos sociolinguístico e psicolinguístico, para dar conta dos referidos fenómenos. Porém, e a exemplo da metodologia aplicada em determinados estudos sobre aquisição de L2/LE

(p.e. Godinho, 2005; Silva, 2010), estas perspectivas podem ser complementadas com outras abordagens, levando-se em consideração os princípios respeitantes à GU, já que este modelo, conforme mencionado no ponto 2.1.1, se revela robusto para uma análise sobre a morfologia concordante, sobretudo quando se trata de estudar padrões colectivos de comportamento linguístico afectados por mudanças linguísticas decorrentes de processos históricos específicos, como é o caso da transmissão linguística irregular.

Apesar de tudo, importa considerar que a Gramática Generativa formaliza as suas constatações, em termos de língua-I, recorrendo a unidades invariáveis e regras categóricas, ao passo que a Sociolinguística Variacionista faz uso de unidades e regras variáveis para análise dos padrões colectivos de comportamento linguístico. Além disso, ao procurar explicar o modo como a língua funciona e entendendo a língua-I como o resultado do desenvolvimento da faculdade inata da linguagem, o generativismo situa as suas observações sobre os fenómenos linguísticos num plano estritamente psico-biológico. Já a Sociolinguística Variacionista, ao lançar a sua atenção sobre a relação entre variação e mudança e operar com esquemas probabilísticos de análise, coloca os seus estudos sobre tais fenómenos num plano sociohistórico. Nesta perspectiva, o nosso modelo de análise será heterogéneo, observando o funcionamento da gramática em concomitância com factores sociais, culturais ou ideológicos envolvidos nos processos históricos de mudança linguística que afectaram, ou estarão a afectar, o PA. Assim, será também possível não só alargar o conjunto de respostas para entendimento das questões inerentes à CPL-var no SN do PA mas estabelecer igualmente quais os fenómenos que concorrem para a implementação (*actuation problem* – Weinreich, Labov & Herzog, 2006 [1968]) de determinadas variantes em detrimento de outras (cf. Lucchesi & Ribeiro, 2009:130) na comunidade de Almojarife. E, ao conjugarmos aspectos da teoria da gramática e análise sociohistórica da mudança linguística, estaremos, em última instância, a procurar contribuir para diminuir a distância teórico-metodológica que separa, actualmente, estas duas perspectivas de estudo.

#### **2.1.4. Variação e sistematicidade na *interlíngua*: aplicação da perspectiva heterogénea ao estudo do PA**

Como a fala de Almojarife encerra variação, adoptámos a metodologia da perspectiva sociolinguística para recolha de dados em contextos heterogéneos (Labov,

1972a). O tratamento dos dados seguiu ainda as propostas labovianas para análise quantitativa na variação linguística, também aplicadas por Dickerson (1974) e Dickerson (1976), para detecção dos contextos linguísticos em que acontece a variação, em primeira mão, e por Preston (1996b), para apurar a direcção em que acontece a mudança linguística, em segunda etapa (cf. Godinho, 2005:61).

A investigação de dados por Labov (1966, 1972a), na perspectiva da sociolinguística quantitativa, permitiu identificar quais as influências internas (p.e. a saliência fónica ou o contexto fonológico em que determinado som se realiza) e externas à língua (p.e. a classe social, a idade, o nível de escolaridade) que actuam sobre as variáveis fonológicas. Seguindo esta metodologia, poderemos, por exemplo, confirmar se a inserção ou apagamento do –s morféxico, marca de PL, nos itens do SN do PA terão subjacentes motivações fonológicas do mesmo tipo das que foram observadas por Labov (cf. Godinho, 2005:61).

A variação nos dados de um determinado falante pode ser livre ou sistemática, com a primeira reflectindo formas que resultam de motivações em competição conotadas aos sistemas de regras da L1 e da L2 (Ellis, 1994:144; Romaine, 2003:411, *apud* Godinho, 2005:62). De acordo com o primeiro autor, a variação livre ocorre por curto período de tempo, restringindo-se aos estados precoces do desenvolvimento linguístico. O falante começa por utilizar um traço de forma não sistemática, mas à medida que a interlíngua se desenvolve, a forma é integrada nesta e passa a ser aplicada em todos os contextos. Portanto, a variação livre vai desaparecendo, à medida que os falantes desenvolvem a sua interlíngua em direcção à LA (Preston, 1966:250, *apud* Godinho, 2005:62).

Como temos vindo a evidenciar, subjazendo à transição de variação livre para variação constante actuam factores situacionais, linguísticos e psicolinguísticos, o que determina que a observação das questões relacionadas com a variação livre e a variação condicionada não podem ser entendidas sem se lançar mão de um modelo teórico adequadamente formulado para tratar a variação e a mudança. O fenómeno da mudança linguística, sociolinguisticamente observado (cf. Weinreich, Labov & Herzog, 2006 [1968]), é também uma problemática da Gramática Generativa, apesar de esta o entender como um processo de aquisição da linguagem com restrições impostas pela GU. Assim, a questão da implementação (*actuation problem*) (Weinreich, Labov & Herzog, 2006

[1968]), que tenta determinar o porquê de acontecer mudança linguística, “*encontra uma resposta no programa gerativista em termos de cadeia de aquisição*”:

- (i) *a aquisição resulta de um processo seletivo e abduativo, de interação entre estruturas cognitivas e tipos de experiência selecionados;*
- (ii) *os dados da experiência relevantes para a aquisição são constituídos de elementos robustos e estruturalmente simples;*
- (iii) *a criança é um aprendiz grau-0 – sensível somente a dados de sentenças raiz e do complementador da sentença encaixada;*
- (iv) *a morfologia desempenha papel importante na definição do valor de um parâmetro;*
- (v) *parâmetros são marcados nos estágios iniciais da aquisição”* (Lucchesi & Ribeiro, 2009:140).

Nesta perspectiva, a reanálise abduativa dos DLP’s, resultante de processos morfofonológicos, ambiguidade estrutural, DLP’s inovadores, factores extralinguísticos ou contacto entre línguas, potencia a mudança sintáctica (Lightfoot, 1991a; Kroch, 2004; Roberts, 2007).

A problemática do encaixamento (*embedding problem*) (Weinreich, Labov & Herzog, 2006 [1968]) é observada sob dois ângulos em termos sociolinguísticos: (i) o modo como se encaixa na estrutura linguística; (ii) a forma como se encaixa na estrutura social. Relativamente à primeira visão, estudos de sintaxe diacrónica têm apresentado três possibilidades, estando a primeira conectada à noção de valor paramétrico (positivo ou negativo) e sua correlação com as hipóteses estruturais (Lightfoot, 1979; 1991a; Rizzi, 1982; Roberts, 2007). Nesta perspectiva, a mudança linguística não afectará a totalidade de um sistema linguístico (Weinreich, Labov & Herzog, 2006 [1968]:185), visto que as propriedades que podem variar entre línguas não são autónomas, antes se organizam em grupos estáveis.

No caso do parâmetro do sujeito nulo, por exemplo, Rizzi (1982) especifica propriedades que diferenciam o PE, o italiano e o espanhol de outras línguas, como o inglês ou o francês. As conclusões demonstram que, uma vez perdida a propriedade correspondente ao valor do parâmetro, as outras propriedades relacionadas com ele se apagam também.

Lightfoot (1979, 1991a), por seu lado, analisa a evolução de verbos plenos do inglês antigo, para verbos auxiliares modais no inglês médio. O autor conclui que a reanálise terá sido impulsionada por um conjunto de mudanças gramaticais distintas

Quanto a Roberts (2007), debruça-se sobre os aspectos que levaram o inglês a divergir, entre 1100 e 1700, quer das línguas românicas quer de outras línguas germânicas. A mudança de uma língua de um tipo para outra de outro tipo ficará a dever-se a uma mudança no seu sistema funcional, cujo efeito dominó activará mudanças paramétricas em cascata. Efectivamente, tais mudanças, ao afectarem os traços de um determinado núcleo funcional, têm como consequência uma extensa reorganização nos sistemas conotados ao mesmo.

No que concerne à problemática do encaixamento na estrutura social (*embedding problem*) (Weinreich, Labov & Herzog, 2006 [1968]), a mesma pode levar a mudanças linguísticas na fala dos adultos, que alteram os DLP's para a aquisição da L1 de uma geração para a outra e afectam algoritmos generativos. Nesta conformidade, a mudança é activada pela alteração nas frequências de uso de uma determinada forma, como acontece na variação que afecta a concordância entre os elementos do SN no PtgL2 adquirido pelos almoxarifanos, e que alteram os dados robustos para aquisição do PA como língua materna pelas gerações precedentes. Este tipo de alterações sociohistóricas não é ocasionado pela gramática em si, mas sim pela mudança de hábitos linguísticos associada a mudanças comportamentais a nível de relações sociais, culturais ou ideológicas. Contudo, poderão ter impacto na gramática ao determinarem também mudanças paramétricas, como o comprova o caso da substituição do pronome pessoal canónico *vós* por *vocês* no PE, que adoptou a forma verbal da 3ª pessoa do plural, reduzindo o paradigma flexional do verbo (p.e. *vós ides* > *vocês vão/eles vão*).

Kroch (2001, 2004), ao propor que se observem os efeitos gerados pelas gramáticas em competição numa determinada comunidade de fala, “*permite conciliar a concepção de uma mudança gramatical abrupta com uma gradualidade que resulta da distribuição social dos usos linguísticos*” (Lucchesi & Ribeiro, 2009:143-144). Efectivamente, a possibilidade de os falantes almoxarifanos poderem optar por estruturas disponíveis nas gramáticas em concorrência na sua comunidade, sobretudo por esta ter na sua génese



uma situação de contacto entre línguas, foi fundamental para a distribuição dos seus usos linguísticos.

Face a estes aspectos, resulta que, para além da adopção de um apropriado modelo teórico, a observação da variação e mudança nos sistemas linguísticos não é possível se não se recorrer também a um dispositivo estatístico adequado para avaliação da mesma. Desta forma, a variação sistemática pode ser analisada recorrendo a variáveis dependentes e independentes, com as primeiras representando o conjunto das variantes linguísticas (os itens em variação) e as segundas dizendo respeito aos prováveis grupos de factores que condicionam as variáveis dependentes. No nosso caso, a variável dependente em análise representa a ausência ou inserção de marca morfológica de número PL nos itens do SN. Por seu lado, as variáveis independentes identificam o conjunto de factores linguísticos e extralinguísticos que beneficiam ou inibem o uso dessa mesma marca. Sujeitando à análise quantitativa os factores prováveis de exercerem influência sobre a variação, se se verificar que os mesmos não têm efeito sobre as variáveis dependentes, então não está configurada uma situação de variação sistemática (cf .Preston, 1996:250, *apud* Godinho, 2005:62).

A CPL-var no SN do PA configura variação sistemática, visto ser um fenómeno abrangente em termos de estratos etários ou de sexo. Mas, esta sistematicidade pode também ser determinada por outros contextos, que não os meramente sociais, como se pode constatar em exemplos do nosso *corpus*, mais concretamente os que dizem respeito à variação que afecta a inserção de PL no nome “trabalhadores”:

[153] PA:  
*não tem possibilidade de pagar os trabalhador antão Banco Mundial*  
[ABILH2]

[154] PA:  
*não tem possibilidade de pagar os trabalhadores principalmente o Stado*  
[ABILH2]

[155] PA:  
*os trabalhadô stão co sacrificia no mato pa dá produção* [CELESH3]

[156] PA:  
*cinquenta metros quadrado daquela terra enquadrá os trabalhadores.*[CASTEH1]

[157] PA:  
*para recolher alguns trabalhadores, os empregados pra levar*  
[CASTEH1]

Em contextos linguísticos idênticos, os falantes aplicam marcação PL diferente, pelo que a explicação para esta terá que ser fornecida em função de motivações sociais, como os anos de escolaridade, e/ou psicolinguísticas, como nos casos em que o falante monitoriza o *output*, após ter prestado atenção às realizações linguísticas do seu interlocutor (Wolfson, 1976:203). Portanto, subjazendo à variação estarão a actuar factores de natureza distinta, isoladamente ou em concomitância (Ellis, 1994:120), pelo que a sua observação terá que ser efectuada de modo heterogéneo. Deste modo, a codificação dos nossos dados, que adoptou a perspectiva variacionista amplamente utilizada em estudos brasileiros, permitiu que constatássemos também coincidências com a matriz proposta por Ellis (1994). Observar a variação numa perspectiva interna à própria língua significa, pois, levar em conta fenómenos que são idênticos em aquisição de L1 e L2 e que se podem conotar às teorias sobre a marcação e aos universais linguísticos. Será o caso, por exemplo, de ocorrer marcação apenas no elemento que é entendido como semanticamente relevante (p.e. *os home branco daqui ficaram* [CELESH3]), um fenómeno que terá respaldo nas línguas africanas (cf. Guy, 1981a:301-302) do grupo níger-congo atlântico (Baxter, 2009:292). Ainda a propósito deste tipo de marcação, Kiparsky (1972:195) advogou a hipótese funcionalista de que acontece uma tendência para os falantes reterem a informação relevante na estrutura de superfície (Condições de Distintividade), evitando o apagamento do morfema de PL apenas nas situações em que ele acarreta a informação semântica da pluralidade (Young, 1993:78).

Outras ocorrências em que se regista variação sistemática poderão estar ligadas às tendências universais detectadas em fonética, como acontece com a não-realização de consoantes finais (Romaine, 2003:413, *apud* Godinho, 2005:63), caso do apagamento do /r/ em final de palavra, um fenómeno vulgar no PB e também no PA (exemplo [155]). Lembremos, a este propósito, que determinadas línguas do grupo níger-congo atlântico, como o kimbundu, para além de não registarem a ocorrência da consoante /r/, apresentam, regra geral, uma vogal em final de palavra. Por outro lado, estes dialectos apenas apresentam vogais foneticamente salientes, não realizando as vogais /a/ [- alt], /e/ [+ alt] e /o/ [+ alt] (Mingas, 2000:36). Fenómenos deste tipo poderão ter tido

influência directa na realização de vogais foneticamente salientes e estruturas do tipo dissilábico CVCV registadas no santomense, apresentando estas uma tendência para a simplificação, quando comparadas com as do português (Ferraz, 1979:26). Assim sendo, tanto o falante de uma língua africana do grupo níger-congo atlântico em situação de aquisição do PtgL2 como a criança em contexto de aquisição do PtgL1 nativizado poderão revelar tendência para perceberem as formas mais salientes, acabando, no caso da pluralização, por inserir marcas nos itens de acordo com o advogado no Princípio da Percepção da Saliência (*Principle of Perceptual Saliency*), que postula que quanto mais distantes forem as formas fonéticas, maior é a probabilidade de elas surgirem marcadas (Wolfram, 1989:189).

De acordo com as Condições de Distintividade, é então de prever que, no caso do PA, a inserção de /s/, marca de número PL, não venha a ocorrer em nomes que sejam antecidos por elementos que já contenham a informação de PL, seja ela semântica (p.e. numerais) seja ela estrutural; em contrapartida, se a informação de PL ainda não tiver ocorrido, isto é, se o nome surgir como primeiro elemento do SN, então é provável que ele seja pluralizado (cf. Godinho, 2005:64). Paralelamente, também é de esperar que os elementos que apresentem maior saliência fónica na oposição singular/plural (p.e. *ovo* > *ovos* ou *patrão* > *patrões*) sejam mais marcados do que os itens em que tal saliência não seja tão óbvia (p.e. *coisa* > *coisas* ou *metro* > *metros*), como sucede quer nas variedades brasileiras (Scherre, 1988; Lopes, 2001; P. Andrade, 2003) quer nas variedades africanas de português (Baxter, 2004, 2009; Figueiredo, 2008, 2009b; Jon-And, 2008, 2009). Contudo, este tipo de marcação tem também sido conotado aos princípios universais (Romaine, 2003:413). Sobre a possibilidade de determinada estrutura da L2 representar ou não uma transferência da L1, voltaremos a pronunciar-nos no ponto 2.4.2.2 do presente trabalho, nomeadamente quando abordarmos as questões concernentes à Teoria da “Processabilidade” (Pienemann, 1998, 2005b).

A observação diacrónica da variação e da sistematicidade na fala da comunidade, verificadas através do chamado tempo aparente, ajuda a detectar possíveis mudanças em curso na mesma. Segundo Ellis (1994:144, *apud* Godinho, 2005:64), o aprendente de uma L2 começa por utilizar, de modo categórico, o mesmo traço para todos os contextos. É o que sucede, por exemplo, com o informante que produziu o exemplo

[155], que nunca pluraliza o nome “trabalhador” no *corpus* que possuímos, como pudemos constatar. Contudo, na geração seguinte vamos encontrar formas marcadas e não-marcadas (exemplos [153] e [154]), o que denota que a variação se torna livre, à medida que os membros mais jovens e mais escolarizados da comunidade vão aprendendo novas formas do mesmo nome. O efeito exercido por determinados factores sociolinguísticos leva ainda a que, posteriormente, a terceira geração evidencie uma maior aquisição da regra, passando a controlar a marcação de modo consciente, sistemático e correcto, conforme o demonstram os exemplos [156] e [157], produzidos pelo informante [CASTEH1] (8º ano de escolaridade, cf. Tabela 8.1), para o qual não foram detectadas formas do nome “trabalhador” sem marca PL, quando esta é exigida.

Trabalhos levados a cabo por Dickerson (1975, *apud* Godinho, 2005:64) ou Plag (2008a, 2008b) demonstram que a ASL é um processo de aprendizagem contínua, apresentando etapas flutuantes e transitórias, mas caracterizadas por sistematicidade. Factores linguísticos e extralinguísticos actuam, desenvolvendo a gramática do falante em direcção ao sistema de concordância da LA (Romaine, 2003:416), de acordo com o observável nos exemplos [153] e [154]. Portanto, quando há consciência da regra de concordância, entram em jogo fenómenos psicolinguísticos, podendo o falante monitorizar o uso das formas. No caso concreto dos informantes do PA, embora se tenha tentado afastar pressões do tipo social que pudessem interferir no seu registo espontâneo, nunca é de descartar a possibilidade de, no decurso das entrevistas, eles terem adoptado diferentes estratégias de comunicação, consoante a pragmática o exigisse. Porém, consideramos que, no cômputo geral, os discursos se adequam às exigências de uma correcta análise em sociolinguística quantitativa para o grau de sistematicidade registado na fala da comunidade (cf. Godinho, 2005:65).

## **2.2. Abordagens teóricas sobre aquisição de L1**

A aquisição de L1 pressupõe, em primeira instância, a aquisição da L1 por falantes monolingues. Contudo, o próprio conceito de monolingüismo puro é visto, à luz da teorização actual, como algo fictício, já que qualquer falante pode modelar e modular o seu registo, adaptando-o às exigências pragmáticas (Berg & Schade, 1992:418). Desta forma, o monolingue recorre a “falas” distintas, caso, por exemplo, esteja a comunicar

com uma criança ou se encontre no tribunal, perante um juiz. Noutras situações, pode mesmo misturar formas ou elaborar paráfrases, usando, para tanto, palavras ou registos linguísticos bastante distintos entre si. No presente trabalho, independentemente da polémica em torno do conceito de monolingüismo, interessa mais procurar entender como se processa não só a aquisição da L1 da criança que está igualmente a tentar comunicar com o adulto mas também o modo como essa aquisição se desenvolve até ao estágio em que ela, depois de atingir a idade adulta, vai fazer uso da recursividade, isto é, vai utilizar o *input* restrito e finito que recebeu para produzir *outputs* infinitos e conscientemente adaptados às diferentes situações de uso com que se vier a confrontar (Hornstein & Lightfoot, 1981:202).

### **2.2.1 Modelo generativista**

Ao elaborar a teoria do inatismo sobre a aquisição da linguagem, Chomsky (1965, 1972) defende que esta capacidade se desencadeia de forma involuntária em todas as crianças, concluindo-se em espaço de tempo breve,<sup>49</sup> mas envolvendo uma grande complexidade de operações. Apesar de existirem diferenças entre as línguas aprendidas, o processo de aquisição actua de forma semelhante, permitindo que todas as crianças, a partir de dados limitados, e com o suporte da GU, não só construam a gramática da língua mas activem também a capacidade da recursividade, isto é, a faculdade da linguagem. Assim sendo, a recursividade torna-se candidata a característica nuclear da linguagem humana, assentando os seus pressupostos quer na Teoria dos Princípios e Parâmetros (Chomsky & Lasnik, 1999 [1993]) quer na Teoria das Regências e Ligações (Chomsky, 1981b) quer ainda em preceitos avançados em posteriores trabalhos de Chomsky (1986a, 1986b). No caso concreto da GU, actua rapidamente, permitindo a aquisição da competência linguística em espaço de tempo muito curto. Tal só é possível porque ela está dotada de propriedades universais formais (referem-se às condições abstractas a que uma gramática obedece para combinar unidades linguísticas), substantivas (apontam para as categorias universalmente presentes nas línguas, isto é, de substância visível e necessárias para se analisar a língua, como acontece com os traços distintivos, nomes, verbos, sílabas singular/plural, masculino/feminino, etc.), implicativas (determinam a existência de uma correlação entre propriedades universais,

como acontece, por exemplo, nas línguas do tipo SVO, que necessitam obrigatoriamente de recorrer ao uso de preposições), ou pertencentes a determinada área da gramática (por exemplo, os universais fonológicos, semânticos, etc.). Como estas categorias universais se manifestam de modo diferenciado nas diversas línguas, a GU tem papel primordial na determinação (e explicação) das propriedades da gramática do ser humano, enquanto a Gramática Generativa tem como principal objectivo o estudo da natureza e propriedades da GU. Tendo em conta estes aspectos, percebe-se o motivo que leva os investigadores que trabalham dentro destes pressupostos a optarem, como objecto de estudo, por descrições linguísticas de gramáticas.

### **2.2.2. A teorização em torno da GU**

#### **2.2.2.1. O Modelo de Princípios e Parâmetros (*Parameter Settings*)**

Estando advogadas as propriedades universais da linguagem humana, Chomsky (1981a), avançou com a Teoria dos Princípios e Parâmetros para tentar descrever a diversidade das línguas. O modelo, composto por princípios extremamente gerais, visava, inicialmente, explicar as diferenças sintácticas entre as línguas, numa dimensão puramente sincrónica. Deste modo, advoga-se que a gramática se organiza em módulos simples e autónomos, mas que abrangem diferentes domínios da linguagem, em virtude de manterem uma rede de interações com as outras componentes desta. Consequentemente, o modelo possibilita que se cruzem estudos puramente gramaticais sobre a estrutura das línguas quer com observações psicolinguísticas sobre aquisição e desenvolvimento da linguagem quer com estudos de gramática comparativa. Com base nos pressupostos deste modelo, a GU contém dois tipos de princípios:

- (i) os princípios rígidos e invariáveis (os universais substantivos e formais), como o Princípio de Projecção, o Princípio de Dependência Estrutural ou o princípio de que todas as línguas possuem SN sujeito e SV predicado (Chomsky, 1965: 27-30);
- (ii) os princípios abertos ou parâmetros, não especificados, e que se vão configurar para um valor num sistema binário (Chomsky, 1981a, 1986b).

Os parâmetros funcionam como “comutadores linguísticos”, cujo valor final é alcançado durante o processo de aquisição através da fixação numa das posições

possíveis, com base na informação colhida no meio linguístico ambiente. Desta forma, a gramática final da língua adquirida é caracterizada por dois aspectos: (i) a aprendizagem lexical da língua, sendo as propriedades fonológicas, sintáticas e semânticas guiadas pelo “dicionário mental”; (ii) os parâmetros da GU, com valores que lhes são atribuídos pelas características da língua.

Os parâmetros pertencem ao âmbito da variação, que tanto pode ser lexical, isto é, aprendida de forma explícita, como do domínio da aprendizagem não-explícita. Fornecem as opções possíveis à GU, bastante limitadas em virtude de serem finitas e binárias: o “parâmetro da direccionalidade” e o “parâmetro do sujeito nulo”. O “parâmetro da direccionalidade” é fixado muito cedo e permite que as crianças não errem ao juntar verbo e objecto na estrutura da produção. Como tal, é o parâmetro que permite estudar as configurações das línguas OV ou VO. Por seu lado, o “parâmetro do sujeito nulo” concede à criança a possibilidade de detectar se a língua que é falada à sua volta admite ou não construções com sujeito nulo.

A ligação entre todos os parâmetros concede uma “gramática nuclear” (*Core Grammar*) à criança, composta por uma complexa rede de conexões entre os princípios rígidos (universais) e os parâmetros, acabando, assim, por determinar as propriedades específicas de cada língua. Por conseguinte, a escolha dos parâmetros não é orientada conscientemente pela criança, correspondendo antes à aquisição da linguagem a partir dos enunciados que recebe da comunidade (os DLP’s), aquisição essa sujeita a um percurso de maturação da GU, que se inicia num estado em que os parâmetros não estão ainda fixados e termina num patamar com parâmetros fixados através de um dispositivo computacional. Em resumo: a GU, para além de definir as propriedades universais da linguagem humana, vai também fixar os parâmetros em determinados valores, originando a gramática particular de cada língua, isto é, a variação entre línguas.

Assim, na perspectiva do Modelo de Princípios e Parâmetros, a mudança sintáctica não é mais do que uma mudança no valor de parâmetro, Efectivamente, no estado inicial da aquisição (FL0),<sup>50</sup> o parâmetro é fixado a partir de dados simples e facilmente acessíveis, isto é, a partir dos DLP’s que a criança tem ao seu alcance para determinar a posição a conceder na referida fixação. No caso dos princípios rígidos, por exemplo aquele que determina que todas as línguas possuem SN sujeito e SV predicado, o

“parâmetro do sujeito nulo”, que vem por defeito com o valor positivo (a língua tem sujeito nulo), possui duas opções: (i) a língua tem sujeito nulo; (ii) a língua não tem sujeito nulo. Estas opções vão possibilitar que a criança escolha entre a realização fonética obrigatória do sujeito (valor *não-pro-drop*) ou realização fonética opcional do sujeito (valor *pro-drop*), fixando aquela que corresponde ao paradigma da sua língua específica. Nas situações que configuram princípios abertos, em que as crianças acedem a frases gramaticais pouco frequentes ou “estranhas”, a fixação pode não corresponder à dos parâmetros efectivos da GU. Ainda assim, em certas circunstâncias, a não ocorrência de determinadas estruturas nos DLP’s pode levar, durante o processo de maturação da GU, à fixação de parâmetros em determinada posição, sem necessidade de se recorrer a correcções ou instruções explícitas (Chomsky, 1981a).<sup>51</sup>

#### **2.2.2.1.2. Teoria das Regências e Ligações**

À luz da Teoria das Regências e Ligações (Chomsky, 1981a, 1986a), cada língua é um procedimento generativo, responsável pela produção de *descrições estruturais* (DE’s) compostas de propriedades semânticas e fonológicas. Deste modo, os parâmetros encontram-se localizados no léxico, no qual, por sua vez, estão alojadas as categoriais lexicais e funcionais das línguas. As primeiras representam os morfemas de conteúdo (categorias como nomes e verbos) e têm papel activo no quadro temático da produção linguística, atribuindo o valor semântico a esta. Como tal, configuram-se como núcleos da sua própria projecção lexical: o SN, o sintagma verbal (SV), etc. Estas categorias encerram em si as propriedades responsáveis pela definição das categoriais lexicais das palavras (p.e. [+V] e [+N]).

Por seu lado, as categoriais funcionais dizem respeito à morfologia flexional que contém a informação gramatical de pessoa, número, concordância, caso, negação, tempo ou aspecto. Portanto, para além de possuírem características semânticas e de flexão, as categoriais funcionais controlam também as formas morfofonológicas que correspondem à realização destas (p.e. *-s*, para marca de flexão em número PL, e *-a*, para marca de flexão em género feminino). Quanto à informação gramatical contida nas categoriais funcionais, é codificada de acordo com o tipo a que diz respeito, por exemplo: FLEX



(categoria funcional da flexão), CONC (concordância) ou T (projecção máxima de informações de tempo) (Godinho, 2005:99).

O dispositivo computacional que determina a fixação dos parâmetros extrai as categorias funcionais do léxico e transforma-as em representações sintácticas abstractas. Deste maneira, na hierarquização da estrutura frásica são as categorias funcionais que dominam as categorias lexicais (o SN é dominado pelo SDET, enquanto o SV é dominado por todas as categorias acima da categoria ASPECTO). Quando o dispositivo computacional extrai as categorias funcionais do léxico, acciona operações que as levam da área dos itens lexicais para o campo dos elementos funcionais (operações de “movimento”), a fim de entrarem em relação próxima com as outras categorias e permitirem que os seus traços interpretativos efectuem o mapeamento dos correspondentes traços não-interpretativos, originando o pareamento entre eles (operação de “pareamento”),<sup>52</sup> isto é, a concordância entre os elementos que constituem a frase (Montrul, 2004:8; Godinho, 2005:100).

Na Teoria das Regências e Ligações, os parâmetros são propriedades abstractas da GU e não estabelecem correspondências biunívocas com propriedades simples (as que são directamente observáveis) do sistema linguístico. Assim, a fixação do parâmetro é estabelecida pela interacção com outros princípios da GU e determina, por sua vez, feixes complexos de propriedades concretas em diversos pontos da gramática. Desta forma, o dispositivo computacional e as características formais das línguas constituem universais, o que não acontece com a edificação das frases, dada a sua dependência do léxico (Godinho, 2005:100).

Pode concluir-se que, em termos gerais, o pressuposto básico da teoria generativa é o de que a língua possui léxico e gramática (o dispositivo computacional). Ao segundo, que é inato, compete gerar as expressões ao nível da representação linguística (as DE's), sendo o léxico, por seu lado, responsável pela caracterização dos itens lexicais que surgem nas DE's (Chomsky, 1996 [1995]:59, *apud* Godinho, 2005:100). Como o léxico (as palavras) não é inato, tem que ser aprendido, necessitando a criança de adquiri-lo, tal como acontece com os parâmetros da sua L1 (a morfologia), já que ambos representam o conjunto das propriedades sintáctico-semânticas abstractas das categorias funcionais (Montrul, 2004:10, *apud* Godinho, 2005:100). Por sua vez, a aquisição de uma L2

significa a aquisição e fixação dos parâmetros da LA, já que os princípios são universais, não necessitando de ser aprendidos (Godinho, 2005:101).

Desta forma, a sintaxe trabalha mais com o léxico, levando a que o falante interprete determinados enunciados, por exemplo “*gostar de*” e “*amar*”, como sinónimos. Contudo, a sintaxe também estabelece relações com a componente interpretativa das palavras (a semântica), a prosódia (a fonologia) e a estrutura informacional (p.e. topicalizações, em que o primeiro discurso determina a escolha do enunciado posterior). Para dar conta da maneira como as diferentes componentes gramaticais interagem entre si, foi concebido o Modelo-T (modelo derivacional – Fig. 9), bastante rico em fontes possíveis de variação, e que tem o léxico como ponto de partida de toda a derivação, sendo as outras componente gramaticais alimentadas por ele (Chomsky & Lasnik, 1999 [1993]). Assim, as DE’s não serão mais do que representações em diferentes níveis da estrutura da língua (Estrutura-P, Estrutura-S, Forma Lógica (FL) e Forma Fonológica (PF)):

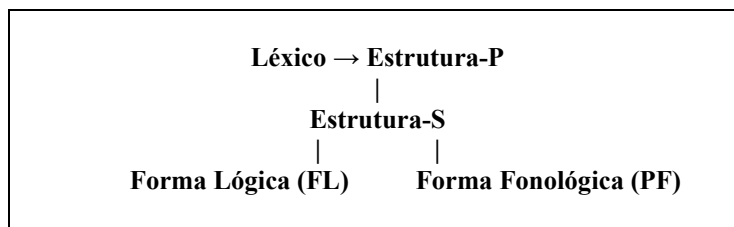


Fig. 9. Modelo de Princípios e Parâmetros (a partir de Chomsky, 1996 [1995:22]).

De acordo com o esquema, a Estrutura-P representa a projecção directa da estrutura argumental, isto é, a estrutura profunda, enquanto a Estrutura-S diz respeito ao nível sobre o qual incidem as regras de interpretação fonológica e de interpretação semântica (a *performane*), ou seja, a estrutura de superfície. O que se passa entre a Estrutura-P e a Estrutura-S é a aplicação das regras sintácticas (essencialmente as regras de movimento), sujeitas a variação paramétrica. Por seu lado, a estrutura de superfície, enquanto sistema de *performance*, alimenta duas vias: (i) a componente fonológica (o módulo articulatorio-perceptual); (ii) a componente lógica (o módulo conceptual-intencional). Como a prosódia pode interferir na interpretação do enunciado, a PF constrói os seus constituintes a partir da sintaxe, mas precisa também dos seus próprios constituintes para elaborar a hierarquia prosódica, uma vez que a sintaxe não lhe

concede todos os elementos de que necessita (Nespor & Vogel, 2007:32). Já em termos semânticos, ocorrem também movimentos não explícitos (subjacências) entre a Estrutura-S e o módulo FL, a fim de se evitar a ambiguidade (Huang, 1982:63; May, 1985:97; Nespor & Vogel, 2007:265). Assim, e apesar de o Modelo-T apresentar vários instrumentos de análise e dos domínios de variação paramétrica, esta não é satisfatoriamente explicada por ele. Paralelamente, também a relação da sintaxe com outros domínios gramaticais, como a morfologia ou a escolha da ordem das palavras com vista a condicionar a estrutura informacional, não se encontra plenamente contemplada no Modelo de Princípios e Parâmetros. Voltaremos a esta questão no ponto 2.6.2.1 do presente trabalho, no qual faremos alusão aos modelos desenvolvidos a partir do Modelo de Princípios e Parâmetros, nomeadamente o Programa Minimalista (Chomsky, 1996 [1995]).

### 2.2.3. Papel da GU na aquisição da L1

Ao partir do princípio que a aquisição da linguagem se processa interna e rapidamente, a linguística generativa entende que a GU é o órgão biológico (o dispositivo computacional) que não só caracteriza o estado inicial da faculdade da linguagem (FL0) mas orienta também todo o processo de aquisição de uma L1 até ao seu estado final (FLf), isto é, até à fixação da gramática do adulto. Deste modo, o Dispositivo de Aquisição das Línguas (*Language Acquisition Device*) inicia-se no estado zero (FL0), completando-se no estágio estável do conhecimento da língua (FLf) (Chomsky, 1965), o que nos permite elaborar a seguinte estrutura:

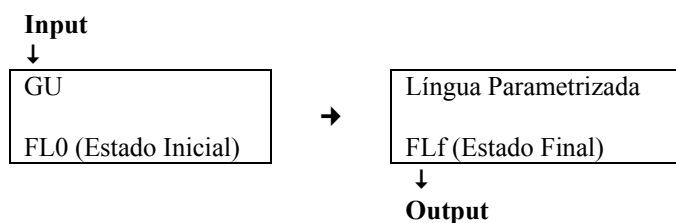
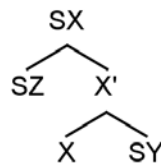


Fig. 10. Modelo de aquisição das línguas.

No dispositivo computacional encontra-se o estado inicial da aquisição (FL0), dotado de princípios rígidos e uma série finita de parâmetros, pelo que:

*“todas as funções a operar vão depender de elementos funcionais e propriedades gerais do léxico. A selecção  $\Sigma$  de opções determina uma língua. Assim, a aquisição de uma língua será o fixing de  $\Sigma$ , i.e., a Gramática da Língua descreve um  $\Sigma$ . As representações nada mais vão ser do que motivadas a condições, por propriedades de interface. As expressões linguísticas vão ser as realizações óptimas das condições de interface. Estas expressões linguísticas serão o objecto formal que satisfaça as condições de interface, da melhor maneira”* (Chomsky, 1996 [1995]:168).

Para representar o “programa” do dispositivo computacional responsável pelas modificações da estrutura projectada a partir do léxico, Chomsky (1996 [1995]:172) propôs a Teoria X-barras, que advoga que são os núcleos projectados que vão ter a principal intervenção nas relações básicas estabelecidas a nível do sistema linguístico, sendo estas relações tipicamente locais (Fig. 11):



**Fig. 11.** Estrutura X-barras.

As relações fundamentais estabelecidas entre os núcleos são então:

- (i) A relação entre SZ e X, isto é, a relação entre Especificador (SPEC) e Núcleo, representando X as restantes categorias;
- (ii) A relação de irmandade a nível temático entre núcleo e complemento, ou seja, entre X e SY.

Como as regras são recursivas, podem combinar-se e gerar um número infinito de estruturas sintáticas. Assim, é possível não só formarem-se estruturas genéricas, tipo SX (Fig. 11), como elaborarem-se também estruturas com múltiplos complementos, árvores sintáticas mais pequenas que omitam partes opcionais, ou estruturas com nós adicionais de SX's ou X's de vários tipos. O diagrama da Fig. 12 ilustra um dos possíveis modos como as regras se combinam para formar uma árvore sintática de estrutura genérica SX:

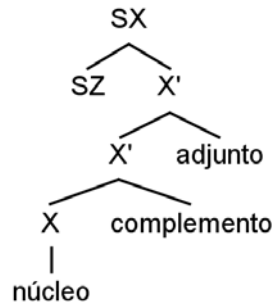


Fig. 12. Estrutura alternativa gerada a partir da estrutura X-barra (a partir de Carnie, 2007 [2002]:118).

Note-se que o complemento (COMP) contendo X' pode ser distinto de um adjunto (ADJT) contendo X', visto o complemento ter X (núcleo) como irmão, enquanto o adjunto possui X' como irmã.

Independentemente das questões relativas à forma de representar o “programa” do dispositivo computacional responsável pelas modificações da estrutura projectada a partir do léxico, é importante reter que a GU actua por fases de desenvolvimento, que se assemelham às dos domínio anatómico, fisiológico ou mental (Chomsky, 1965). No entanto, a aquisição da língua é um processo impreciso, daí nada garantir que os diferentes aprendentes, a partir dos mesmos DLP's, estruturam internamente a mesma gramática (Kroch, 2004:1). De acordo com o modelo generativista de aquisição de L1's, a transmissão/aquisição linguística ocorre segundo o seguinte esquema:

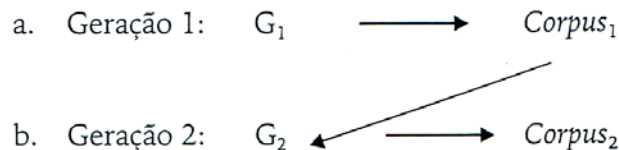


Fig. 13. Modelo de transmissão/aquisição de L1 (Lucchesi & Ribeiro, 2009:144)

Como a criança (G<sub>2</sub>) só tem acesso a DLP's a partir do ambiente linguístico em que está inserida, isto é, à língua-E (*Corpus*<sub>1</sub>) e não ao conjunto de parâmetros valorados para determinadas propriedades morfosintáticas que configura a língua-I (G<sub>1</sub>), vai elaborar sistemas gramaticais hipotéticos e intermédios (*Corpus*<sub>2</sub>), traduzidos em desvios e variação, até atingir um FLf, mais ou menos rígido (a gramática do adulto), e que lhe permitirá gerar as produções linguísticas (*output*) adequadas às situações de

comunicação (Montrul, 2004:4). Ainda assim, o desenvolvimento da linguagem da criança guia-se por uma ordem que vai do genérico para o específico, começando na fase pré-linguística, mas tendo uma progressão bastante rápida. Depois de ultrapassadas as fases dos sons controlados por estímulos<sup>53</sup> e do balbúcio, começa a articular as primeiras palavras<sup>54</sup>, geralmente monossilábicas e com a forma CV, por volta do primeiro ano. O desenvolvimento gramatical é nulo, mas as palavras demonstram que a criança já percebeu que os sons se relacionam com os significados. A partir dos dezoito meses emerge a fase das duas palavras, que revelam não só a existência de relações semânticas e sintáticas mas também entoação sem pausas. Estas proto-frases caracterizam-se pelo recurso a nomes, que constroem frases do tipo SO ou para demonstrar a relação sujeito-lugar. A criança ainda não utiliza os pronomes nem recorre a marcas sintáticas e morfológicas, mas algumas das palavras apresentam flexões, isto é, denotam ter-se iniciado a aquisição gramatical. Após começar a encadear mais do que duas palavras para transmitir mensagens principais, a criança rapidamente inicia a construção de frases e, com dois anos e meio de idade, em situação de desenvolvimento normal, produz já enunciados semelhantes aos dos adultos nativos, revelando direcionamento para os usos da LA, traduzido em domínio flexional nas construções gramaticais e aplicação das estruturas básicas da L1 (Cerqueira, 1994; Faria, Freitas & Miguel, 1998, 2001; Freitas & Miguel, 1998; Costa & Santos, 2004 [2003]; Ferrari-Neto, Corrêa & Augusto, 2005; Castro & Ferarri-Neto, 2007). À medida que vão adquirindo mais linguagem, complexificam as construções gramaticais e, cerca de dezoito meses depois, a variação é esporádica e os enunciados apresentam uma gramática já bastante próxima da dos adultos. Nesta fase, as realizações linguísticas das crianças já não são imitações da fala dos adultos, mostrando que está completado o percurso da aquisição que possibilitará à criança passar a funcionar como um “*telégrafo para o infinito*” (Fromkin & Rodman, 1993 [1974]:357).

No início da aquisição da L1, a GU disponibiliza todos os elementos e possibilidades gramaticais que possui, mas com a exposição ao *input* específico de determinada língua, que leva a criança a testar, em fases sucessivas, as hipóteses de utilização do mesmo, ela apreende as regras que a levam a usar a língua adequadamente. A variação registada, que produz erros próprios do processo de desenvolvimento da linguagem, afecta as

partes não-substantivas do léxico e as propriedades gerais dos itens lexicais, sendo, por norma, ultrapassada relativamente cedo. Pelo contrário, na variação linguística referente aos aprendentes de uma L2, os erros mantêm-se durante tempo prolongado e podem mesmo nunca vir a ser corrigidos, fossilizando-se. Deste modo, as línguas não são mais do que subconjuntos particulares da GU, que governa todos os seus estágios de aquisição (Godinho, 2005:97), consoante as necessidades específicas em que aquelas são utilizadas.

Partindo destes princípios, os estudos generativistas têm centrado a sua atenção em dois objectivos: por um lado, perceber e avançar explicações acerca das propriedades universais da faculdade da linguagem, isto é, determinar os princípios abstractos que concedam justificação para as representações fonológicas e sintáctico-semânticas da linguagem; por outro lado, entender e explicar como se processa a variação sistemática entre línguas registada durante o processo de desenvolvimento da aquisição. Assim, os linguistas avançaram com quatro hipóteses, listadas por Godinho (2005:102):

1. A teoria continuísta (*Continuity Hypothesis*), que entende que as gramáticas das crianças e dos adultos são em tudo semelhantes (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996);
2. A teoria não-continuísta, que advoga que a gramática dos adultos é acentuadamente diferente da das crianças (Slobin, 1973, 1985);
3. A teoria da maturação (*Maturation Hypothesis*), que entende que os princípios da GU actuam logo a partir dos primeiros enunciados da criança, a que se seguem, mais tardiamente e de modo descontínuo, outros aspectos operativos (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994; Meisel, 1994a, 1997).
4. A teoria da aquisição gradual, que se aproxima da anterior, mas postulando que a construção gramatical se dá faseadamente, a partir da aquisição de novos itens lexicais e morfológicos (Montrul, 2004).

Com base nestes pressupostos, muitos têm sido os estudos levados a cabo para se perceber como se processa a aquisição das diferentes L1's e/ou L2's, em contextos de monolingüismo, bilingüismo ou plurilingüismo. No caso concreto da aquisição do

PtgL1, poder-se-ão apontar como relevantes, entre outros, os trabalhos de Cerqueira (1994), Faria, Freitas & Miguel (1998, 2001), Costa & Santos (2004 [2003]), Ferrarri-Neto, Corrêa & Augusto (2005) e Castro & Ferrarri-Neto (2007), enquanto que, para o PtgL2/PtgLE, podem ser indicados as pesquisas de Fransceschina (2002), Godinho (2005, 2008), Matos (2008), Baxter (2009) e Silva (2010).

O modelo abstracto de parâmetro permite não só que a Gramática Comparativa reduza o feixe de diferenças superficiais entre dois sistemas linguísticos a um único parâmetro mas também que a Gramática Histórica passe a conceber mudanças em termos da alteração de valor de um ou mais parâmetros em determinado sistema linguístico. Nesta perspectiva, o nosso trabalho segue o pressuposto de que a GU está subjacente a todo e qualquer contexto de aquisição de L1. Contudo, a própria tradição racionalista entende que o desenvolvimento das capacidades linguísticas, visto apenas sob a perspectiva de *“uma pseudogramática generativa reduzida a uma mera análise das expressões em árvores, representa uma posição de efeitos absolutamente nefastos e revela uma incompreensão total tanto dos fundamentos básicos da gramática generativa como do processo de aquisição e desenvolvimento da linguagem da criança”* (Raposo, 1992:36-37). Portanto, factores sociais, como o meio ambiente, a idade da aquisição, a estabilidade emocional e até a saúde física do falante, têm também influência na parametrização, permitindo aos estudos variacionistas fazerem previsões acerca do realinhamento das propriedades sintácticas dos dialectos sujeitos a contacto linguístico (Kato & Tarallo, 1989:37).

#### **2.2.4. GU e estudos de aquisição**

Tanto a Teoria dos Princípios e Parâmetros como a Teoria das Regências e Ligações constituem a base actual dos estudos sobre aquisição, já que fornecem um quadro teórico bastante fundamentado para o entendimento do modo como esta se processa. De acordo com estes princípios, os dados do *input*, mesmo que não constituam uma aprendizagem explícita, interagem com os princípios (as propriedades) e parâmetros da GU, “engatilhando” nos primeiros para activar os segundos. Portanto, a gramática é fixada, permitindo que a língua reflecta a interacção entre o mecanismo genético humano e o ambiente sociolinguístico aquando da aquisição (Lightfoot, 1999a:51). Como o produto



final – a competência linguística –, representa a conjugação do processamento dos DLP's<sup>55</sup> com o inatismo do indivíduo para a faculdade da linguagem, as características do *input* que a criança recebe são de importância vital para a construção do FLf (Godinho, 2005:91).

A fixação da gramática complexa é uma realidade na fase adulta, independentemente da pobreza dos estímulos, podendo estes ser condicionados pela pressão de factores extralinguísticos aos quais os falantes estão sujeitos: estatuto social e profissional, idade, escolaridade, contacto com outras línguas, etc. Deste modo, a GU é, ao mesmo tempo, flexível (acomoda a variação registada entre as diferentes línguas) e rígida (segue um rumo específico em direcção ao conhecimento final), permitindo que se adicionem às propriedades universais inerentes a ela própria quer as propriedades específicas das diferentes línguas quer as condições sociolinguísticas em que ocorre o *input*. Com base neste pressuposto, estudos recentes têm evidenciado que a GU condiciona a aquisição da linguagem tanto nos contextos monolíngues de aquisição da L1 como nas situações de bilinguismo e de ASL por adultos (White, 1989, 2003; Bruhn de Garavito, 1994; Meisel, 1994a; Hawkins & Chan, 1997; Franceschina, 2002, 2003; Montrul, 2004; Godinho, 2005, 2008; Silva, 2010).

Tendo como ponto de partida os pressupostos teóricos abordados, ao analisarmos a aquisição e desenvolvimento geracional de um aspecto particular da morfossintaxe do português, no caso concreto a CPL-var no SN de uma variante reestruturada resultante de transmissão linguística irregular, dentro da perspectiva da GU, estaremos não só a procurar descrever tal fenómeno mas também a dar o nosso contributo para validação de uma teoria linguística que leva em conta quer os princípios universais quer os aspectos particulares da aquisição das línguas, tal como tem sido feito em outros trabalhos que abordam particularidades específicas de aquisição de L2/LE (Godinho, 2005; Silva, 2010). Paralelamente, a conjugação de análises teóricas com metodologias quantitativas permitirá entender melhor o fenómeno da referida variação, já que cada uma delas nos concede instrumentos importantes para se concluir melhor acerca do trajecto e desenvolvimento diacrónico das características estruturais no interior do SN do PA, com vista a configurar uma possível mudança em curso.

### 2.3. Abordagens teóricas sobre aquisição bilingue

Os falantes de Almojarife são bilingues em português e santomense, pressupondo-se que tenham estado em situação de maior ou menor exposição aos dois dialectos como L1's, ainda que, por vezes, esta evidência seja negada por preconceito sociolinguístico. Independentemente da discussão estabelecida em torno do próprio conceito de bilinguismo, restringi-lo-emos, no âmbito do presente trabalho, a duas perspectivas: (i) considerar o indivíduo bilingue como um falante com uma modelagem única e específica, entendido como um todo (Grosjean, 1982, 1985); (ii) observar o bilingue levando em linha de consideração o papel desempenhado pela aquisição da sua L2. Assim sendo, o bilinguismo acaba por reflectir quer as situações de domínio de duas ou mais línguas quer a adopção de diferentes registos das mesmas, de acordo com o exigido pela pragmática (Dufour & Kroll, 1995:136).

O facto de um cada vez maior número de sociedades tender para o multiculturalismo levou a Neuropsicologia, a Psicologia Cognitiva e a Psicolinguística a procurem entender não só a forma como as línguas do bilingue estão representadas na sua rede cognitiva mas também o modo como ele acede à informação lexical (Oliveira, 2002:87). Os achados revelam que os bilingues desenvolvem estratégias de processamento da informação de acordo com o contexto em que se dá a aquisição de ambas as línguas e que os factores linguísticos (p.e. as diferentes estruturas das diversas línguas) e extralinguísticos (p.e. a idade e modo de aquisição ou a ordem de aprendizagem das línguas) intervêm directamente no modo como as duas línguas são adquiridas pelos bilingues (Hamers & Blanc, 2000 [1989]:115). Assim, as produções linguísticas de cada uma das línguas do bilingue representam *“uma matriz de traços distintivos, incluindo especificações semânticas, e vão ser conectadas a um sistema comum de representações conceptuais. A mensagem, antes de ser codificada e decodificada em cada língua, não faz parte da competência linguística implícita mas de um sistema conceptual.”* (Oliveira, 2002:89). Estes aspectos parecem indiciar que, no bilingue, existem duas representações lexicais específicas, uma para cada língua, enquanto a representação conceptual é comum a ambos os idiomas (Potter *et alii*, 1984:28). Contudo, vários autores discordam deste pressuposto, alegando que, tal como acontece na ASL, uma das línguas (a L2) é mediada pela outra (L1). Assim, uma das línguas do bilingue exerce dominância sobre a

outra (Kroll & Scholl, 1992:196), levando os aprendentes em situação inicial de aquisição das L1's a produzirem erros de morfologia flexional em ambas. A este propósito, lembremos que o presente trabalho tem igualmente como pressuposto a hipótese de que a L1 é o estado inicial para aquisição de uma L2, isto é, de que a L1 exerce dominância sobre a LA. Todavia, e como referiremos no ponto 2.4.1.1, não descartamos também a possibilidade de o estágio final de sistematicidade revelado pelos falantes de Almojarife aliar outro tipo de interferências às provenientes da L1, sejam elas de carácter cognitivo (maturidade, quantidade e qualidade do *input*, frequência de uso da LA, conhecimento de outras línguas, etc.) sejam do tipo extralinguístico (motivação, idade, sexo, escolaridade, aspectos culturais, etc.).

Paradis (1987:117) entende que o bilingue pode funcionar como um nativo nas duas línguas, uma vez que os itens lexicais armazenados contêm quer a forma quer o significado das palavras. Deste modo, nos dois armazenamentos lexicais estarão também presentes quer a forma (fonologia e propriedades sintácticas) quer os significados das palavras, enquanto o armazenamento conceptual, que é comum, contém as representações mentais das coisas observadas e dos eventos experienciados. Deste modo, o modelo entende que, aquando do uso da palavra, compete aos itens lexicais activar toda a panóplia de traços conceptuais que permitem obter uma representação mental do seu referente (Hagège, 1996:112). Este aspecto determina uma perfeita coordenação das duas línguas no bilingue, fazendo-o actuar com um nativo em ambas e levando-o a distinguir as características léxico-semânticas encerradas em termos equivalentes em tradução, mas que activam representações mentais distintas (Paradis, 1987:166; Oliveira, 2002:91).

### **2.3.1. Acesso ao léxico por parte do bilingue**

Alguns estudiosos entendem que cada palavra contém diversos tipos de informação a que correspondem várias representações na rede cognitiva do falante. Contudo, estas estão todas conectadas (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), o que levanta a questão acerca do modo como o bilingue acede ao léxico. Experiências com recurso ao *priming*<sup>56</sup> revelam que o *priming* semântico acontece entre as línguas do bilingue, em virtude de elas acederem a uma representação conceptual

comum. Por outro lado, o *priming* de repetição não ocorre, evidenciando que uma palavra e a sua tradução não compartilham representação mental. Para os autores que aplicaram este tipo de metodologia (Grainger & Beauvillain, 1988:59; Jin & Fischler, 1990:1153), a situação só é possível porque, a cada uma das línguas, correspondem entradas lexicais independentes, isto é, porque os sistemas lexicais do bilingue estão separados.

Fenómenos característicos dos falantes bilingues, como a alternância de códigos,<sup>57</sup> revelam a capacidade dos bilingues em alternar as suas línguas nas situações normais de fala (Romaine, 1989; Muysken, 1995; Myers-Scotton, 1997). No entanto, esta capacidade de alternância não é arbitrária (Pfaff, 1979; Clyne, 1980; Poplack, 1980b), sendo governada por regras de restrições semântico-estruturais, resultantes das interconexões estabelecidas entre as duas gramáticas do falante (Pfaff, 1979:307). Por conseguinte, a alternância de códigos incide unicamente sobre a Estrutura-S, não actuando sobre o significado das palavras e das frases, uma vez que este é armazenado a nível metalinguístico, ou seja, na Estrutura-P (Clyne, 1980:403). Este aspecto leva também o bilingue a buscar primeiro as palavras no léxico menos activado e, só posteriormente, no da língua de base (Soares & Grosjean, 1984:385, exemplo [35]).

Para Green (1986:213), as línguas do bilingue estão sujeitas a três níveis de activação: (i) nível da selecção, que leva a língua seleccionada a controlar o *output* da fala; (ii) nível da activação, no qual a língua activa colabora no processamento da fala, trabalhando paralelamente à língua seleccionada, mas sem aceder ao canal de fala em uso; (iii) nível da latência, correspondente ao da língua em repouso, ou seja, a língua que se encontra armazenada na memória a longo termo, mas sem intervenção activa no processamento imediato da linguagem. Deste modo, no momento em que o falante vai seleccionar uma das línguas para a usar numa determinada situação, as outras são activadas ou colocadas em suspenso. Estes aspectos permitem ao bilingue seleccionar mais do que uma língua para uso (Green, 1986:219), pelo que as palavras escolhidas durante o processo de fala são da língua seleccionada, mas a partir da língua activa, ou mesmo da língua latente. Por sua vez, a activação da língua implica que se accionem não apenas os seus itens lexicais mas também todo um conjunto de lexemas e unidades sublexicais (p.e. sílabas e fonemas). Assim sendo, a tipologia morfológica de qualquer

língua, em virtude de implicar o envolvimento da estrutura interna das palavras na produção linguística, tem influência determinante na alternância de código. A noção de equivalência entre gramáticas das duas línguas facilita o uso das mesmas pelo bilingue, levando a que ocorram não só alternâncias de código mas também interferências lexicais na sua fala. Este aspecto é tanto mais compreensível se considerarmos que existe equivalência não apenas de categorias (itens lexicais, estruturas fráscas, traços morfossintáticos) mas igualmente de relações entre categorias sintagmáticas (ordem das palavras na frase) ou paradigmáticas (oposições equivalentes). Realizações dos falantes do PA evidenciam isso mesmo, como acontece no caso das construções relativas ou relativas cortadoras (exemplo [160]), em que ocorrerá interferência/transferência da conjunção usada no seu substrato, o santomense (exemplos [158] e [159]), dada a semelhança fônica entre as formas do substrato e da LA ([ku] e [kə]):

[158] FORRO:  
*floli se ku sa i* (Ferraz, 1979:74)  
 flor MSPEC REL estar aqui  
 “Esta flor que está aqui”

[159] FORRO:  
*om ku Zõ sa ka fla n-e* (Ferraz, 1979:71)  
 homemREL João estar ASP falar em-ele  
 “o homem de que/de quem o João está a falar”  
 “o homem que o João está a falar”  
 “o homem que o João está a falar nele/dele”

[160] PA:  
*dia co ele tem, ele dá, mas dia que ele ãõ tem, ele ãõ tem mesmo*<sup>58</sup>  
 [SALVH3]  
 “nos dias (em) que ele tem, ele dá, mas nos dias (em) que ele não tem, ele não tem mesmo”

No santomense pode também ser observado que a conjunção *ku* cumpre ainda outras funções gramaticais, como acontece nas construções integrantes, (exemplo [161]):

[161] FORRO:  
*m-pesa kwe ka sobe* (Ferraz, 1979:65)  
 eu-pensar INTEGRANTE (KWE < KUE) KA chover  
 “eu penso que vai/irá chover”

Idêntica situação ocorre no PA, com a mesma conjunção estendendo a sua função a situações de subordinação:

[162] PA:  
*Mas eu ño tenho canoa, co aparado ño tem dinheiro pa fazê mas canoa*  
[SALVH3]  
“Mas eu ño tenho canoa, se fico parado ño tenho dinheiro para fazer  
mais canoas”

[163] PA:  
*é de vez quando co gente quero co vem fazer, manda-mo co ê faz*  
[SALVH3]  
“é de vez em quando que as pessoas querem que eu faça, mandam que eu  
faça”

Deste modo, alternâncias de códigos e interferências dependem de um conjunto de princípios linguísticos abstractos, patentes em diferentes comunidades linguísticas e, ao que tudo indica, baseados em aspectos cognitivos (Myers-Scotton, 1997:187). Portanto, estes fenómenos, apesar de emergirem na produção linguística, têm o léxico como base. E tal só é possível porque as estruturas gramaticais estão contidas em lemas (as entradas abstractas no léxico mental do falante) e incluem os aspectos fonológicos de um elemento de informação lexical. Por outras palavras: compete ao nível conceptual, o mais abstracto da produção da língua, a selecção dos lemas que vão ser activados, pelo que, neste tipo de fenómenos, a decisão sobre a língua que vai conceder os lemas a serem activados se encontra feita *a priori*. Conclui-se então que, seja na fala monolíngue seja na fala bilingue, a estrutura dos enunciados vai estar dependente das condições pragmáticas e sociopragmáticas em que as referidas falas acontecem.

### **2.3.2. Papel da GU na aquisição bilingue: a *Hipótese da Diferenciação da Linguagem***

Embora as duas línguas do bilingue sejam adquiridas de modo idêntico, na realidade não ocorre simetria entre ambas. De facto, cada uma das línguas está conectada a determinados indivíduos ou actividades do quotidiano, pelo que, com o desenvolvimento da criança, uma delas acaba por ser preferencial e passar a exercer dominância sobre a outra. Não obstante ainda não se ter determinado como e quando se dá a dominância de uma língua sobre a outra na aquisição bilingue, é aceite que a frase é inicialmente

construída de modo parcial na língua dominante e, só posteriormente, traduzida e articulada na língua dominada. O fenómeno da dominância é também frequente em falantes políglotas, como sucede com os aprendentes de PtgLE que tenham previamente adquirido o inglês como L2. De facto, é comum ouvir-se estes falantes realizarem frases do tipo “*Ele introduziu-me ao amigo dele*”, na qual acontece uma clara predominância do verbo inglês “*to introduce*” sobre o seu congénere português “apresentar”. Como se verifica, o que ocorre aqui é todo um conjunto de estratégias de compreensão baseado na língua dominante e não na língua em uso (Bates *et alii*, 1982:274).

Por norma, as crianças monolíngues aprendem a L1 com os pais. Contudo, no caso das crianças bilingues, a sua exposição às duas línguas é diferenciada, já que o contacto pode ser mais intenso com um dos pais, caso, por exemplo, o outro tenha que se ausentar de casa para trabalhar. Estudos feitos com crianças em início de aquisição bilingue, entre dois e três anos de idade, cujos pais detêm e falam diferentes L1’s, têm evidenciado que este tipo de aquisição bilingue é complexo, uma vez que as referidas crianças apresentam uma mistura de sons e palavras das suas duas línguas nas mesmas produções orais, mesmo que estejam a comunicar com falantes que usem apenas uma língua. Uma das justificações para as crianças produzirem este tipo de misturas assenta no facto de elas não deterem vocabulário suficiente numa das línguas, ou mesmo nas duas, para se expressarem completamente numa delas (Genesee, 1989:176). Portanto, a estratégia para tornar a comunicação eficaz é recorrer ao vocabulário que elas conhecem, mas que pertence à outra língua (Vihman, 1985:316).

Face a estes aspectos, e tendo em conta que a GU modela a aquisição em situações nas quais a criança é exposta simultaneamente a duas línguas, Genesee (1989:172) e Meisel (1994a:93) propõem a Hipótese da Diferenciação da Linguagem, que assenta na Teoria da Maturação (*Maturation Hypothesis*) de aquisição de L1’s (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994; Meisel, 1994a, 1997) e defende que a GU orienta os parâmetros das duas gramáticas (Pfaff, 1979:315), que actuam logo a partir dos primeiros enunciados da criança (Köppe, 1994b:225), desenvolvendo-se, depois, de modo paralelo, mas interagindo (e mesmo influenciando-se) mutuamente (Sridhar & Sridhar, 1980:414). Assim, e de acordo com a hipótese maturacional, a fase FL0 apenas envolve o módulo pragmático da linguagem, não

detendo o módulo do conhecimento gramático, já que o desenvolvimento deste está sujeito a maturação, isto é, à aquisição das categorias funcionais.

Em estudos circunscritos ao quadro dos Princípios e Parâmetros (Chomsky, 1981a), levados a cabo no âmbito do projecto DUFDE (*Deustsh und Französisch – Doppelter Erstspracherwerb*),<sup>59</sup> ou seja, com crianças em aquisição bilingue simultânea de francês e alemão como L1's, Köppe (1994b) estudou a ordenação das palavras em estágio precoce de desenvolvimento da gramática infantil, com vista a detectar a fase em que a criança tem acesso ao movimento SN e respectiva elevação do sujeito. Baseando-se no Princípio do Comprimento Médio dos Enunciados (*Mean Length of Utterance*) e atentando ao momento em que a criança passa a combinar verbos com SN's, Köppe (1994b) buscou determinar se o movimento constitui um princípio da GU disponibilizado na fase FL0, ou, pelo contrário, se encontra sujeito a maturação.

A conclusão de Köppe (1994b:233) é que os princípios da GU estão presentes desde sempre nas duas línguas do bilingue, passando a operar a partir do momento em que as categorias funcionais emergem. A evidência empírica para este pressuposto é fornecida através do modo como as crianças conseguem analisar precocemente as propriedades lexicais de diferentes tipos de verbos, o que lhes permite projectar as diferenças estruturais a nível do SV. O pressuposto de Köppe (1994b), claramente continuísta, leva a entender que a gramática das crianças bilingues não só é igual à das monolíngues como também o é em relação à dos adultos, uma vez que tem todos os princípios disponíveis desde o começo da elaboração gramatical (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996).

Também no âmbito do projecto DUFDE, Meisel (1994a) investigou o modo como se processa a aquisição de categorias funcionais e quais as consequências desta para o desenvolvimento das gramáticas em questão. Para tanto, Meisel (1994a:92) partiu do pressuposto que a gramática dos pequenos falantes tem como base inicial as categorias lexicais, sendo as funcionais implementadas gradualmente. Deste modo, o autor opina a favor da Teoria da Maturação (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994), que se contrapõe à Teoria Continuísta (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996), defensora de que “*a linguagem*



*primitiva da criança já é constrangida por princípios da GU e pelo valor não-marcado dos parâmetros” (Kato, 1995:69).*

Observando a aquisição de finitude, concordância e tempo nas gramáticas das crianças, Meisel (1994a:127) conclui que estas não possuem categorias funcionais na fase FL0, as quais vão apenas emergir faseadamente no tempo, acabando por determinar o desenvolvimento das categorias gramaticais associadas à flexão (concordância, tempo e finitude). Assim sendo, a criança não tem inicialmente acesso à gramática, pelo que recorre a princípios semântico-pragmáticos para organizar a sua fala. Por volta dos 2;0 anos, a maturação neurológica da criança disponibiliza-lhe a GU, passando a sua gramática a operar, nesta fase, dentro da Hipótese Continuista (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996). Por outras palavras: Meisel (1994a) não acredita que exista qualquer dependência entre o processo que operava antes dos 2;0 anos e aquele que passa a agir a partir da maturação, ou seja, os princípios semântico-pragmáticos não estão sujeitos a processo de gramaticalização.

As observações de Meisel (1994a:128) levaram-no a concluir que, na gramática infantil, a concordância é a primeira categoria gramatical que permite distinguir as formas verbais finitas das infinitas. Deste modo, as marcas de tempo só aparecem após a criança apresentar todas as marcas de concordância pessoal e grande maioria das concordâncias de número. Estes aspectos fizeram Meisel (1994a:128) concluir também que o processo de maturação se fica a dever à instanciação de traços, já que a finitude é definida inicialmente nos moldes da concordância e, só depois, em termos de tempo.

Com base nos achados de Köppe (1994b) e nos postulados da Hipótese da Diferenciação da Linguagem (Genesee, 1989; Meisel, 1994a), é lícito inferir que, dentro do mesmo espaço de tempo em que as crianças monolingues se tornam competentes numa língua, as bilingues aprendem duas e utilizam-nas hábil e apropriadamente em meios socialmente distintos. Portanto, monolingues e bilingues apresentam progressos similares no desenvolvimento da linguagem, aproximadamente nas mesmas idades. As diferenças na aquisição da linguagem a nível individual reflectem apenas que algumas crianças, sejam elas monolingues sejam elas bilingues, adquirem as suas primeiras palavras ou usam estruturas orais complexas mais precocemente do que outras.

Ainda no âmbito do projecto DUFDE, Müller (1994b) analisa a questão do restabelecimento dos parâmetros, a que nos referiremos mais detalhadamente no ponto 2.8 do presente trabalho. Observando a ordem dos constituintes em orações encaixadas do alemão e do francês, a autora entende que a GU contém uma restrição para o accionamento paramétrico. Nesta conformidade, Müller (1994b:268) postula que os parâmetros não podem ser reaccionados durante o desenvolvimento linguístico, indo ao encontro dos pressupostos avançados pelos defensores de que, apesar de os parâmetros não poderem ser restabelecidos, a GU continua a intervir no processo de ASL (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashes, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003).

Face ao exposto sobre aquisição bilingue, o nosso trabalho segue também o pressuposto de que a GU não está apenas subjacente a todo e qualquer contexto monolingue de aquisição de L1, já que estende a sua acção aos domínios da aquisição bilingue ou plurilingue. Contudo, teremos também em conta que, neste tipo de aquisição, o contexto sociohistórico é determinante para que uma das línguas acabe por se tornar preferencial e passe a exercer dominância sobre a outra. Na comunidade de Almojarife, a primeira fase da aquisição, por monolingues, conduz ao aparecimento do santomense como língua preferencial de comunicação interétnica, restringindo o uso das L1's do grupo níger-congo atlântico a contextos meramente familiares. Posteriormente, a aquisição do português em situação de transmissão linguística irregular vai determinar, em contexto de bilinguismo, a emergência do PA, marcado por elevada variação, dada a defectividade dos DLP's. Posteriormente, a variação dilui-se ligeiramente em direcção ao sistema da LA, como bem o revela a aquisição da regra de concordância pelas gerações mais novas e mais escolarizadas de Almojarife, ou seja, pelas gerações expostas a maior contacto com a LA.

#### **2.4. Abordagens teóricas sobre aquisição de L2's**

Presentemente, as teorias sobre a intervenção da GU na ASL apoiam-se em diferentes perspectivas: (i) acesso total à GU, isto é, a versão forte sobre o que é transferido (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006); (ii) acesso parcial à GU, ou seja, a versão fraca sobre o que é transferido (Lefebvre, 1986;

1998; Jake & Myers-Scotton, 1998; DeGraff, 1999a, 1999b; Siegel, 2004a; 2004b); (iii) influência da L1 nos estágios iniciais de aquisição da L2, entendendo-se aqui uma nova perspectiva sobre a abordagem contrastiva entre L1 e L2, não limitada à questão da transferência de hábitos (Lightbown & Spada, 2010 [2006]).

#### **2.4.1. Modelo generativista e papel da GU na aquisição de L2's**

Como mencionámos nos pontos 2.2.3 e 2.3.2, a aquisição de L1's (contexto monolíngue ou bilingue) processa-se interna e rapidamente, ainda que as crianças produzam variação no processo de desenvolvimento da linguagem durante as fases precoces da aprendizagem, o que não acontece na ASL. No que concerne a esta, actualmente bastantes linguistas defendem que a melhor forma de a entender passa pela perspectiva da GU. Porém, a ASL, depois de ultrapassado o período crítico do desenvolvimento, terá que ser explicada com recurso a outro quadro teórico, pois os mecanismos da GU poderão já não se encontrar disponíveis nesta fase (Clahsen & Muysken, 1986; Muysken, 2001). Efectivamente, se se levar em conta que a faculdade da linguagem é biologicamente determinada e que está presente na aquisição de uma L1, é também de considerar que esses princípios tenham um papel interventivo na ASL (Flynn, 1989:92), ainda que o contexto desta se afaste radicalmente daquele que determinou a emergência da gramática da LA. Assim, ambos os tipos de falantes passam por vários estágios de construção da gramática, acabando a aquisição das regras reguladas pela GU por reflectir as competências linguísticas nesses mesmos estágios. Assim, as diferentes gramáticas que a interlíngua em maturação conhece, (IL1, IL2, IL3...), desde a fase FL0 até ao patamar FLf, traduzem-se na competência final (ILn) (cf. esquema de Godinho, 2005:98), a qual só difícil e esporadicamente alcançará o mesmo nível de *performance* dos falantes nativos da LA.

Na ASL, as opiniões quanto ao que será o seu estágio inicial divergem em duas correntes: uns defendem que é a própria GU, tal como acontece na aquisição da L1 (Flynn, 1989; White, 1989, 1990); outros entendem que é a L1, mantendo-se a GU activa enquanto a interlíngua se desenvolve, a fim de substituir a L1 quando esta revela lacunas ao nível da assimilação do *input* ou do fornecimento de representações (Montrul, 2004). Por seu lado, os que compartilham a visão de que o estágio inicial da ASL é a GU

acham que esta possui um dispositivo computacional inato de princípios universais e invariáveis (cf. Chomsky, 1996 [1995]), comuns a todos os idiomas, por forma a regerem a forma e o funcionamento da gramática, imprimindo as diferenças estruturais entre as línguas. Tal é possível porque o léxico é dotado de parâmetros que actuam sobre as categorias lexicais e funcionais específicas das línguas. As primeiras (categorias lexicais), como o seu próprio nome indica, detêm propriedades que definem a categoria lexical da palavra, fornecendo o conteúdo semântico às produções linguísticas; por seu lado, as segundas (os morfemas funcionais) possuem características universais semânticas e de flexão, realizadas através de formas morfofonológicas, e interagem com a morfologia flexional para concederem a informação gramatical (género, número, concordância, caso, negação, tempo e aspecto). Dispondo destes dados, e tendo em conta que as categorias lexicais são dominadas pelas funcionais, isto é, que as diferentes unidades linguísticas estabelecem relações hierárquicas entre si, compete ao dispositivo computacional filtrar estas, com vista a transformá-las em representações sintácticas. Assim, ao contrário do que sucede com o dispositivo computacional ou com as categorias funcionais, a construção da frase não é universal, pois depende do léxico da língua (ponto 2.2.2.1.2 do presente trabalho).

Face à teoria de que o estágio inicial da ASL é a GU, conclui-se que o dispositivo computacional, pelo facto de ser inato, não necessita de ser aprendido. O mesmo não acontece com o léxico particular (as palavras e a morfologia) das diferentes línguas, que precisa de ser aprendido durante o processo de aquisição. Ao adquirir a L1, a criança procede também à aquisição do léxico e das categorias funcionais desta (os parâmetros que regem o conjunto de propriedades sintácticas e semânticas abstractas da L1). Na ASL, ocorre idêntico processo, pois é necessário adquirir o léxico e os parâmetros da LA.

#### **2.4.1.1. GU e teorias em torno da aquisição de L2's**

Tendo em conta que, no estado inicial da ASL, o aprendente já possui a gramática da GU, com princípios e parâmetros estabelecidos, das quatro teorias que foram mencionadas acerca da disponibilidade da GU na aquisição de L1's (ponto 2.2.3) apenas a Hipótese Continuista (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996)

e a Hipótese da Maturação (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994; Meisel, 1994a, 1997) estão implicadas na ASL (Godinho, 2005:102). De acordo com a Hipótese Continuista, a gramática das crianças é igual à dos adultos, tendo todos os princípios disponíveis desde o início da elaboração gramatical. Deste modo, e dado que a criança possui competência gramatical total na língua a que se encontra exposta, se os parâmetros não forem fixados, é porque se verificam dificuldades na aquisição. Caso não ocorram estas, as crianças produzem frases gramaticais, conduzidas pelos parâmetros filtrados pela GU.

Por seu lado, a Hipótese da Maturação entende que os princípios e parâmetros estão geneticamente programados para serem fixados de modo gradual, isto é, em diferentes estágios de maturação. Como tal, apenas determinadas categorias gramaticais operam na fase inicial da elaboração de frases, actuando outras em período mais tardio, e segundo uma ordem específica. Assim, interessa perceber em que fase da maturação ocorre a transformação da GU em gramática da língua a ser aprendida.

A Teoria da Maturação advoga ainda que existe uma diferença acentuada na ASL entre crianças e adultos. Assim, enquanto nas primeiras ocorre uma fase de maturação semelhante à da aquisição da L1, nos segundos essa fase já se encontra concluída por terem efectuado a aquisição da L1 (Yip, Rutherford & Clahsen, 1995:25). De facto, estudos sobre aprendentes da mesma L2, mas detentores de distintas L1's, apontaram para diferentes potencialidades de ASL (Schachter, 1989; Franceschina, 2002; White *et alii*, 2004). Estes resultados permitiram também que se avançasse com três hipotéticas premissas direccionando a ASL, podendo as mesmas partilhar aspectos comuns ou distinguir-se nas suas etapas até ao estágio final (Godinho, 2005:103-104): (i) a GU actua a partir do estado inicial da aquisição; (ii) a GU actua a partir do estado inicial da aquisição, mas assentando na L1; (iii) registam-se importantes transferências da L1 para a L2. Actualmente, é consensual considerar que a GU actua através da L1 na ASL, divergindo as teorias apenas acerca do momento da operacionalidade: estado inicial ou intervenção apenas quando se verificar necessidade de tal?

Considerando que a GU opera por intermédio da L1 na ASL, compreende-se o porquê de a interlíngua deter características formais da L1 e da L2. Este aspecto, por outro lado, vem confirmar a existência de diferenças e semelhanças nos processos de aquisição de

uma L1 e de uma L2. A nível das diferenças, poder-se-ão destacar (Godinho, 2005:104-105): o facto de o aprendiz da L2 já ter desenvolvido um sistema linguístico (a L1); o facto de os adolescentes e adultos possuírem maturidade cognitiva; o facto de o estágio final da L2 poder revelar variações (erros) fossilizadas; e ainda o facto de os aprendentes da LA nunca atingirem o nível de proficiência dos falantes nativos desta. Quanto às semelhanças entre aquisição de uma L1 e ASL, podem apontar-se as seguintes (Godinho, 2005:105): ambas têm como objectivo a aquisição de um sistema complexo, a partir de *input* reduzido; os aprendentes das duas cometem o mesmo tipo de variação, fazendo pressupor a existência de um sistema linguístico interiorizado de regras sistemáticas (os princípios e parâmetros da GU), que vai governar as construções linguísticas. Face a este quadro de possibilidades, os generativistas, no que concerne à ASL, têm procurado explicar como se desenvolve a competência linguística do aprendiz, seja em situação de aprendizagem formal seja em contexto natural de aquisição.

Assim, e ainda relativamente à questão da disponibilidade/indisponibilidade da GU na aquisição de L2's, são quatro as posições assumidas actualmente pelos linguistas (cf. Godinho, 2005:106-111; Figueiredo, 2009a:31-32; Silva, 2010:40-44):

- (1) A GU actua na ASL, mas apenas de modo parcial, porque, apesar de os princípios se manterem acessíveis, os parâmetros não podem ser restabelecidos (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashes, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003). Uma evidência para sustentar esta hipótese é fornecida fonologicamente, por exemplo, pela perda de capacidade discriminatória a nível dos padrões trocaicos e jâmbicos que ocorre em idade pós-adolescência.
- (2) A GU actua de forma igual (total) na aquisição da L1 e da L2, sendo ambos os processos bastante similares. Desta forma, os aprendentes da L2 têm acesso total aos princípios e parâmetros desde o início da aquisição e até que esta se complete (Cook, 1994; White, 1989, 2003; Bruhn de Garavito & White, 2002; Montrul, 2004; Cook & Newson, 2007 [1988]). As competências em L1 e L2 são paralelas (duas configurações na GU), só assim se compreendendo o porquê de os falantes avançados de uma L2 deterem representações mentais complexas desta e que se

podem diferenciar das que são elaboradas pelos falantes da LA. Portanto, a relação entre a GU e a aquisição da L1 traduz-se na competência em L1 (Gramática *n*), enquanto a relação entre GU e L2 produz a competência em L2 (IL *n*). Porém, o estágio final de aquisição, e conseqüente variação da interlíngua, estão dependentes da maior ou menor pobreza dos estímulos durante a aprendizagem. Defendem os seguidores do acesso total à GU (*Full Transfer/Full Access Model – FTFA*) que, mesmo em situação de pobreza de estímulos muito acentuada, a ASL não cessa. Assim, e como esta não pode ser atribuída ao *input* que provém do meio envolvente, a ASL tem que ser vista como uma propriedade genética do ser humano. Como se vê, a perspectiva do acesso total à GU engaja na Teoria Continuista da disponibilidade da GU na aquisição de L1's (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996), levando a entender que existe continuidade no desenvolvimento linguístico que configura a ASL.

- (3) A GU actua na L2 através da L1,<sup>60</sup> pois é hoje considerado que a maioria dos aprendentes de uma L2 parte do princípio que esta é igual à L1. Argumentam os autores da corrente do acesso indirecto à GU que o ponto de partida para a ASL é a gramática da L1 e que as propriedades computacionais desta se transferem em bloco para as fases iniciais da ASL, excepto as matrizes fonéticas e os itens léxico-morfológicos (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006). Assim, numa fase inicial da ASL, os aprendentes cometem erros de morfologia flexional, mas, com a aquisição intermédia de morfologia flexional relevante e de itens lexicais, é possível refixar os parâmetros e atingir um estágio final de aquisição. Portanto, a posterior exposição ao crescendo de *input* reconstrói, sucessivamente, a interlíngua. Erros em fase tardia são justificados pela FFFH (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), visto as categorias funcionais deixarem de estar disponibilizadas depois do período crítico do desenvolvimento (Lenneberg, 1967; Chomsky, 1972; Larsen-Freeman & Long, 1991; Ellis, 1994). Contudo, é tido como certo que a construção gramatical é orientada pela GU, que mantém sempre os seus princípios activos, permitindo que os aprendentes adquiram valores

paramétricos a partir dos valores de parametrização da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006).<sup>61</sup>

- (4) A GU deixou de estar disponível para actuar na ASL, pois já desempenhou o seu papel na aquisição da L1 ou aquisição bilingue e impede os aprendentes da L2 de acederem aos princípios universais ou de estabelecerem os parâmetros (Clahsen & Muysken, 1986; Muysken, 2001). Neste aspecto, destaque-se a posição radical de Meisel (1997, *apud* Godinho, 2005:106), que corrobora Selinker (1972:216) ao considerar que a interlíngua não é um sistema natural nem sistemático, uma vez que os aprendentes de L2's são incapazes de estabelecer correctamente a dependência estrutural da sintaxe complexa. Assim sendo, a L2 não tem qualquer relação com a faculdade da linguagem, sendo adquirida com recurso a capacidades cerebrais distintas desta. O pressuposto assenta nos princípios advogados pela perspectiva psicolinguística de interlíngua (Selinker, 1972), já que a competência em L2 é bastante distinta da competência em L1, podendo a LA ser adquirida por indução, o que não acontece com a L1.

O Programa Minimalista (Chomsky, 1996 [1995]), a que faremos referência no ponto 2.6.2.1, advoga que os traços não interpretáveis na fase FL0 da aquisição, como acontece com os traços de género e número, são apenas usados para activar operações sintácticas. Como tal, podem ser apagados, caso a concordância se dê com traços valorados a partir de outros elementos nucleares da estrutura frásica. Este aspecto pode determinar uma competição de gramáticas na interlíngua, caso se entenda que a L1 é o estado inicial para a ASL e ambas as línguas possuam sistemas distintos de marcação em género e número, que é exactamente o que sucede a nível do santomense e do português. Assim, a concordância entre os constituintes frásicos está dependente de aspectos como a Condição de Visibilidade,<sup>62</sup> que leva a que os elementos componham o seu valor paramétrico a partir do valor paramétrico do núcleo, produzindo-se a concordância entre todos os itens de acordo com o valor patente neste. Seguidamente, os elementos são projectados para a FL, levando os seus radicais terminais definidos para um determinado valor, no caso o valor do núcleo. Ora, aqui levanta-se um problema a nível da construção sintáctica na interlíngua, já que o padrão de marcação pode ser diferente na L1 e na L2,



levando o falante a enfrentar dificuldades na fixação dos valores na interlíngua. Daí que, tal como advogam os defensores da hipótese do restabelecimento de parâmetros (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), os aprendentes de L2's, em idade pós-crítica elaborem construções em que é evidente a simbiose entre sintaxe da L1 e itens lexicais da LA (formas morfofonológicas da L2 em especificações de traços da L1) (DeGraff, 1999b:482).

O facto de as construções divergirem nas produções da L1 e da L2 revela que elas são condicionadas por princípios universais, o que mostra a importância da GU na ASL. Este aspecto leva também a que se direcione atenção aos traços funcionais, já que se estes exibirem proximidade na L1 e LA, ocorre também um desempenho mais aproximado entre aprendentes da L2 e falantes nativos da LA (Hawkins & Chan, 1997:220). Por outro lado, reside também nos traços funcionais a justificação para o facto de os aprendentes de uma L2, ainda que continuamente expostos a dados da LA, não conseguirem adquirir as mesmas representações sintácticas dos falantes nativos desta. Deste modo, apesar de os falantes da L2 alcançarem competência no que respeita às propriedades sintácticas distribucionais dos itens funcionais, poderão não as reproduzir de modo sistemático, porque não conseguem fazer corresponder as características formais às formas morfofonológicas (Godinho, 2005:109). Neste caso, estamos perante uma situação que alguns autores (Prévost & White, 1999:219) consideram como superficial (*Missing Surface Inflection Hypothesis* – MSIH).

Como se vê, qualquer dos modelos proposto acerca do acesso à GU na ASL pode ser válido para diferentes aprendentes, ou mesmo para distintos aspectos da língua em relação ao mesmo aprendente (Cook, 1994; Cook & Newson, 2007 [1988]). Assim, o problema do investigador será escolher, entre todos eles, os que melhor resposta dêem às situações com que se depara no processo de análise.

#### **2.4.2. Aquisição de L2's e modelos psicocognitivos**

Os seguidores do cognitivismo entendem que o conhecimento linguístico é indissociável das capacidades cognitivas do aprendente (Godinho, 2005:81). Assim, a ASL representa uma construção de sistemas do conhecimento, direccionada para a compreensão e uso. Dado que a capacidade de retenção do conhecimento é limitada, o

aprendente necessita da experiência e da prática para ir, progressiva e automaticamente, utilizando os conhecimentos que vai interiorizando (McLaughlin, 1987; McLaughlin & Heredia, 1996, *apud* Godinho, 2005:81).

#### **2.4.2.1. Conexionismo (*Connectionism*) e modelos de processamento paralelo distribuído**

Alguns pesquisadores entendem a capacidade de aquisição como o resultado de uma série de conexões que o cérebro estabelece a nível das suas redes neurofisiológicas. Estas ligações ocorrem em módulos informativos e tornam-se mais fortes ou mais fracas, caso haja ou não activação das conexões mentais e neurológicas que se interligam, levando-as a excitarem-se ou inibirem-se umas às outras, consoante os elementos linguísticos ocorram conjunta e frequentemente, ou não. Como não existe um módulo neurológico específico para a aprendizagem da linguagem, esta deve ser demarcada do conjunto de regras abstractas que lhe é atribuído, pois a aquisição não passa de um mero processamento cerebral do *input* (Gass & Selinker, 2001:216, *apud* Godinho, 2005:82).

Segundo o modelo, os aprendentes de uma língua são sensíveis ao *input* das suas formas regulares, delas extraindo e testando padrões probabilísticos de aplicação. Por conseguinte, os pressupostos do Conexionismo não lidam exclusivamente com a construção do conhecimento linguístico (a competência) mas debruçam-se também sobre a forma como se acessa a este conhecimento e se aplica o mesmo (a *performance*). Para observar a forma como se estabelecem as conexões no processo de aquisição, Rumelhart & McClelland (1986) testaram hipóteses com recurso a modelos computacionais programados para reproduzirem as redes neurofisiológicas do cérebro humano e suas activações, à medida que a aprendizagem vai prosseguindo. Trabalhando com crianças, os pesquisadores observaram a aquisição das formas verbais regulares e irregulares do pretérito no inglês, conseguindo que o programa indicasse o modo como as conexões iam, progressivamente, sendo activadas. Estas observações seriam continuadas por Pinker (1991:487), o qual concluiu que somente os verbos irregulares são recuperados a partir de associações de memória, já que os verbos regulares são produzidos em resultado do processamento das regras de sufixação.

Quanto à ASL, Sokolik & Smith (1992:58), com recurso a métodos conexionistas, investigaram a atribuição do género nos nomes em francês, concluindo que a diferença

registada entre aprendentes do francês L1 e L2 é explicada com a troca de duas variáveis no modelo: (i) nas situações em que o computador atribuiu um estado zero de conexão aos aprendentes de L1, assumiu também que os aprendentes de L2 cumpriam a mesma tarefa, aplicando padrões conectivos pré-existentes, os quais interferiam na aquisição; (ii) o modelo considerou ainda que a ASL apresentava um grau de aprendizagem mais baixo. Como se pode constatar nas considerações em (i), ocorre um paralelismo com os pressupostos generativistas de que o ponto de partida para a ASL é a gramática da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006).

Por seu lado, Ellis & Schmidt (1997), partindo dos pressupostos advogados por Rumelhart & McClelland (1986) e Pinker (1991), analisaram a aquisição da morfologia PL em adultos, verificando que a mesma resulta da associação de determinados mecanismos a nível cerebral. Nesta conformidade, as estruturas complexas são construídas a partir de processamentos simples, ou seja, envolvem automatizações, tal como fora advogado em trabalhos psicolinguísticos prévios (Schiffrin & Schneider, 1977; Schneider & Schiffrin 1977; McLaughlin, 1987, 1990).

Contudo, o modelo conexionista apresenta desvantagens, já que se torna questionável determinar até que ponto se podem isolar variáveis que interagem em contextos naturais e, conseqüentemente, até que ponto os resultados obtidos desta forma espelham o que acontece na vida real. Por outro lado, o que se observa com recurso a esta metodologia tende a ser demasiado específico e localizado, daí resultando o perigo de se ignorar o modo como os diferentes aspectos da aquisição interagem. Assim, o conhecimento sobre os eventos, conceitos e linguagem surge representado de modo bastante difuso no sistema cognitivo. Este aspecto, associado ao facto de a aprendizagem ser apresentada como um sistema cumulativo, dificulta ainda a explicação acerca do modo como ocorre a aquisição das categorias funcionais marcadas por sintaxe complexa.

#### **2.4.2.2. Teoria da “Processabilidade” (*Processability Theory*)**

A morfologia flexional pressupõe a distinção entre três categorias: flexão inerente, flexão contextual e flexão morfofonémica. A primeira não é requerida pela sintaxe, mas tem relevância sintáctica (Booji, 1995:2). É o caso da flexão da categoria de número dos nomes, do grau comparativo e superlativo dos adjectivos e do tempo e aspecto dos

verbos, em que a presença de traços morfossintáticos num determinado lexema afecta a sua referência. Quanto à flexão contextual, é orientada pela sintaxe e abarca a pessoa e número dos marcadores verbais que concordam com o sujeito e/ou objectos, marcadores de concordância dos adjectivos e marcadores estruturais de caso dos nomes (Booji, 1995:2). Portanto, os traços morfossintáticos expressam-se no lexema como consequência de uma parte da construção sintáctica, não afectando a sua referência. Por fim, a flexão morfofonémica não expressa traços de significado ou traços morfossintáticos funcionais, possibilitando apenas que os traços morfossintáticos possam ser realizados (Kihm, 2010:1). É o que acontece com a flexão de género e número, cujos exponentes não podem ser afixados ao radical verbal sem a inserção de uma vogal temática, mas que não expressa qualquer tipo de traço: (p.e. cant-a-s/\*cant-s).

Assentando em modelos psicolinguísticos que visam explicar o modo como a produção do discurso se desenvolve, a Teoria da “Processabilidade” (Processability Theory) defende que o desenvolvimento morfossintático da interlíngua segue um trajecto universal perfeitamente definido (Pienemann, 1998, 2005b). Observando a aquisição da ordem das palavras e de estruturas gramaticais no alemão L2, Pienemann (1998) defende que estas obedecem a um processamento rígido. Assim, as novas estruturas só são aprendidas caso as etapas de aquisição prévia estejam concluídas, isto é, que os processamentos linguísticos arquitectados pelo processador da língua estão dependentes de uma hierarquia universal inata, não necessitando o aprendente de utilizar estratégias ou habilidades para produzir as realizações da L2. Desta forma, as predições sobre o desenvolvimento da ASL podem ser testadas empiricamente. Tomando-se como exemplo a frase “O filho deu um gato à mãe”, poder-se-á elaborar o diagrama da Fig. 14, que ilustra o modo como se processa a aquisição dos elementos frásicos.

A produção do primeiro SN da frase só é possível após ser gerada a mensagem pré-verbal na unidade que Levelt (1989:23) classificou de conceptualizador. Esta transmite o conceito “FILHO” ao codificador gramatical, responsável pela produção da pertinente estrutura gramatical, ou seja, o SN. O processamento de parte da estrutura frásica é feito com itens lexicais resgatados do léxico, pelo que a selecção do lema “FILHO” determina a categoria N, que fornece o DLP em torno do qual é construído o SN.

CONCEPTUALIZADOR

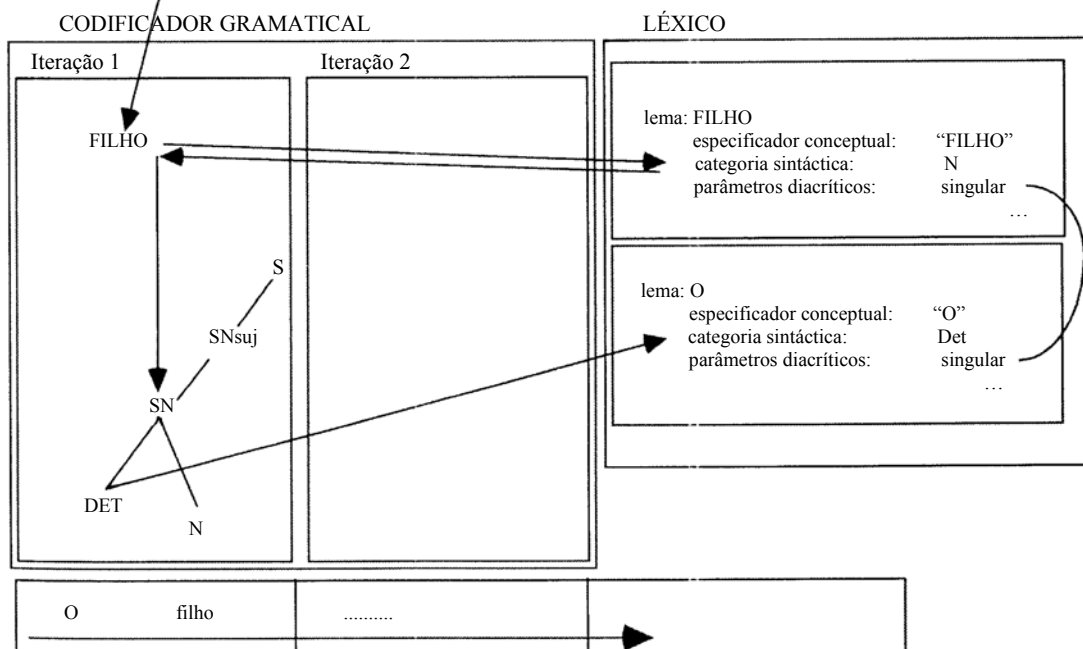
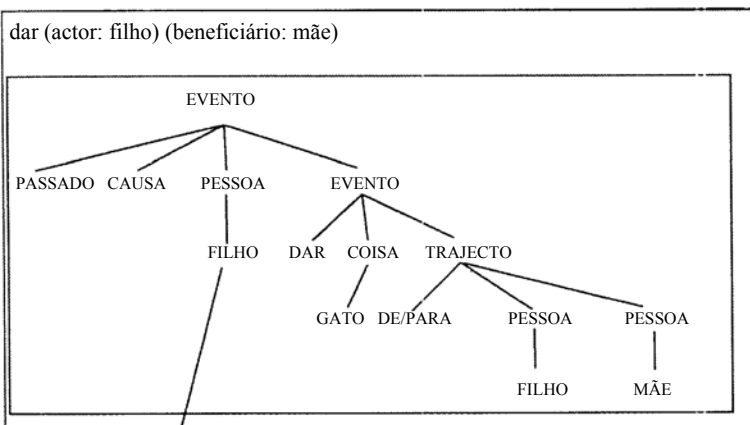


Fig. 14. Incrementação estrutural da interlíngua (Pienemann, 1998:68)

O diagrama revela que o processo de construção frásica implica, necessariamente, as seguintes tarefas (cf. Levelt, 1989:236):

- fornecimento da estrutura sintáctica na qual poderão ser inseridos os itens lexicais;
- busca dos traços morfossintácticos que estabeleçam o pareamento com a estrutura conceptual (p.e. [+ definido]);
- selecção do traço [singular] para o lema “FILHO”;
- pareamento dos traços de itens que possam ser inseridos sob o nó do DET do SN com os do lema seleccionado que detém a função de núcleo;

- resgatar do léxico os lemas pertinentes (p.e. o determinante “O”);
- por fim, transferir a estrutura resultante para unidades de processamento posterior, como o codificador fonológico e o articulador.

O exemplo abordado ilustra uma das características fundamentais da produção do discurso: a incrementação. Esta leva a estrutura a ser gradualmente construída com o processamento da conceptualização, operando os processamentos subsequentes conjuntamente com *outputs* prévios ainda incompletos, que obrigam a disponibilizar na memória, a curto termo, *outputs* intermédios incompletos. Verifica-se então que produção linguística envolve bastantes processamentos paralelos, consideráveis constatações memorizadas a curto termo e a disponibilidade para lidar com todo o tipo de estruturas rotineiras de processamento específico (Pienemann, 1998:7), tais como:

- acesso ao lema;
- processamento categorial;
- processamento frásico;
- processamento oracional (processamento-S);
- processamento da oração subordinada, se for caso de tal.

Como o processamento frásico é responsável pela troca da informação gramatical no interior da frase, os traços morfossintácticos dos constituintes são pareados durante a sua realização (Fig. 15). Note-se que a troca de informação intra e interfrásica seria impossível se não ocorressem, antes, o processamento do acesso ao lema – fornece o material lexical e os seus respectivos traços diacríticos – e o processamento da categoria – concede a informação da categoria sintáctica que permite a elaboração da estrutura posterior –. Por seu lado, o processamento-S implica o acordo entre o sujeito e o verbo. Este só é possível caso se respeite a estrutura sintáctica da direita, isto é, caso o sujeito estabeleça pareamento com a informação gramatical pertinente (a terceira pessoa do singular). Verifica-se então que o processamento-S é aquele que permite a troca gramatical entre os SV e SN sujeito:

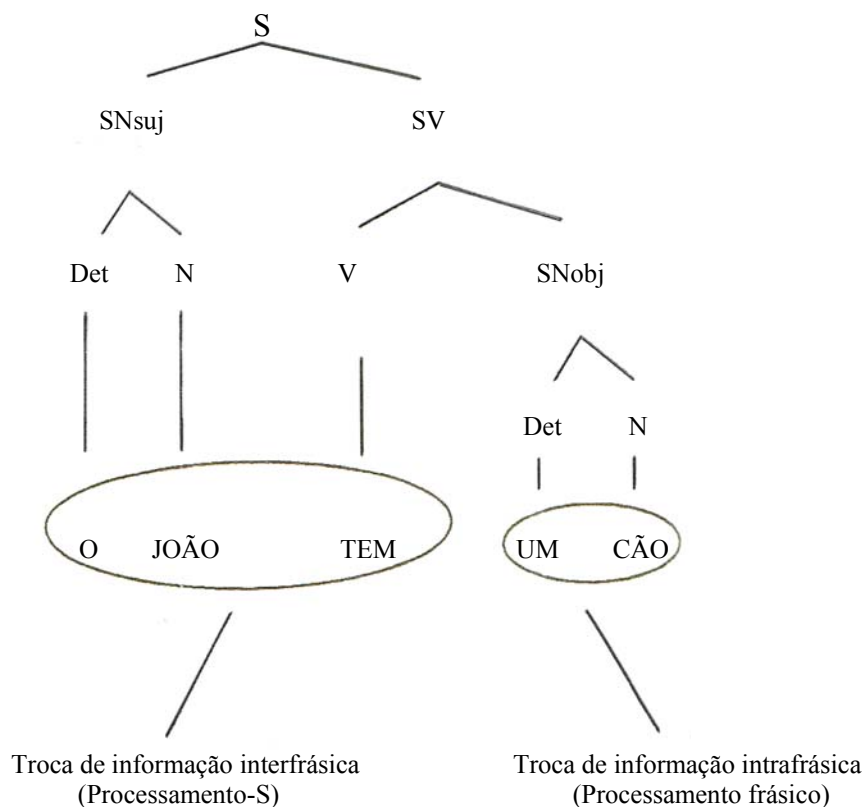


Fig. 15. Processamento-S e Processamento Frásico (Pienemann, 2000:101).

Seguidamente, na Tabela 2.1 apresenta-se uma hipotética hierarquia do desenvolvimento dos estágios de ASL em inglês:

**Tabela 2.1.** Hierarquia hipotética de articulação dos processamentos na aquisição da interlíngua (Pienemann, 2000:108).

	T1	T2	T3	T4	T5
<b>Processamento-S'</b> (S embebido)	-	-	-	-	+
<b>Processamento-S</b>	-	Simplificado	Simplificado	Troca de informação interfrásica	Troca de informação interfrásica
<b>Processamento frásico</b>	-	-	Troca de informação frásica	Troca de informação frásica	Troca de informação frásica
<b>Processamento categorial (categoria lexical)</b>	-	Morfemas lexicais	Morfemas lexicais	Morfemas lexicais	Morfemas lexicais
<b>Acesso (palavra / lema)</b>	+	+	+	+	+

Na tabela acima verificam-se 5 momentos de aquisição (T1 a T5), em que ocorrem outros tantos estágios diferenciados de desenvolvimento da interlíngua. As evidências

empíricas têm revelado que um processamento-S simplificado já se encontra disponível em estágios bastante precoces da aquisição. Em T1, o aprendiz apenas consegue produzir realizações holofrásticas, isto é, frases compostas por um único lema de categoria indefinida e resgatado do léxico. No estágio seguinte, o falante resgata morfemas lexicais e constrói frases bastante simples, tipo NVN (SVO) ou NNV (SVO ou SOV), mas que demonstram que os referidos morfemas já exibem uma categoria específica nesta fase da aquisição. No período T3, ocorre apenas concordância interna no SN, o que deixa perceber que acontece troca de informação intrafrásica, não havendo ainda concordância interfrásica, isto é, concordância entre sujeito e verbo. Na fase T4 verifica-se o processamento-S completamente desenvolvido, o que não acontece no PA. Por fim, no estágio T5 já é possível ao falante intercalar frases.<sup>1</sup>

Na Tabela 3.1 é possível observar a articulação dos processamentos que correspondem aos estádios de desenvolvimento da sintaxe mencionados anteriormente:

**Tabela 3.1.** Articulação dos processamentos na aquisição da interlíngua inglesa (Plag, 2008a:124).

<b>Estágio</b>	<b>Articulação do processamento</b>	<b>Processamento da L2</b>	<b>Morfologia</b>	<b>Sintaxe</b>
<b>T1</b>	Palavra / lema	Palavras	Formas invariáveis	Constituinte único
<b>T2</b>	Processamento categorial	Morfemas lexicais; Pronomes possessivos	Plural dos nomes	Ordem canónica
<b>T3</b>	Processamento frásico	Troca de informação intrafrásica	Concordância no SN	Frases iniciadas com ADV / “do”; Topicalização
<b>T4</b>	Processamento-S / ordem da frase	Troca de informação interfrásica	Neg+ V	Inversão “yes/no”; Inversão da cópula
<b>T5</b>	Processamento-S / ordem da frase	Troca de informação interfrásica	Concordância SV (3SG –s)	“Do” / Auxiliar 2ª posição
<b>T6</b>	Processamento das orações subordinadas	Orações principais e subordinadas	—	Cancelamento da inversão

A pluralização dos nomes acontece em T2, enquanto a concordância a nível do SN ocorre na fase T3, ou seja, quando se inicia a troca de informação gramatical intrafrásica. Por seu lado, a concordância sujeito-verbo apenas se verifica no estágio T5, isto é, numa fase em que a troca de informação interfrásica já está consolidada. Observando a distinção entre flexão inerente e flexão contextual, verifica-se que a falta



de flexão é uma característica da fase T1, mas que se pode prolongar pelos estágios seguintes, dependendo do tipo de flexão que se pretende observar. A flexão inerente instancia-se a partir do estágio T2 (p.e. pluralização dos nomes nus), enquanto a aplicação da concordância requer a articulação de processamentos posteriores, acontecendo unicamente em estágios mais avançados.<sup>64</sup>

Diversos estudiosos têm testado os princípios da Teoria da “Processabilidade” em distintas L2’s com falantes de diferentes L1’s, confirmando o processamento e as hierarquias do desenvolvimento morfossintático (Kawaguchi, 2005; Mansouri, 2005; Pienemann, 2005a; Pienemann, Biase & Kawaguchi, 2005; Pienemann *et alii*, 2005b; Zhang, 2005). Este aspecto dá consistência aos seus princípios, validando a sua aplicação nos estudos que visam observar a forma como surge e se desenvolve a morfossintaxe flexional de todo o tipo de línguas, incluindo as adquiridas em contexto massivo ou emergentes em situação de transmissão linguística irregular.

#### **2.4.2.3. O Modelo dos Quatro Morfemas (*The 4M Model*)**

Apesar de, presentemente, as teorias sobre ASL versarem mais sobre a essencialidade da GU na ASL, o Modelo dos Quatro Morfemas, baseada no modelo psicolinguístico de Levelt (1989:182) e avançado por Myers-Scotton & Jake (2000a, 2000b) e Myers-Scotton (2002), tenta descrever a produção normal e espontânea dos adultos. Exaustivamente descrita e analisada por Lopes (2001:48-59) e Godinho (2005:84-86), esta teoria procura descrever o modo como se organiza o léxico mental e a forma como se acede a ele para se conseguir a produção linguística. Em termos de ASL, o modelo prevê que os morfemas de conteúdo da LA sejam adquiridos antes dos morfemas de sistema. Por outro lado, tanto os morfemas de sistema semanticamente mais relevantes como os de realização morfológica na L1 que se assemelhem aos seus correspondentes na LA são igualmente adquiridos antes dos morfemas menos salientes, ou que ocorram nas estruturas de superfície apenas para preencher requisitos estruturais da L2 (Wei, 1996:91).

Lopes (2001), P. Andrade (2003) e Baxter (2004) chamam a atenção para as demonstrações de Myers-Scotton (2001) acerca do modo como a CPL-var no SN envolve o recurso aos morfemas sistémicos estruturais (*early system morphems*) e aos

morfemas de nível funcional sistémico (*late system morphems*), os quais originam a redução de estruturas gramaticais durante a fase da aquisição (cf. nota de fim de texto 17). De facto, no caso de algumas línguas crioulas, e tal como analisaremos mais detalhadamente no ponto 4.1.2.6.2 do presente trabalho, alguns elementos do português, caso dos artigos definidos, poderão, de acordo com o sistema de pluralização das línguas africanas do grupo níger-congo atlântico, ter sido fonologicamente interpretados como prefixos classificadores dos nomes e, depois, transpostos para as estruturas morfosintáticas do PtgL2 nos seus estágios iniciais de aquisição em situação de transmissão linguística irregular.

### **2.4.3. Modelo da mudança evolucionista da língua**

De acordo com os defensores deste modelo (Givón, 1998; Croft, 2000; Mufwene, 2001; Clements, 2009), a evolução da língua estabelece analogias com a evolução das espécies, constituindo-se de traços lexicais e estruturais (análogos aos genes em biologia) e sendo os conjuntos desses traços (*feature pools*) usados para codificar a comunicação humana (relações comunicativas) no seio das comunidades de falantes, ou seja, entre elementos que interagem rotineiramente num certo grupo, recorrendo, para tal, a uma determinada língua. Neste aspecto, e à imagem daquilo que é defendido pelos conexionistas, não existe um módulo neurológico específico para a aprendizagem da linguagem, pelo que esta não se configura como um conjunto de regras abstractas (Gass & Selinker, 2001:216), sendo antes determinada pelo comportamento de seres humanos individuais, enquanto elementos integrantes de grupos utilizadores de um idioma específico (Clements, 2009:210).

Para Givón (1984:41-42), à linguagem humana competem duas funções primordiais: (i) representar o conhecimento; (ii) transmitir o conhecimento representado. Partindo destes pressupostos, o autor propõe que a linguagem seja vista em termos de aproximação à evolução-adaptativa manifestada biologicamente, uma vez que a estrutura e a função da primeira, a exemplo do que sucede com as componentes da segunda (anatomia e fisiologia), se complementam e não podem ser observadas em separado. Por seu lado, o processo que decorre entre o estado FL0 e o estado FLf da aprendizagem da língua é consequência de uma incrementação gradual da aquisição.

Assim sendo, a língua é um sistema dinâmico e em constante mutação, que tem a finalidade de representar o conhecimento face ao propósito de comunicar. Em tal sistema coexistem, então, variadas estruturas em determinado momento da sua etapa evolutiva (Pienemann, 1998:7).

Não obstante, as distinções não se estabelecem apenas em termos de linguagem, já que são também extensíveis aos sistemas de comunicação. Efectivamente, em termos de propriedades estruturais, funcionais e cognitivas, tais sistemas evidenciam diferenças entre os estágios de comunicação pré-gramatical – a interacção é feita com recurso a léxico codificado e combinatórias de regras rudimentares – e de comunicação gramatical – a interacção é estabelecida com recurso a léxico e regras gramaticais elaborados –. O desenvolvimento que permite a evolução do código pré-gramatical ao gramatical só é possível com base nas repetições que acontecem no discurso, sendo este essencial para a compreensão da gramática (Hopper, 1987, 1988, 1998).

Os padrões de reconhecimento e o papel que desempenham na forma, função e mudanças da língua, envolvem a noção de protótipo porque, com recurso a ele, é possível perceber quer a frequência das ocorrências no uso da língua quer o modo como as estruturas e classes emergem nesta. Estudos estatísticos sobre a estrutura de orações absolutas no espanhol (López Meirama, 1997) evidenciam que cerca de 85% dos sujeitos em posição inicial são altos na hierarquia de definitude (SN's realizados como nomes próprios, pronomes pessoais ou nomes com um determinante deítico), enquanto 90% dos sujeitos não-iniciais são baixos na hierarquia (SN's sem determinante). Por outro lado, observações sobre a ordem dos constituintes no castelhano medieval mostram que os casos de sujeito pospostos são baixos na hierarquia, uma vez que só ocorrem quando o objecto é um demonstrativo ou um relativo (Suárez Fernández, 2008:293, nota 86). Portanto, no caso do espanhol, o tópico protótipo é alto em definitude, tal como a ordem protótipo do sujeito é alta pré-verbalmente.<sup>65</sup>

Croft (2000) e Mufwene (2001) constroem a noção de língua como espécie assente em propriedades comportamentais, da mesma forma que as espécies são definidas em biologia. A aproximação é feita sob a perspectiva genética das populações, considerando-se os aspectos sociohistóricos que determinam variação como a ecologia das línguas. No entender de Croft (2000:13-15), tal como as espécies irmãs isoladas

(*sibling species*), que se reproduzem separadamente, mas não se distinguem estruturalmente, e as espécies politípicas (*polytyp species*), que são estruturalmente distintas entre si, também uma língua detém propriedades estruturais essenciais e imutáveis que a identificam, distinguindo-a ou aproximando-a de outras. Nesta conformidade, independentemente da possibilidade de as espécies linguísticas se reproduzirem isoladamente, não são definidas por traços estruturais abstractos, mas antes por propriedades fundamentais compartilhadas pelos membros da sociedade que a utilizam.

A classificação taxionómica das espécies é baseada em semelhanças e diferenças dos traços estruturais das línguas, enquanto a classificação filogenética corresponde ao quadro teórico que caracteriza as populações. Desta forma, a associação filogenética de duas populações acontece em termos históricos, uma vez que a mesma só é possível caso algum traço inovador venha a ser compartilhado por ambas (Croft, 2005:15). Os traços que distinguem ou aproximam duas línguas podem ser do tipo cultural e lexical (p.e. o hindu e o urdu são dialectos idênticos da mesma língua, mas os seus falantes distinguem-se por questões religiosas), do tipo cultural, social e político (p.e. os falantes do mandarim e cantonês compartilham o mesmo sistema político, cultural, social e de escrita, mas falam diferentes dialectos da mesma língua) ou até de identidade nacional e política (p.e. o alemão e o holandês falados na fronteira dos dois países são mutuamente inteligíveis pelos habitantes raianos, apesar de serem assumidos como dialectos de línguas distintas, isto é, como variedades do alemão *standard* e do holandês *standard*).

No caso concreto do PE e do PA, duas línguas politípicas da família romance, as diferenças são de vários níveis, mas mais na oralidade do que na escrita. Fonologicamente, o sistema de sílabas átonas é mais reduzido no segundo do que no primeiro, enquanto as consoantes apicais [t], [d], [s] e [z], antes das vogais [+ altas] [ə] e [i], são palatalizadas em [k], [g], [j] e [ç] no PA (Ferraz, 1979:22). Por outro lado, na morfossintaxe dos almorarifanos acontece variação na concordância de género e número no SN, enquanto os pronomes pessoais sujeito podem ser usados com função de objecto (exemplo [148]). Também a colocação dos clíticos varia nas duas línguas, uma vez que, no PA, podem ser colocados pré-verbalmente (exemplo [151]), enquanto no PE o são pós-verbalmente. Outra das distinções entre as duas línguas verifica-se a nível do

sistema de conjugação verbal, mais elaborada no PE do que no PA (exemplos [136], [137] e [138]). Note-se que as diferenças se estendem ainda ao plano lexical (exemplo [145], nota de fim de texto 46). No entanto, apesar das muitas distinções entre as duas variedades, ambas são consideradas como elementos integrantes de uma única língua, em virtude de Portugal e São Tomé e Príncipe deterem traços culturais e sociais estreitos. Por outro lado, se se levar em consideração que as características apontadas para o PA são também válidas para determinadas variedades do PB, então estas configurarão casos de espécies irmãs isoladas (*sibling species*) relativamente à fala de Almoxarife.

Assim, a interação comunicativa depende não apenas das similaridades estruturais das variedades faladas mas também do comportamento social dos falantes. Este aspecto determina que a definição de língua assente mais em critérios sociais do que linguísticos (Chambers & Trudgill, 1980:3). Nesta perspectiva, se a espécie biológica é definida como um conjunto de indivíduos isolados que se reproduzem, por um lado, e se a espécie linguística é caracterizada como um conjunto de indivíduos isolados que comunicam entre si, por outro lado, então as noções biológicas de raça e *deme* geográficas corresponderão às redes sociais de um determinado grupo de indivíduos isolados que comunicam entre si, e não tanto com indivíduos externos à sua rede social (Clements, 2009:8).

Para Mufwene (2001:151-152), a língua é como um parasita, ou seja, um género de espécie simbiótica, que necessita de um hospedeiro (o falante) para com ele poder coexistir. Em adição, a transmissão linguística tem relevância secundária no seu modo vertical, ou seja, de pais para descendentes, mas primordial na sua forma horizontal. Esta transmissão assemelha-se à das plantas consideradas superiores, ou seja, à que ocorre de forma poliplóide, levando as línguas a poderem ter uma origem múltipla, marcada por estágios iniciais incipientes na sua génese, como acontece com os pidgins, que geralmente se desenvolvem de modo dinâmico e recorrente em apenas duas ou três gerações. Paralelamente, este desenvolvimento é acompanhado de mudanças genéticas e epigenéticas, conduzindo a uma extensa reestruturação em todos os níveis do genoma, e que originam repadronização, novos padrões de expressão genética, invasão intergenómica e evolução coordenada. Um exemplo de que a transmissão vertical não é

relevante para a evolução linguística é concedido pelo pidgin falado pelos africanos transplantados para o sudoeste de Portugal nos finais do século XV e princípios do século XVI, que não evoluiu para crioulo (Clements, 2009:43-48). Daí inferir-se também que, muitas vezes, não são as crianças a fonte das inovações que activam a variação conducente à mudança. Por conseguinte, a emergência da variação em L1 e L2 ancora em factores diversos e envolvendo vários tipos de desencadeamento (*bootstrappings*).<sup>66</sup>

Tendo em conta que os sistemas complexos emergem de modo simples e que, pouco a pouco, vão desenvolvendo complexidades (Ellis & Schmidt, 1997), conclui-se que as estruturas sintácticas se processam com base no léxico. Trabalhos sobre aquisição verbal (p.e. Bresnan, 1982, 2001; Langacker, 1987; Goldberg, 1995, 1999) têm evidenciado a forte relação entre significado, por um lado, e estrutura argumentativa dos verbos e estrutura sintáctica, por outro lado. A relação entre aquisição vocabular e aquisição gramatical é tão evidente que, segundo os defensores do modelo evolucionista, a segunda não será viável de modo independente, antes configura um desenvolvimento do léxico, sendo este possível apenas através da frequência e rotina de uso dos verbos em questão. Por conseguinte, a gramática emerge de modo primário durante a aquisição, sendo afectada pela combinação do *input* linguístico, pelas exigências das situações comunicativas e por constrangimentos cognitivos.

No caso específico da ASL, estes aspectos aliam-se aos sistemas de marcação universal para orientarem o tipo de desenvolvimento e forma que determinado pidgin ou crioulo vai evidenciar (Clements, 2009:26). Estudos levados a cabo sobre a escolha das formas copulativas em CP's e crioulos de base lexical espanhola revelam que os falantes dão primazia às estruturas do tipo CV (p.e. a forma *sã* em vez de *é*, no maquista), em virtude da sua força lexical (Clements, 2009:22). Neste caso, a saliência perceptual impõe-se à frequência na LA, fazendo com que uma estrutura não-marcada universalmente, por ser mais facilmente processável (Thomason & Kaufman, 1988:331), seja responsável pela emergência da estrutura silábica de uma variedade que nasce por contacto linguístico. Por outro lado, as relações de dominância entre os grupos de falantes de diferentes línguas em contacto podem também ser determinantes para que traços relativamente marcados, como a flexão morfológica, se instalem nas variedades

submetidas a contacto. Estes aspectos conduzem-nos, assim, às questões das transferências e empréstimos (Thomason & Kaufman, 1988:73-74) por nós debatidas nos pontos 1.1 e 1.3.2.3.1 do presente trabalho, já que ambas se reflectem na actual estrutura sintagmática do PA.

Após os considerandos tecidos em torno dos pressupostos teórico-metodológicos sobre aquisição de L1, aquisição bilingue e ASL, passaremos agora a dirigir a nossa atenção para os aspectos que determinam a concordância PL entre os elementos do SN no português, antes de abordarmos as questões em torno da aquisição massiva de L2's e, conseqüentemente, do português pelos falantes da comunidade de Almojarife.

## **2.5. Variação: pressupostos teóricos**

A Sociolinguística visa o estudo dos padrões de comportamento linguístico observáveis dentro de uma comunidade de fala, formalizando-os analiticamente com recurso a um sistema heterogéneo, constituído por unidades e regras variáveis (Lucchesi & Ribeiro, 2009:127; Lucchesi & Araújo, 2010). Partindo-se destes princípios, tem-se buscado uma resposta satisfatória para a questão da mudança linguística, enquanto fenómeno que alia determinados pressupostos incontornáveis (Labov, 1972a, 1974, 1982, 1994; Weinreich, Labov & Herzog, 2006 [1968]): (i) os sistemas linguísticos são heterogéneos e plurais, a exemplo do que acontece com as comunidades que os usam; (ii); não há mudança sem variação, mas pode ocorrer variação sem mudança; (iii) conseqüentemente, os padrões linguísticos registadas nos distintos momentos da variação reflectem o estágio e percurso da mudança em curso. Nesta conformidade, os elementos de uma determinada comunidade, apesar de conjugarem, entre si, normas e atitudes face ao uso da fala comunitária, não recorrem todos às mesmas formas linguísticas. Estas, no entanto, não deixam de compartilhar traços, os quais, por seu lado, marcam diferença relativamente aos traços utilizados noutras comunidades.

A exemplo do que sucede com qualquer outra língua, o PA regista também formas que correlacionam aspectos linguísticos e sociais usados heterogeneamente pelos seus falantes, isto é, pelos membros que compõem a pequena comunidade de Almojarife, São Tomé. Deste modo, determinados usos linguísticos produzidos pelos almojarifanos, ainda que distintos entre si, apresentam equivalência semântica a nível de vocabulário,

da sintaxe e morfossintaxe, do subsistema fonético-fonológico e no campo pragmático-discursivo (cf. Mollica, 2003a:9). A presença e ausência de marcas de concordância nominal em idêntico tipo de construções (exemplos [152a] e [152b]) ou de fonemas que podem ter diferentes realizações fonéticas que se alternam no mesmo contexto linguístico (exemplos [211] e [212]) ilustram bem a variabilidade linguística patente em Almojarife, sendo tais alternâncias sistemáticas, as variantes linguísticas, determinadas pela pressão que factores estruturais (as variáveis linguísticas) quanto sociais (as variáveis extralinguísticas) exercem, em simultâneo, sobre os usos da língua, aumentando ou diminuindo a sua frequência de ocorrência e impedindo que os mesmos sejam aleatórios (Tarallo, 1986:8). Consequentemente, estes usos em co-ocorrência e concorrência revelam a capacidade que a heterogeneidade linguística possui para não produzir alterações significativas nas mensagens que o locutores pretendem transmitir aos alocutários. Não obstante, os condicionalismos que actuam sobre as variantes em competição poderão determinar que uma delas apresente um maior uso em determinados contextos, podendo este ser analisado e previsto com recurso a metodologia estatística cientificamente apropriada.

Ao seleccionarmos como objecto de estudo a análise da variação registada num idioma que tem na génese da sua emergência uma situação de contacto entre línguas, estamos não só a considerar como socialmente relevante a linguagem da pequena comunidade de Almojarife mas também a atribuir à sua variação um papel que pode ajudar a dar contributo valioso para os estudos sociolinguísticos, nomeadamente no que concerne às questões relacionadas com a mudança linguística e factores que a determinam. Para tanto, procuraremos fazer a descrição estatisticamente fundamentada da CPL-var no SN do PA, isto é, tentaremos analisar, apreender e sistematizar como este fenómeno se caracteriza de acordo com as propriedades da língua, por um lado, e se compromete, a nível do seu *status* positivo ou negativo, a nível do sistema linguístico da comunidade, por outro lado. Para procurarmos entender este grau de comprometimento, calcularemos a influência que cada factor, interno ou externo ao sistema linguístico de Almojarife, possui na realização da CPL-var no SN do PA. Estes aspectos permitirão detectar se as variantes em competição na realização da pluralização a nível do SN configuram um caso de mudança em processo (verifica-se uma tendência de



predominância de uma variante sobre a outra) ou se, pelo contrário, manifestam uma situação de variação estável (o quadro da variação tende ainda a manter-se por longo período de tempo) ou de recuo na inovação. Portanto, para se concluir sincronicamente acerca do processo de mudança em diacronia que estará, ou não, actuando sobre o fenómeno que nos propomos analisar, será necessário avaliar o grau de estabilidade ou mutação patente nos usos alternativos (*evaluation problem*, cf. Labov, 1972a, 1982; Weinreich, Labov & Herzog, 2006 [1968]), detectando-se, com exactidão, quais as variáveis internas à língua (p.e. fono-morfossintácticas ou lexicais), inerentes ao indivíduo (p.e. idade ou sexo), de tipo social (p.e. escolarização) ou meramente contextuais (p.e. tensão discursiva), que estarão afectando sistematicamente a variante, no sentido de a difundir e implementar entre os falantes almoxarifanos (*actuation problem*, cf. Labov, 1972a, 1982; Weinreich, Labov & Herzog, 2006 [1968]). Ainda assim, é também importante realçar que a complexidade dos condicionamentos actuantes sobre os usos que determinam a variação contínua da CPL-var no SN do PA não possibilita que se prevejam todos os tipos de agentes que estabelecem correlação com as suas variantes linguísticas, ou seja, não é possível demarcar nitidamente as fronteiras em que a referida variação ocorre.

No PA, a estigmatização linguística e o grau de monitoramento maniqueísta certo/errado exercido sobre alguns dos seus falantes complementaram-se no sentido de parametrizarem determinadas formas linguísticas, que tinham sido adquiridas informalmente, ao encontro do padrão da variedade de português mais culta falada na ilha. No entanto, concorrendo com estas são também realizadas algumas variantes motivadas, que constituem inovações relativamente ao PE. Estas, por seu lado, não deixam também de ser realizações coesas, configurando padrões estruturais e estilísticos que estão na base da caracterização do idioma usado pela comunidade almoxarifana. Portanto, estamos perante variações que, de certa forma, escaparam à monitorização e são usadas regularmente num eixo diatópico e diastrático marcado por grau considerável de isolamento geográfico e social. Desta forma, as variações acabaram por se implementar com padrões cujas tendências são preditivas e mensuráveis, já que não deixaram de se estruturar de acordo com as propriedades sistémicas das línguas, que

impõe também restrições categóricas às variantes, impedindo que as mesmas se tornem aleatórias (Tarallo, 1986:8).

No presente trabalho, a Teoria da Variação abordada instrumentaliza a análise sociolinguística e quantitativa do tipo laboviano, já que a mesma se tem revelado coerente e metodologicamente eficaz em diversos estudos anteriores que pretenderam efectuar descrições linguísticas numa perspectiva idêntica à que adoptámos. Todavia, e como temos vindo a sustentar no presente capítulo, não deixaremos de ter igualmente em consideração outros pressupostos teóricos, conquanto os mesmos dêem contributo importante para a análise que pretendemos levar a cabo.

No caso concreto dos factores estruturais, por exemplo, dever-se-á ter em conta o seu encaixe na estrutura relevante (cf. Naro, 2003a:16), o qual poderá originar a falta de concordância de número quando acontece maior ou menor diferença fónica entre as formas singular e PL. Face a questões como esta, o desafio em Teoria da Variação será determinar o quanto cada um deste tipo de oposições concorre para a realização de uma ou de outra variante em competição. Os dados orais do nosso *corpus* exibem categorias morfológicas em que o fenómeno da concordância PL ocorre com ou sem marca formal em diferentes posições na cadeia sintagmática, dificultando a verificação do efeito das ditas categorias de um modo isolado, isto é, sem se recorrer, em simultâneo, à observação do seu posicionamento na estrutura frásica. Este aspecto alerta para a impossibilidade de se medir o efeito de uma determinada categoria na CPL-var do SN do PA sem levar em consideração o efeito de outras categorias que, com ela, concorrem para a motivação do fenómeno em observação.

### **2.5.1. Variáveis extralinguísticas, variação e mudança**

Agentes externos à língua, como marcadores regionais de comunidades geograficamente circunscritas ou formas usadas especificamente por determinados estratos sociais, detêm um papel importante no controlo da variação, impedindo que esta seja caótica e aleatória, por um lado, para se tornar regular, sistemática e previsível, por outro lado. Contudo, estas variáveis não operam isoladamente, antes co-actúan com as variáveis linguísticas e também com outras variáveis extralinguísticas. De entre as segundas, variáveis como a idade do falante, seu nível socioeconómico ou de

escolarização, ou ainda acesso que ele tem aos media, poderão determinar a ocorrência de distintos padrões linguísticos (popular e culto) no seio da comunidade em que está inserido, como sucede em Almocharife.

Embora sejam expressivos os estudos que correlacionam diferentes variáveis extralinguísticas com fenómenos de uso na fala de diferentes comunidades, o debate acerca dos seus efeitos na padronização do registo apresentado está longe de se encontrar esgotado. Observando o efeito que factores não linguísticos exercem sobre formas estigmatizadas do inglês vernacular de comunidades negras urbanas, Labov (1972b) concluiu que tais formas resultavam de preconceitos relacionados com as pressões étnica, social e escolar. Por seu lado, Sankoff, Kemp & Cedergren (1978), debruçando-se sobre estruturas alternativas do francês de Montréal, Canadá, com vista a determinar o maior ou menor grau de marcação social nas mesmas, verificaram que este era influenciado pela escolarização e estatuto profissional dos falantes. A importância das variáveis sociais na determinação de formas linguísticas usadas pelos mesmos falantes seria ainda coadjuvada e confirmada em outros estudos (p.e. Laberge, 1977; Clermont & Cedergren, 1979; Kemp, 1979, 1981).

Para além de observarem o efeito que factores sociais exercem sobre formas linguísticas, os estudos labovianos permitiram também que se fizessem inferências sobre o desenvolvimento diacrónico da língua, a partir dos resultados das análises sincrónicas, já que estas reflectem em tempo real, em um determinado ponto da estrutura da gramática da fala da comunidade, o processo da mudança em curso projectado no tempo aparente. O pressuposto da projecção do tempo real no tempo aparente é explicado por Chambers & Trudgill (1980:165) com recurso à variável social idade, alegando os autores que, por exemplo, a fala dos indivíduos de 40 anos de idade não só reflecte directamente, e no tempo actual, a produção dos indivíduos de 20 anos há 20 anos mas também pode ser comparada com a dos falantes de 20 anos da actualidade. Caso ocorram discrepâncias entre os usos das faixas etárias com 20 e 40 anos, essas discrepâncias reflectem o progresso registado nas realizações durante o período de tempo que separa as duas faixas etárias. Ainda assim, é preciso considerar que os sistemas linguísticos não são sempre perfeitamente estáveis nem totalmente predizíveis quanto às tendências de mudanças futuras, já que estas podem ser afectadas por

processos inexistentes no momento do estudo, mas que venham a ocorrer posteriormente (Labov, 1981:177).

Ainda no que concerne à variável social idade, podem registar-se duas situações (Chambers & Trudgill, 1980:91-93): (i) se as faixas etárias intermédias apresentarem o maior uso das formas inovadoras, o padrão curvilíneo denuncia estar-se perante uma situação de variação estável; (ii) se a aplicação da inovação incidir mais fortemente nos estratos etários mais novos, o padrão inclinado evidencia uma mudança em progresso. Todavia, estas inferências terão que ser sustentadas pelos resultados de outras variáveis sociais, como por exemplo, a escolarização.

Assim, no que às variáveis escolarização e/ou actividades de letramento diz respeito, normalmente associadas aos indivíduos das classes mais altas, tem sido confirmada a influência de ambas no sentido da aquisição da norma culta (p.e. Bortoni-Ricardo, 1994; Silva & Scherre, 1998 [1996]; Godinho, 2005; Silva, 2010), configurando-se situações de variação estável. Porém, alguns estudos têm revelado que o emprego de determinadas estruturas linguísticas por parte de falantes das classes média ou baixa, apesar de pertencerem a grupos mais escolarizados, pode apresentar flutuações, não se moldando ao sistema de ensino (Silva & Scherre, 1998 [1996]). Este aspecto revela não apenas que se está perante falantes que tendem a impulsionar processos de mudança mas também que, afinal, o efeito de indicadores sociais sobre o perfil sociolinguístico dos falantes é bastante mais complexo do que à primeira vista pode parecer (Mollica, 2003b:29). De facto, numa sociedade linguisticamente complexa, como a de São Tomé, serão imensos os indicadores sociais com influência nas formas de uso, competindo ao investigador seleccionar aqueles que mais impacto poderão ter sobre a variação que se pretende observar.

Nas sociedades mais industrializadas, a classe social parece ser o factor mais marcado linguisticamente (Chambers, 2002 [1995]), podendo os seus índices ser avaliados em função de aspectos como a ocupação, a educação ou o poder económico. Contrariamente, círculos sociais fechados, localizados e economicamente mais desfavorecidos, como acontece com a comunidade de Almojarife, tendem a incrementar relações familiares, de vizinhança ou de bairro. De qualquer forma, seja nuns seja nos outros, há tendência para se constituírem estereótipos, pelo que os índices de classe

social passam tanto por factores objectivos como por factores subjectivos. Paiva & Scherre (1999), em estudos pioneiros sobre usos do PB (PEUL – Programa de Estudos sobre os Usos da Língua), tentaram determinar a correlação estabelecida entre variáveis não convencionais, como a origem social, os bens materiais e os bens culturais, por um lado, e os usos linguísticos prestigiados e não prestigiados, por outro lado. Não obstante, os resultados levantaram a possibilidade de tais categorias não poderem ser medidas com recurso a critérios linguísticos, já que a língua é apenas mais um no conjunto dos elementos que constituem o património pessoal do indivíduo, enquanto membro integrante de uma determinada comunidade.

Do exposto é possível concluir que ocorrem incertezas quanto à possibilidade de certos factores extralinguísticos representarem uma opção válida como identificadores e determinadores do estatuto social de determinado indivíduo ou comunidade. Paralelamente, dúvidas parecem não existir no que concerne a outros factores, enquanto limitadores ou impeditivos da ascensão social do indivíduo e, conseqüentemente, de uma maior aquisição dos usos linguísticos de maior prestígio. Este é exactamente o caso da pequena comunidade de Almojarife, submetida a uma situação de isolamento cada vez mais acentuado, a qual determina que aspectos como as condições miseráveis de habitação, o acesso precário à saúde e escolarização ou a fome impeçam, aos seus membros, o acesso ao estatuto de cidadãos plenos.

### **2.5.2. Variáveis linguísticas e contexto**

Os trabalhos pioneiros de Labov (1972a, 1972b) sobre a centralização de ditongos em Martha's Vineyard, costa ocidental dos Estados Unidos, e sobre o uso do /r/ em Nova Iorque, com recurso a dados colectados em situações reais de comunicação, concederam primazia à observação do modo como condicionamentos externos à língua interferiam na produção dos referidos fenómenos. Centralizando-se atenção nos factores sociais, com recurso a métodos de análise quantitativa, buscava-se não só uma teorização oposta à preconizada pelos princípios generativistas mas procurava-se evidenciar também que os usos linguísticos das comunidades são heterogéneos. Nos estudos referidos, notava-se também uma apetência pela selecção de fenómenos morfofonológicos como objecto de observação. Este aspecto justificava-se, já que, em colectas de dados reais, os sons e

fonemas se aprestam como entidades que garantem um número de realizações mais elevado do que o de qualquer outra entidade linguística. Paralelamente, os estudos com recurso a variáveis de natureza fonológica vinham já de longa data, sendo uma constante em trabalhos estruturalistas, embora sem recurso a métodos quantitativos.

Tendo-se evidenciado eficientes para a observação da variação sistemática e quantificável dos fenómenos fonológicos, os estudos labovianos deram o mote para a análise, na mesma linha, da variação registada na sintaxe, na semântica e até no discurso, áreas onde, apesar de tudo, os fenómenos não são tão recorrentes como na fonologia. Além do mais, a busca de formas sintácticas alternantes esbarrava, segundo alguns autores, na impossibilidade de se manter idêntico significado em ambas as realizações. Lavandera (1984) propõe, então, que se tratem as variáveis linguísticas levando em conta o enfraquecimento da equivalência semântica e dando primazia à “comparabilidade funcional”. Portanto, frases que transmitem a mesma intenção comunicativa, ainda que distintas (p.e. *Está calor aqui dentro.* > *Por favor, ligue o ar condicionado.*) devem ser tratadas como variantes de uma mesma variável. Contudo, outros estudiosos dos fenómenos da variação (Weiner & Labov, 1983) entendem que as condições para análise das variantes sintácticas se devem restringir ao valor referencial, como sucede nas alternâncias das vozes activa e passiva, em que a noção de verdade não sofre alteração.

Seja num caso seja no outro, o estudo variacionista ver-se-á sempre confrontado com diferenças conectadas aos valores semânticos, aos contextos discursivos e ao tipo de interacção entre locutor e alocutário. Todavia, tais diferenças poderão ser controladas com recurso a factores que o estudioso entende estarem correlacionados com o fenómeno em análise. Portanto, é possível determinar os referentes portadores de informação nova ou velha, testando o peso que exercem no uso das formas comunitárias e aquilatando a sua influência na mudança em curso. Ainda assim, tal só poderá acontecer determinando se o traço discursivo ou semântico a observar constitui uma das variáveis admitida como escolha alternativa num mesmo contexto, isto é, a variação.

A questão da escolha contextual ganhou projecção em análises variacionistas, sendo determinante para o estudos das variantes sintácticas, o que significa dizer que estas não podem ser observadas a partir de produções linguísticas que configuram realizações

isoladas (Klein-Andreu, 1983; Bentivoglio, 1987; Kroch, 1989a). Estruturas em que acontece o deslocamento de um SN para o início da frase ou ocorrem realizações com marcadores discursivos (p.e. *sabe?*, *assim*, etc.) dificultam a decisão sobre o que constituirá, de facto, o conjunto de variantes a analisar. Nesta conformidade, estudos recentes sobre variação não se limitam a seleccionar fenómenos gramaticais de contornos perfeitamente marcados, antes observam domínios abrangentes. É o que sucede com variantes que se manifestam através de distintas realizações linguísticas, (p.e. *nós* > *a gente* > *a malta* > *o pessoal*). Nestas situações, não se estabelece “*apenas uma escolha entre duas alternativas, mas uma escala de possibilidades para aquilo que poderíamos chamar de um domínio funcional, e aí estudar a variação. Neste caso, o significado referencial seria preservado, mas a questão do contexto exigiria um cuidado especial, ao levar em conta a situação discursiva*” (V. Silva, 2003:71).

Em virtude de enxertarem um no outro, significado e contexto determinaram que estudos variacionistas do PB refinassem a observação das variantes sintáticas e discursivas, levando a avanços consideráveis nas áreas da linguística textual e linguística funcional, as quais seleccionam fenómenos de âmbito e estrutura superiores ao das sentenças: (i) o *status* informacional dos itens lexicais (Braga, 1984; Berlinck, 1989) e oracionais (Paiva, 1991); (ii) aspectos da coesão textual (Omena, 1978, V. Silva, 1988), relacionados a fenómenos de nível morfossintáctico e oracional; (iii) relação semântica de contraste, aplicada a fenómenos morfossintácticos (Mollica, 1984; V. Silva, 1988), oracional (Oliveira & Braga, 1997) e discursiva (Gryner, 1990). Paralelamente, o interesse pelos fenómenos não-fonológicos internos à língua ganhou novo fôlego à luz da Teoria da Variação, até porque estes constituem um desafio acrescido, quando se trata de os correlacionar às variáveis sociais. Quanto aos resultados, têm evidenciado que este tipo de fenómenos não está tão sujeito à influência das variáveis sociais, alertando, consequentemente, para o facto de a sua variação advir mais da pressão que é exercida pelos níveis sintáctico, semântico e discursivo-pragmático (Braga, 2003).

### **2.5.3. Modelos matemáticos para reprodução do efeito global dos dados empíricos**

Nos estudos pioneiros sobre CPL-var no SN do PB (Braga, 1977; Scherre, 1978), postulou-se que a maior distinção fónica singular/plural favorecia a inserção de marca

formal de PL, um aspecto que, no entanto, era contrariado pelos valores estatísticos enviesados, já que o percentual de marcação era bem mais elevado nas formas em que a referida oposição era simples, como acontecia com os artigos ou demonstrativos (p.e. *o>os* ou *este>estes*). Assim, só observando a forma como as distinções fónicas co-actuaem com a posição linear de cada elemento flexionável do SN é que foi possível perceber o modo como ambos os factores intervêm na configuração da regra variável de marca de PL. De facto, as altas frequências reveladas pelas oposições simples estavam a inflacionar os resultados em virtude de se encontrarem concentradas na primeira posição da cadeia linear e não por razões que pudessem ser conectadas à sua estrutura morfológica. Removendo os dados das ocorrências nesta posição foi então possível confirmar que, de facto, a maior distinção fónica singular/plural favorece a inserção de marca formal de PL e que, conseqüentemente, *“as frequências brutas, embora concretas e intuitivamente bastante ‘reais’, podem ser falaciosas, porque seu cálculo não leva em conta as inter-relações existentes entre as categorias que atuam numa regra variável”* (Naro, 2003a:19).

As tentativas pioneiras de prover um modelo matemático consistente com os dados linguísticos iniciaram-se com o modelo aditivo proposto por Labov (1969), no seu clássico estudo sobre a variação patente no uso da cópula e do verbo auxiliar *“to be”* por parte de falantes do inglês registado nos *“ghettos”* urbanos de diversas comunidades negras. No modelo laboviano, a frequência geral da marca em observação procurava representar a soma dos factores contextuais, isto é, a co-actuação dos factores oposição simples, posição, número total de categorias relevantes e média global de aplicação da variante sob estudo.

Posteriormente, Naro (1981) testaria o modelo para analisar a concordância variável na fala de adultos cariocas alfabetizados. Neste estudo, a média global de aplicação da variante servia como ponto de referência para os diversos factores, cada um dos quais registando maior ou menor efeito na regra da concordância, ou seja, aumentando ou diminuindo a frequência da variante em determinado contexto. Não obstante, o modelo representa apenas uma soma de números, que pode ultrapassar os 100% ou ir abaixo dos 0%, aspecto que inviabilizaria a sua aplicação na elaboração de uma hipótese que



definissem, com exactidão, a força da actuação conjunta de categorias presentes num dado contexto, de modo a reproduzir o efeito global da variação.

Antes de Naro (1981) ter testado o modelo, Cerdergren & Sankoff (1974), partindo do pressuposto que a *performance* é modelada pela conjugação de factores generativos e extralinguísticos, desenvolveram um modelo matemático a partir das propostas de Labov (1969), que visava testar, em registos do espanhol, francês e inglês, as frequências previstas pelas regras variáveis. O modelo, denominado de aplicativo de multiplicação, propunha que a observação da variação se efectuasse a partir de probabilidades, e não de frequências. Assim, a variação representa o produto da actuação simultânea de factores independentes, e não a soma destes. No entanto, o modelo apenas se revelaria satisfatório para observar a actuação conjunta de dois factores altamente desfavorecedores (o modelo multiplicativo de aplicação das regras linguísticas), já que, para um estudo apropriado da co-actuação dos factores favorecedores, seria necessário recorrer a um modelo multiplicativo de não-aplicação das regras linguísticas.

Com vista a ultrapassar este impasse, Rousseau & Sankoff (1978) elaboraram um modelo logístico que conjuga propriedades dos três modelos anteriores (aditivo, multiplicativo de aplicação e multiplicativo de não-aplicação) e se destinava a observar, estatisticamente, o modo como as categorias estabelecem inter-relações em áreas distintas da linguística, como sucede na Biologia. Por outro lado, o modelo apresentava a vantagem de se encontrarem também respostas satisfatórias a partir da desvinculação dos factores entre si. Este aspecto revela-se importante porque a variação em direcção à mudança ocorre em virtude de, num determinado momento do tempo, uma das duas formas alternativas (a inovadora vernacular) passar a ser favorecida em detrimento da outra (a conservadora). Estudos empíricos de determinados fenómenos linguísticos (p.e. Silva, 1982; Kroch, 1989a) têm evidenciado que o processo de substituição conhece três etapas distintas, seguindo o trajecto de curva em  $\sim$ : (i) o estágio inicial de mudança, em que uma das ocorrências em uso se sobrepõe lentamente à outra; (ii) o estágio intermédio de mudança, em que a ocorrência que se sobrepõe evidencia um aceleração notório de uso; (iii) o estágio final de mudança, no qual a aceleração se retrai, até se consumir a mudança, traduzida no desaparecimento da forma em competição menos usada. Como o modelo logístico permite recorrer a probabilidades do

*input* (a média geral de aplicação da regra), é possível medir a tendência de presença da variante, isolando-a do efeito dos factores. Portanto, a metodologia da Teoria da Variação, adoptada no âmbito do presente trabalho, com recurso ao modelo de regressão logarítima múltipla VARBRUL (ponto 3.2 do presente trabalho), constitui uma ferramenta poderosa e fiável quando se trata de observar e descrever o dinamismo da variação sincrónica de determinados fenómenos linguísticos, enquanto retrato dos processos de mudança no tempo real. No entanto, as limitações da ferramenta poderão esbarrar nas do próprio pesquisador, caso este não detecte os factores relevantes, não levante e codifique acertadamente os dados empíricos e não interprete correctamente, à luz do quadro teórico disponível, os resultados obtidos.

Face ao exposto, temos que a caracterização de um processo de variação estável ou de mudança em curso independe dos resultados isolados de cada variável em observação. Nesta conformidade, para caracterizarmos a actual situação linguística do PA, no que concerne à CPL-var no SN, procuraremos ancorar a coerência argumentativa na representação que elaboraremos do processo como um todo, a partir das evidências empíricas que nos forem fornecidas pelos resultados das distintas variáveis (cf. Lucchesi, 2004b; Lucchesi & Araújo, 2010), as quais irão ser constituídas e pormenorizadamente descritas no capítulo 3. Deste modo, acreditamos vivamente que, superadas as dificuldades evidenciadas neste ponto do nosso trabalho, iremos também dar o nosso contributo para o avanço da Teoria Variacionista, já que a discussão em torno dos resultados alcançados com recurso ao tratamentos dos dados empíricos produzidos pelos informantes de Almojarife permitirá que aprofundemos igualmente questões de natureza teórica. Para tanto, não nos limitaremos a fazer dos factores internos um ponto exclusivo do nosso debate nem relacionaremos a sua variação restritamente às questões de nível sintáctico, semântico e discursivo-pragmático, uma vez que procuraremos identificar também até que ponto os factores sociais do tipo clássico (p.e. idade, sexo do informante ou anos de escolaridade) se aliam a estes para determinarem o uso da CPL-var no SN da fala de Almojarife, ou seja, para moldar o fenómeno que constitui o objecto principal da análise sociolinguística desenvolvida no âmbito do presente estudo.

## 2.6. Concordância de número PL a nível do SN

### 2.6.1. Posicionamento e flexão em número dos constituintes do SN: visões descritivistas

Para elaboração do ponto 2.6 e seus subpontos, tomámos como base de orientação o capítulo 3 do trabalho de Godinho (2005:112-165) e leituras sugeridas no mesmo (p.e. Pérez-Pereira, 1989; Cerqueira, 1994; López-Ornat, 1997; Lléo, 1997; 1998, 2001; Faria, Freitas & Miguel, 1998, 2001; Marrero & Aguirre, 2003; Name & Corrêa, 2003), tendo ambos sido complementados quer com dados e conclusões de outras leituras quer com inferências nossas. Relativamente às generalidades acerca da flexão em número do português, as mesmas estão contempladas em qualquer gramática prescritiva (p.e. Cunha & Cintra, 1997 [1984]:181-189, para os nomes; Cunha & Cintra, 1997 [1984]:252-253, para os adjectivos; Bechara, 1999 [1928]:117-131, para os nomes; Bechara, 1999 [1928]:145-146, para os adjectivos). Godinho (2005:112-119) lista alguns exemplos para confirmar que, no próprio português padrão, a marcação do PL com o sufixo *-s* acontece de forma aparentemente sistemática. A listagem, apesar de exaustiva, não é completa. Assim, a propósito da flexão em número de nomes e adjectivos, julgamos ser importante chamar a atenção para mais algumas especificidades da norma padronizada não contempladas na referida análise. Deste modo, nem todos os substantivos são passíveis de marcação PL em português, como acontece com os nomes que terminam em *-s* átono (exemplo [164a]). De igual forma, também determinados monossílabos com igual terminação se mantêm invariáveis (exemplo [164b]),<sup>67</sup> o mesmo não acontecendo com termos que finalizam em *-s* tónico (exemplo [164c]):

- [164] PTG:
- |    |                 |   |                    |
|----|-----------------|---|--------------------|
| a. | <i>o atlas</i>  | > | <i>os atlas</i>    |
| b. | <i>o cais</i>   | > | <i>os cais</i>     |
| c. | <i>o ananás</i> | > | <i>os ananases</i> |

Analisando o efeito do *contexto fonológico posterior* sobre a *saliência fónica* dos itens terminados em *-s* no singular, Braga & Scherre (1976), Lopes (2001) e Baxter (2004) referem que estes são responsáveis pela inibição da pluralização no elemento seguinte, já que os falantes fazem uma interpretação analógica do /s/ final com a marca de PL, deixando de aplicar o PL depois, por este se revelar redundante (Kiparsky,

1972:195). Contudo, este é um fenómeno que não ocorre no PA, como procuraremos demonstrar no ponto 5.2.3.10.1 do presente trabalho.

No português padrão, alguns nomes, classificados como não-contáveis, não possuem a noção de separação em partes diversas para uma determinada matéria ou para uma ideia abstracta. Recebem a denominação de *singularia tantum* porque, apesar de possuírem forma PL, só se usam, normalmente, no singular. Estão incluídos neste critério: (i) os nomes que pertencem a categorias semânticas que exprimem a ideia de matéria, como os metais (exemplo [165a]); (ii) os nomes que transmitem noções abstractas, como os pontos cardeais (exemplo [165b]), as qualidades do ser humano (exemplo [165c]), <sup>68</sup> as ciências (exemplo [165d]), as artes (exemplo [165e]) ou as religiões (exemplo [165f]); (iii) os nomes que, traduzindo a ideia de pluralização, fornecem também o conceito de não poderem ser contados, ou seja, os nomes massivos (exemplo [165g]):<sup>69</sup>

[165]	PTG:			
a.	<i>o ouro</i>	>	–	
b.	<i>o norte</i>	>	–	
c.	<i>a gentileza</i>	>	–	
d.	<i>a história</i>	>	–	
e.	<i>a pintura</i>	>	–	
f.	<i>o budismo</i>	>	–	
g.	<i>a relva</i>	>	–	

Os não-contáveis incluem ainda alguns nomes colectivos que, na forma singular, encerram a noção de pluralidade, fazendo referência a uma colecção ou conjunto de objectos. Entre os colectivos há duas categorias: os universais, que não são contáveis e só se pluralizam se houver necessidade de fazer referência à classe (exemplo [166a]); e os particulares, que se contam e são passíveis de pluralização (exemplo [166b]):<sup>70</sup>

[166]	PTG:			
a.	<i>O casario de Macau.</i>	>	<i>O casario de Macau e da Taipa.</i>	
b.	<i>A turma B do segundo ano.</i>	>	<i>As turmas B e C do segundo ano.</i>	

Alguns nomes são usados no singular com o valor de PL genérico, visto encerrarem representações abstractas de pluralização (Pereira, 1984:46), como acontece com as palavras *homem* ou *coisa*:

- [167] PTG:  
 a. *O homem é um ser inteligente.*  
 b. *A coisa agora está preta.*

Não obstante, e como referimos no ponto 2.6.1, Scherre (2001:98-99) adverte para a possibilidade de a representação abstracta de pluralização traduzir realizações com carga semântica de pluralidade que é percebida pelos falantes a partir do contexto, originando-se, assim, a não-realização da marca formal de PL. Sobre este tipo de nomes nos pronunciaremos, de novo, no ponto 4.1.2.4.6.

Outra situação de nomes usados no singular com o valor de PL genérico ocorre com os substantivos etnónimos (referentes a povos, tribos, castas ou agrupamentos em que é prevaiente o conceito de etnia), apesar de admitirem a forma PL:

- [168] PTG:  
*o bantu* > *os bantu*

Como apontámos no ponto 1.2.1.1, a pluralização destes nomes apresenta uma estrutura idêntica à das variedades africanas e brasileiras de português reestruturado. Contudo, no caso do português, não estamos perante uma marcação que define o padrão PL sistemático da língua, ao contrário do que sucede nas referidas variedades.

Continuando a observar formas de marcação plural no português, alguns empréstimos lexicais acolhidos por este idioma conservam também o PL original:

- [169] PTG:  
*o curriculum* > *os curricula*

Determinados nomes podem possuir duas ou três formas de PL, como acontece com muitos dos que terminam em *-ão*:

- [170] PTG:  
 a. *o vulcão* > *os vulcões* > *os vulcões* > –  
 b. *o sacristão* > *os sacristãos* > – > *os sacristães*  
 c. *o guardião* > – > *os guardiões* > *os guardiães*  
 d. *o aldeão* > *os aldeãos* > *os aldeões* > *os aldeães*

A análise da influência da variável independente *saliência fónica* na marcação PL no SN do MRJ revelou que os itens finalizados em *-ão* tendem a inibir a pluralização,

tendo Scherre (1988:124) atribuído esta à incerteza provocada nos falantes pelas distintas formas de se pluralizarem os referidos elementos. Estas questões encontram-se debatidas no ponto 5.2.1 do presente trabalho.

A pluralização de alguns nomes diminutivos do português, com recurso ao sufixo –*zinho*, apresenta grau de dificuldade acrescido para os falantes em aquisição do PtgL2, já que estes itens flexionam em número colocando os dois elementos no PL, mas suprimindo o –s do nome grau normal:<sup>71</sup>

[171] PTG:  
*o cãozinho* > *os cãesinhos*

Em português existem também nomes que, por norma, se utilizam unicamente na forma PL, isto é, os *pluralia tantum*:

[172] PTG:  
*as alvíssaras* > –

Fazemos ainda referência a alguns nomes, por norma concretos, cuja flexão em número acarreta uma alteração semântica do valor da palavra:

[173] PTG:  
*a costa* (orla marítima) > *as costas* (parte do corpo humano)

Com substantivos próprios<sup>72</sup> que configuram nomes e apelidos é possível a pluralização quando se deseja fazer referência a um conjunto de indivíduos que possuem um mesmo antropónimo (exemplo [174a]), ou se quer mencionar, não um indivíduo, mas vários que detêm características semelhantes à da individualidade referida (exemplo [174b]):

[174] PTG:  
a. *Os Antónios constituíram uma Associação.*  
b. *Ainda há muitas **Teresas de Calcutá** por esse mundo fora.*

Os apelidos, por seu lado, podem também ocorrer no PL, caso façam referência ao conjunto de indivíduos de uma determinada família (exemplo [175a]), ou se pretenda reportar a obras ou objectos produzidos pelo detentor do apelido que os elaborou ou manufacturou (exemplo [175b]):

[175] PTG:

- a. *Os **Freires** têm muitas casas em Leiria.*
- b. *O Museu de Arte Moderna tem muitos **Picassos** em exposição.*

De particular relevância para o nosso estudo são os nomes metafônicos, visto agregarem dupla pluralização, isto é, um PL morfémico (inserção de *-s* final) e um PL fônico (abertura vocálica), como acontece com os nomes do seguinte tipo: *jogo>jogos*, *ovo>ovos* ou *olho>olhos*. De facto, é comumente aceite que os falantes em situação de ASL revelam tendência para fazer incidir um maior peso de marcação PL sobre os itens fonicamente salientes. Estes aspectos encontram-se debatidos no capítulo 5, no qual procederemos à análise das variáveis fonológicas do PA.

No que respeita à flexão em número dos compostos morfológicos, remetemos a informação pertinente para Villalva (2003a:928-929). No entanto, convém referir que a flexão em género e número não se limita aos nomes, já que outros elementos que constituem o SN, como os indefinidos, os quantificadores ou os adjectivos, podem também apresentar este tipo de flexões.

Como no português é exigida a concordância entre os constituintes que compõem o SN (nomes, determinantes, quantificadores e adjectivos), os falantes de L1's africanas do grupo níger-congo atlântico ou do santomense, em aquisição do português, são confrontados com dificuldades na aprendizagem da regra, o que vai originar variação na interlíngua. No que concerne à concordância entre o núcleo e todos os constituintes do SN no PE, verifica-se que os determinantes e os quantificadores não colocam problemas maiores, reservando-se ao adjectivo algumas questões mais pertinentes. Godinho (2005:114-115) remete para Casteleiro (1981) as explicações detalhadas sobre a sintaxe do adjectivo, mas menciona as três situações em que se dá a ocorrência do adjectivo na frase: (i) adjectivos que ocorrem antes e depois do nome com alteração de sentido; (ii) adjectivos que só ocorrem antes do nome; (iii) adjectivos que só ocorrem depois do nome. Contudo, a propósito da primeira situação, importa também referir que a flutuação da posição do adjectivo, pré ou pós-nominal, nem sempre acarreta alteração semântica. De facto, adjectivos que não encerram um traço mais informativo específico (avaliativo), são colocados antes ou depois do nome apenas por mera questão estilística ou com a intenção de focalizar:

- [176] PTG:
- a. *A situação **actual** do Primeiro-Ministro é periclitante.*
  - b. *A **actual** situação do Primeiro-Ministro é periclitante.*

Sobre as outras duas situações referidas, constata-se que os adjectivos unicamente marcados por um traço mais qualitativo (conotativo) ocorrem antes do substantivo, como acontece com o adjectivo *mero* (Godinho, 2005:115). Contudo, neste aspecto discordamos com o outro exemplo avançado, ou seja, o do adjectivo *simples*,<sup>73</sup> já que este pode também ocorrer em posição pós-nominal:<sup>74</sup>

- [177] PTG:
- a. *Só gostava de iogurtes **simples**.*
  - b. *Uma **simples** testemunha bastou para o condenar.*

Por seu lado, os adjectivos de característica informativa, ou seja, denotativa, (por exemplo, os que se referem a nacionalidades ou cores), só podem ser posicionados pós-nominalmente (exemplos [178a] e [178b]), excepto em ocorrências literárias, já que estas subvertem regras sintácticas por questões estéticas:

- [178] PTG:
- a. *Uma mulher **chinesa** / \*Uma **chinesa** mulher.*
  - b. *Uma estante **metálica** / \*Uma **metálica** estante.*

Note-se que alguns destes adjectivos, se colocados pré-nominalmente, mudam de traço sintáctico (exemplo [179a]), passando a adjectivos substantivados que exigem a anteposição de um determinante. Classificados como modificadores internos, estes adjectivos possuem um traço formal rígido que obriga a que ocorram depois do nome. Caso tal não aconteça, deixam de ser itens subordinados para se tornarem o elemento central do SN (exemplo [179b]):

- [179] PTG:
- a. *O céu **azul**.*
  - b. *O **azul** do céu.*

Já as situações que envolvem o uso de dois ou mais adjectivos no SN implicam um posicionamento ordenado, no qual apenas um poderá ser pré-verbal. Contudo, caso ocorram dois adjectivos de traço informativo, ambos terão de ser situados depois do nome (exemplo [180a]). De facto, a colocação de qualquer dos adjectivos em posição



pré-nominal torna a frase agramatical (exemplos [180b] e [180c]). Quanto à ordem entre os adjetivos, é hierarquicamente determinada pelo maior ou menor grau informativo que encerram em si, surgindo em primeiro lugar o mais informativo. Assim, se a ordem hierárquica dos adjetivos for alterada, a frase torna-se agramatical (exemplo [180d]):

- [180] PTG:
- a. *O comboio **eléctrico amarelo**.*
  - b. *\*O **eléctrico** comboio **amarelo**.*
  - c. *\*O **amarelo** comboio **eléctrico**.*
  - d. *\*O comboio **amarelo eléctrico**.*

Nos casos em que houver um adjetivo de traço mais denotativo (p.e. *castanha*) e outro de traço mais conotativo (p.e. *pesada*), o segundo poderá ocorrer em posição pós (exemplo [181a]) ou pré-nominal (exemplo [181b]). Se optarmos pela primeira hipótese, o adjetivo de traço mais conotativo terá que obedecer à hierarquia formal, isto é ser colocado depois do adjetivo de traço mais informativo, pois o contrário torna a frase agramatical (exemplo [181c]). A colocação do adjetivo mais denotativo em posição pré-nominal produz também uma frase agramatical (exemplo [181d]):

- [181] PTG:
- a. *Uma pasta castanha **pesada**.*
  - b. *Uma **pesada** pasta castanha.*
  - c. *\*Uma pasta **pesada** castanha.*
  - d. *\*Uma **castanha** pasta **pesada**.*

Situações há em que se verifica também a presença de um adjetivo de traço informativo (p.e. *enjoativo*) e outro que não altera o traço semântico com a deslocação para a posição pré-nominal (p.e. *intenso*). Nestes casos, o adjetivo que não altera o traço pode ser usado antes (exemplo [182a]) ou depois do nome (exemplo [182b]). No segundo caso, e também por razões hierárquicas, ocupa o segundo lugar da estrutura em virtude de possuir um traço conotativo, o que não acontece com o adjetivo de traço mais informativo. A deslocação deste adjetivo para a posição pré-nominal (exemplo [182c]) ou pós-adjectivo que não altera o traço semântico (exemplo [182d]), torna as frases agramaticais:

- [182] PTG:
- a. *A sopa tem um **intenso** cheiro enjoativo.*
  - b. *A sopa tem um cheiro enjoativo **intenso**.*
  - c. *\*A sopa tem um **enjoativo** cheiro intenso.*
  - d. *\*A sopa tem um cheiro intenso **enjoativo**.*

O ordenamento de um SN em que ocorra a existência de um adjectivo modificador interno que não altera o traço semântico se colocado antes do nome (p.e. *política*) e um outro adjectivo que também não altera o traço semântico com a deslocação para a posição pré-nominal, mas que não funciona como modificador interno (p.e. *actual*), é também hierarquicamente definido. Assim, o primeiro adjectivo terá que ocorrer sempre pós-nominalmente (exemplo [183a]), já que o seu deslocamento para antes do nome origina uma frase agramatical (exemplo [183b]). Como se vê ainda no exemplo [183a], caso os dois adjectivos ocorram depois do nome, o ordenamento hierárquico impõe que o adjectivo modificador interno ocupe a primeira posição, visto deter também um traço mais informativo. Pelo facto de o adjectivo que não funciona como modificador interno poder ser deslocado para a posição pré-nominal, a construção frásica do exemplo [183c] é gramatical:

- [183] PTG:
- a. *A situação **política** actual do país é estável.*
  - b. *\*A **política** situação actual do país é estável.*
  - c. *A **actual** situação política do país é estável.*

As questões que se prendem com o posicionamento do adjectivo no português poderão ter implicações na aquisição dos mesmos por parte de falantes das línguas do grupo níger-congo atlântico ou do santomense, já que, nestes idiomas, a posição do adjectivo detém rigidez formal, ocorrendo sempre em posição pós-nominal (ponto 1.3.2.2 do presente trabalho).

Após termos apontado algumas especificidades do SN do português, numa abordagem descritivista, importa também observar quer o funcionamento do fenómeno da concordância de número PL na referida língua quer o modo como se dá a aquisição das suas categorias funcionais. Este aspecto impõe-se, dado que a correlação entre a variação e a caracterização das estruturas sintagmáticas, que seguidamente iremos

abordar, ajudará a formular hipóteses (capítulo 3), relativamente às questões que estarão subjacentes à aquisição dessas mesmas estruturas.

### 2.6.2. O quadro teórico da concordância

A concordância de número detém regras bastante rígidas no português padrão,<sup>75</sup> podendo o seu sistema de flexão verbal, por exemplo, fazer com que o sujeito pronominal não seja realizado (Cunha & Sintra (1997 [1984]:494), identificando-se apenas pelos traços de pessoa e número da concordância (Raposo, 1992:43).<sup>76</sup> Segundo a visão generativista, mais concretamente de acordo com os pressupostos do Princípio da Projecção Alargado da Teoria X-barras (Grimshaw, 2005 [1991]), existe uma categoria SN na posição de sujeito das frases, pelo que, em português, este SN pode ter um nome como núcleo ou um pronome foneticamente realizado ou não, e que constitui uma categoria vazia sujeito ou *pro* (Campos & Xavier, 1991:20). Portanto, a concordância é um fenómeno de *interface* por excelência, uma vez que traduz uma manifestação morfológica que é consequência de uma relação sintáctica (Costa & Silva, 2006a:1) Este aspecto determina também que as manifestações de concordância entre diferentes elementos sintácticos sejam distintas nas diferentes línguas (Béjar, 2003; Sigurðsson, 2004), já que algumas destas não aplicam uma das propriedades mais salientes deste fenómeno: a do traço da redundância, que repete informação já expressa em determinadas categorias (Kiparsky, 1972:195). É o que acontece, por exemplo, no santomense, em que a marcação PL a nível do SN pode ocorrer apenas no elemento pré-nominal, mantendo-se inalterado o radical de todos os restantes elementos da cadeia sintagmática. Partindo do pressuposto que a ASL se dá a partir da gramática da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), as produções dos falantes do PA vão apresentar competição entre as gramáticas das línguas em contacto (santomense e PE). Desta forma, a CPL-var no SN do PA é bastante semelhante à de outras línguas que passaram por transmissão irregular, como o PT ou o PVB, nas quais todo o tipo de flexões continua a ser consideravelmente reduzido. Face a estes aspectos, torna-se legítimo questionar sobre o quanto da concordância pode ser estabelecido na gramática, ela mesma, e o quanto da concordância deve ser visto como

*interface*, isto é, como mero reflexo morfológico de uma relação sintáctica, para o qual a própria sintaxe não tem explicações a oferecer.

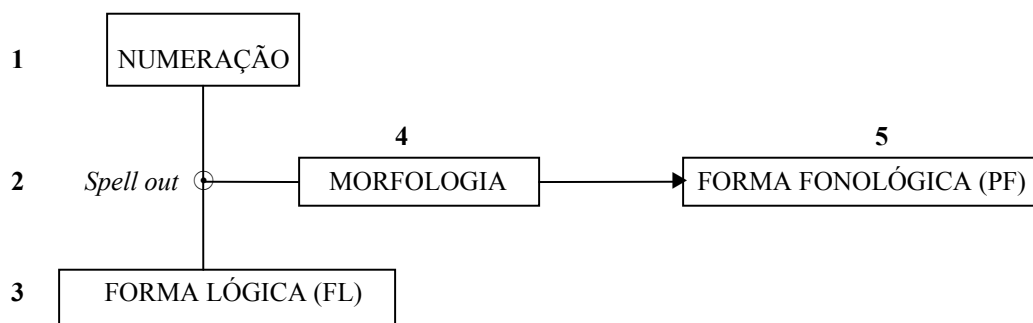
### **2.6.2.1. Concordância e núcleos funcionais**

Segundo a corrente tradicional lexicalista, as relações de concordância só se estabelecem depois de se juntarem todos os elementos que compõem a frase. Nesta conformidade, a formação de palavras ocorre no léxico, já que as regras morfológicas são de natureza distinta das sintáticas, actuando em elementos primitivos diferentes dos das segundas. Portanto, a morfologia opera sobre o radical e os morfemas para produzir palavras, enquanto a sintaxe incide sobre as palavras já flexionadas para introduzi-las nas frases. Esta visão configura o Princípio da Integridade Lexical (PIL), preconizado por Harris (1982:26), que assegura que a sintaxe opera a nível das partes da palavra, impedindo que, por exemplo, os afixos sejam desagregados desta com recurso a regras sintáticas.

Contudo, para os não-lexicalistas, parte da formação das palavras não é feita na morfologia, mas sim na própria sintaxe. Baker (1985), observando o sistema de concordância no chamorro, língua malaio-polinésia das ilhas Marianas, avançou com o Princípio do Espelho (PEH), que estabelece um estreito paralelismo entre morfologia e sintaxe, mas não especifica os princípios gerais segundo os quais os afixos são concatenados na estrutura sintáctica do chamorro. Posteriormente, Baker (1988) estudaria também o falar dos índios hopi, Arizona, língua que não possui determinantes e na qual o verbo e o seu complemento aparecem incorporados um no outro, formando uma única palavra.<sup>77</sup> O facto de os determinantes não serem incorporados, permite perceber que este tipo de operação obedece a uma condição sintáctica, e não morfológica. Deste modo, a sintaxe opera a nível de palavras e morfemas, verificando-se que uma palavra complexa pode ser formada através de regras sintáticas e, mais especificamente, com movimento nuclear, através da incorporação de um radical lexical num morfema. Por outro lado, o processo deixa também entender quais são as restrições às incorporações, isto é, qual é o tipo de nomes que pode ser incorporado. Como os movimentos sintáticos operam de baixo para cima, não há hipótese de ocorrer incorporação dos nomes sujeito.

Este tipo de abordagem permite observar o comportamento de elementos morfológicos e sintáticos, em especial a questão do PEH (Baker, 1985), que estabelece que a derivação morfológica reflecte a derivação sintáctica, e vice-versa. Se a estrutura morfológica de uma palavra complexa resulta de um movimento nuclear, o PEH deixa perceber que “*a ordem dos morfemas numa palavra complexa reflecte a incorporação sintáctica de núcleos que correspondem a esses morfemas*”<sup>78</sup> (Baker, 2002:326). Note-se que a abordagem abrange também os efeitos do PIL, uma vez que o resultado do processo de incorporação é ainda uma categoria a nível da palavra. Este aspecto demarca o modelo da incorporação das outras abordagens sintáticas que interagem com a morfologia, nas quais é assumido que uma palavra complexa é formada de palavras lexicais distintas na sintaxe, e que estas só formam palavras a nível fonológico. Neste caso, os efeitos do PIL terão que receber explicações independentes.

Por seu lado, Halle & Marantz (1993) avançam com a proposta da MD. Essencialmente não-lexicalista, o modelo é consensual, visto não retirar protagonismo à componente lexical (Fig. 16).



**Fig. 16.** *Spell-out* em morfologia distribuída (adaptado de Cook, 1996:320-326).

De acordo com esta proposta, há um *input* para a componente morfológica, que lê o que a sintaxe produziu. Por conseguinte, os processos que não são resolvidos pela sintaxe, são-no pela morfologia, em virtude de esta ser sensível às questões da adjacência. Para Bobaljik (1995:243), o modelo demonstra que a afixação não é feita sintacticamente e, como a operação sintáctica está a ser sensível a questões morfológicas, então o processo também não é morfológico. Nesta perspectiva, a sintaxe não trabalha com palavras, mas sim com formas abstractas (traços). Estas questões levam ainda a concluir que a morfologia actua como filtro da sintaxe, já que cada

morfema corresponde a um encaixe (*slot*) sintáctico, mas cuja afixação não está dependente dos movimentos.

As propostas da MD propõem, duas linhas orientadoras-base (Embick, 2000:192):

- (i) As palavras e subpartes que as compõem são constituintes sintácticos mais pequenos, mas que não diferem, no essencial, dos constituintes maiores, como as orações ou frases. Assim, embora cada constituinte tenha propriedades *sui generis*, todas as suas estruturas internas são elaboradas por uma única componente gramatical: a sintaxe.
- (ii) O Léxico não existe na MD, estando as funções que determinadas correntes linguísticas lhe atribuem distribuídas entre outras componentes da gramática.

Segundo Noyer (2006, *apud* Silva, 2010:19/20 ), a estrutura gramatical em MD pressupõe uma divisão em duas partes (Fig. 17): (i) a *informação listada* (uma lista morfémica, um vocabulário e uma enciclopédia), que é depositada em alguns repositórios; (ii) um *dispositivo generativista*, subdividido em operação sintáctica, propriamente dita, e diversos mecanismos pós-sintácticos (abaixamentos, ornamentações, empobrecimentos, deslocamentos locais e linearização).

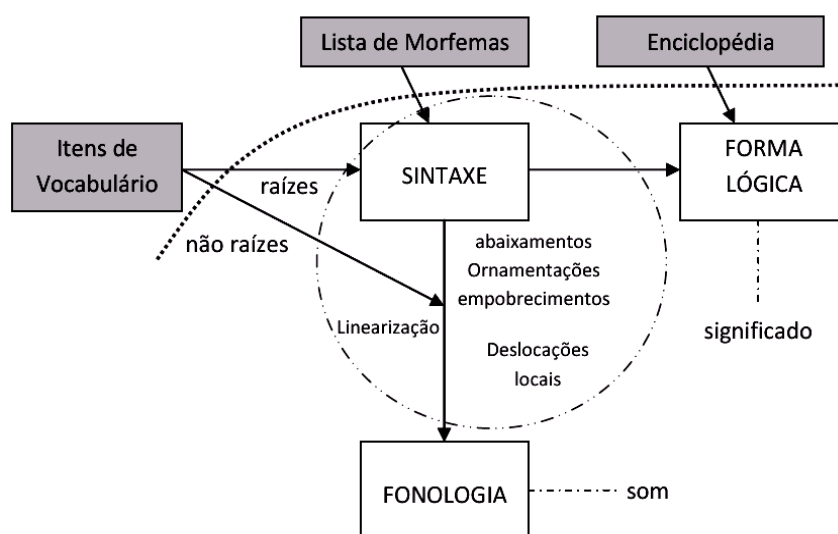


Fig. 17. Modelo da morfologia distribuída (Noyer, 2006:734, *apud* Silva, 2010:20).

O *dispositivo generativista* elabora as expressões linguísticas que provêm dos morfemas seleccionados pela lista morfémica, interpretando fonológica e semanticamente estas expressões com informação concedida pelo vocabulário e pela enciclopédia. Por seu lado, os morfemas contidos na lista morfémica, que constituem os terminais da estrutura morfossintáctica e não contêm traços fonológicos, são de dois tipos: radicais (morfemas-L) e categorias funcionais de sintaxe (morfemas-F). Os primeiros representam itens de classe aberta de categorias indeterminadas, cujas funções são determinadas pelos seus contextos sintácticos (Noyer, 2006:734). Quanto aos segundos, dizem respeito a categorias como o tempo, categorias vazias (CV), CONC ou DET. Portanto, os itens vocabulares (radical das palavras, morfemas de PL ou morfemas de tempo) estabelecem a relação entre representantes fonológicos e morfemas, inserindo-os de acordo as exigências contextuais. Por fim, as entradas da enciclopédia relacionam interpretações com expressões linguísticas estruturadas, que podem ser palavras ou frases.

Halle & Marantz (1993, *apud* Silva, 2010: 21/22) estabelecem, então, três propriedades em MD:

- (i) A Inserção Tardia (*Late Insertion*), que especifica que todas as expressões fonológicas devem ser inseridas somente depois da sintaxe. Deste modo, as categorias sintácticas são abstractas e não possuem conteúdo fonológico, daí resultando que, antes da inserção, elas detenham apenas traços morfossintácticos. Quanto ao momento da inserção do vocabulário, não tem gerado consenso entre os teorizadores. Embick (2000:193) entende que os representantes de morfemas radicais são inseridos anteriormente ou, talvez, durante a sintaxe propriamente dita, enquanto os outros elementos morfemas são-no depois desta. Por um lado, os pressupostos da MD partilham afinidades com os de outras teorias, como a “*Extended Word*” ou o “*Paradigm*” (Anderson, 1982), enquanto, por outro lado, estabelecem semelhanças com os conceitos tradicionais da Morfologia de Base Lexema-Morfema (Beard, 1995);
- (ii) A Subespecificação (*Underspecification*) de itens vocabulares, que determina que as expressões fonológicas não precisam de ser totalmente especificadas para as

posições sintáticas nas quais irão ser inseridas. Este conceito, bastante controverso, rejeita os pressupostos do recurso a um mecanismo de filtragem (Lumsden, 1987; Wunderlich, 1996; Halle, 2000);

- (iii) A Estrutura Sintáctica Hierarquizada *Top-down* (*Hierarchical Structure All the Way Down*), isto é, a elaboração sintáctica dos elementos da sintaxe e da morfologia que se combinam e entram no mesmo tipo de estrutura de constituintes (árvores binárias). A estrutura hierarquizada manifesta-se em todos os níveis da análise estrutural, não ocorrendo morfemas e mantendo-se uma separação restrita entre as regras morfofonológicas e as que atendem aos alomorfes, que podem alterar as propriedades fonológicas dos elementos. Como tal, a inserção do vocabulário limita-se a fornecer elementos. É o que sucede, por exemplo, com a derivação, em inglês, que afecta o termo *foot*, que pressupõe a inserção do item [fut] numa posição de radical. A especificação do contexto de morfema PL implica, desta forma, que se insira um elemento zero na posição canónica de morfema PL. Posteriormente, acontece o reajustamento da forma derivante [i], sendo esta derivação gerada num contexto especial de PL (Noyer, 2006:735).

As propriedades referidas permitem estabelecer o modelo da Fig. 16, em que a sintaxe (1) combina os traços morfossintáticos, cabendo ao *spell-out* (2) a função de conectar a componente morfológica (4), na qual acontecem as operações morfológicas ou de reajustamento à PF (5), ocorrendo as operações fonológicas. Deste modo, encontra-se descrita a teoria que representa, de forma global, o modo como a gramática funciona, no sentido de levar um determinado conceito mental a entrar num *interface* pós-sintáctico conceptual, do qual se processará uma instanciação resultante da construção final na PF. Por seu lado, a FL (3) é apenas um nível de representação que exhibe as relações estruturais relacionadas com o significado das expressões, pelo que não expressa nem representa qualquer significado.

Como se verá no ponto 2.9 do presente trabalho, os pressupostos da MD dão contributo valioso para o entendimento acerca da CPL-var originada pela assimetria de determinados itens gramaticais em posições pré e pós-nucleares, tanto no PB (Costa & Silva, 2006b) como no PA. De facto, tais pressupostos consideram que a estrutura



interna das palavras acontece a partir quer das operações sintáticas, especialmente em sequência do movimento de núcleo (ou Mov-X°), quer de outros mecanismos pós-sintáticos específicos. Portanto, o *merge*<sup>79</sup> morfológico permitirá que os morfemas operem relações de adjacência linear a outros constituintes, daí resultando uma relação de adjunção relativamente aos seus núcleos. Paralelamente, operações de empobrecimento (*impoverishment*) são também de considerar nesta fase, uma vez que as mesmas, ao apagarem os valores de determinados traços morfossintáticos antes da inserção do vocabulário, levam à neutralização sistemática de determinadas categorias (Noyer, 2006:736).

Os estudos da concordância quer entre sujeito e verbo quer a nível dos elementos do SN levaram também à necessidade de se observar a dependência entre concordância (nomeadamente a concordância ao nível do SN) e outros fenómenos sintáticos. Em algumas línguas, como o árabe, a concordância pode ser reflexo das configurações, isto é, resultar de correlações entre as marcas sintáticas e as configurações. De facto, este idioma permite que se altere a ordem dos constituintes, criando-se frases do tipo SV ou VS. Contudo, a configuração entre os constituintes determina que, na estrutura SV, o verbo concorde em número com o sujeito pré-verbal, enquanto na configuração VS essa marca se fixa unicamente no sujeito pós-verbal, permanecendo o verbo no singular (Brustad, 2000:363). Ora, se a concordância é manifestação de algo sintático, mas surge codificada na parte funcional, então esta reflecte aquilo que se passa na sintaxe. Daí existirem línguas que permitem que o verbo concorde com o objecto, como acontece nas estruturas clitizadas do francês (exemplos [184a] e [184b]), e que confirmam a ocorrência de dependência entre as marcas de concordância e a configuração específica:

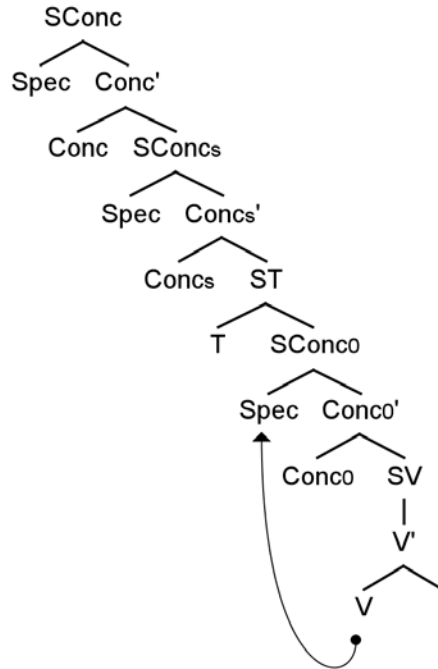
- [184] FRANCÊS:
- a. *J'ai ouvert **les fenêtr**es.*  
“Eu abri as janelas”
  - b. *Je **les ai ouvert**es.*  
“Eu abri-as”.

No árabe é ainda possível constatar que a concordância pode ter maior ou menor visibilidade, dependendo da configuração estrutural da realização. Deste modo, a falta de visibilidade morfológica não significa, necessariamente, ausência de processo sintático.

Tendo em conta o pressuposto de Pollock (1989:386), que propõe que o verbo é formado na sintaxe,<sup>80</sup> muitos linguistas trabalharam a questão da concordância partindo do princípio que esta é uma manifestação directa da categoria sintáctica CONC. Porém, levando em consideração que em algumas línguas há uma concordância visível com o objecto, e assumindo que o Caso é uniformemente atribuído na configuração nuclear SPEC, Chomsky (1993), seguindo Kayne (1989), propõe a distinção entre CONC-S(ujeito) e CONC-O(bjecto). Assim, estes dois núcleos são responsáveis pelo licenciamento do Caso, sendo a concordância estabelecida quando um SX e um núcleo inserem uma relação nuclear de SPEC no domínio dessas categorias.<sup>81</sup>

Este tipo de visão sobre a concordância levantou um conjunto de problemas, sendo alguns destes abordados pelo próprio Chomsky (1996 [1995]), quando avançou com os pressupostos que estão na base do Programa Minimalista. Como o espírito do programa é limpar tudo o que se revelava redundante no Modelo de Princípios e Parâmetros, o autor propõe que se reduzam as operações de “movimento” dos constituintes (*move*) e de concatenação da estrutura (*merge*) ao mínimo indispensável (*Derivation by Face*), isto é, que se considere apenas o que é interpretável e que interessa para as *interfaces*, descartando-se aquilo que não é interpretável para estas. Observada sob a perspectiva minimalista, a sintaxe é, então, a aplicação conjunta das duas operações, *merge* e *move*, com a primeira a combinar dois elementos sintácticos para produzir um outro elemento sintáctico, e a segunda a deslocar visivelmente (*overtly*)<sup>82</sup> um item sintáctico de uma posição para outra (Fig. 18).

A proposta de alteração do Modelo-T não só permite ao objecto deslocar-se para receber o caso como também deixa o sujeito ocupar mais do que uma posição, evidenciado uma estrutura dissociada entre este e o caso nominativo. O modelo prevê que, algures na aplicação das regras entre o léxico e a FL, há um ponto, o *spell out*, em que a derivação sintáctica é enviada para a PF. Uma das consequências inerente a este pressuposto é a de que, independentemente da língua em questão, os movimentos são visíveis, apenas variando o momento em que ocorre o *spell out*. Como tal, a interpretação das produções linguísticas, a FL, é igual para todas as línguas.



**Fig. 18.** Estrutura minimalista da concordância sintagmática (Chomsky, 1996 [1995]:173): deslocação do objecto para receber caso.

Por seu lado, a noção de economia e derivações avançada pelo Programa Minimalista determina que uma operação sintáctica só acontece no caso de ser motivada. Numa língua natural existem dez categorias de mecanismos gramaticais minimalistas<sup>83</sup> que são automaticamente postas em acção. Se estes traços estiverem ausentes ou forem insuficientes no *input*, as crianças accionam os princípios cognitivos inatos para os desenvolverem na sua linguagem. Com vista a definir-se o que é ou não é um movimento motivado, foi, então, criado um sistema de *traços-phi*, associado a qualquer categoria sintáctica. Estes podem ter uma força forte (o elemento sintáctico tem movimento) ou fraca (o elemento sintáctico não tem movimento), indo as forças determinar o factor de variação entre as línguas. Na estrutura minimalista da concordância sintagmática (Fig. 18), CONC-S e CONC-O são mnemónicas informais para distinguir os dois *papéis-phi* de CONC. Assim, este representa um conjunto de *papéis-phi* (género, número, pessoa), ou seja, os sistemas comuns de CONC para sujeitos e objectos, uma vez que CONC-S e CONC-O podem representar selecções distintas de níveis de acordo. Por outro lado, os *traços-phi* associados às categorias linguísticas são ainda de dois tipos: interpretáveis, que têm relevância na FL (p.e. o número); e não-interpretáveis (p.e. o género e o caso).

Na visão de que as categorias vêm equipadas com traços, a sintaxe serve basicamente para eliminar, através das operações de movimento, aqueles que não são interpretáveis, ocorrendo o apagamento quando traços semelhantes se encontram. Como tal, o falante selecciona como gramatical aquilo que é mais económico, esvaziando a componente sintáctica e colocando-a no léxico. Nesta perspectiva, o movimento existe para seleccionar as operações que não interessam para as *interfaces*, pelo que constitui uma propriedade mais económica (preferida) da língua do que a inserção de material lexical. Assim, as *interfaces* ganham lógica e tudo o que não cabe no modelo mínimo cai na sua área. Este aspecto determina que alguns movimentos, como o movimento- $\bar{A}$ , que tem por base inúmeras motivações (p.e. questões de escopo ou discursivas), afectem qualquer tipo de sintagma, enquanto outros, como o movimento-A, motivado pelo caso e cujo alvo é a posição SPEC, afectem apenas os SN's argumentais.

A partir dos pressupostos minimalistas, foi possível procurar entender a estrutura das *interfaces*. Deste modo, não só a morfologia, que não é contemplada no Modelo-T, pôde ser tratada de forma mais clara, como também se abriu caminho para que, posteriormente, se observasse a *interface* entre sintaxe, morfologia e semântica.

### **2.6.3. O SN no português europeu**

Qualquer construção linguística configura um sintagma, ou grupo, cujo elemento central, o núcleo, é realizado com recurso a uma categoria sintáctica (nuclear), vulgarmente designada por “parte do discurso”, “classe de palavras” ou “classe morfossintáctica” (Brito, 2003a:326). No caso do português, a estrutura interna do SN inclui obrigatoriamente um núcleo e, opcionalmente, outros dois tipos de constituintes: complementos e especificadores. Se as palavras inseridas no SN configuram unidades pertencentes ao reduzido grupo de palavras ou unidades morfológicas da língua, cujo significado remete para as noções mais abstractas desta, estamos na presença de categorias funcionais. Opostamente, se as unidades incluídas no SN fazem parte do extenso e renovável inventário do vocabulário ou léxico da língua, cujos significados remetem para entidades, situações, propriedades ou relações entre entidades, estamos perante categorias lexicais, como acontece com os nomes, os adjectivos, as preposições, os verbos e os advérbios. Estas categorias, por seu lado, constituem o núcleo das respectivas categorias sintagmáticas em que estão incluídas: sintagma nominal (SN),

sintagma adjetival (SADJ), sintagma preposicional (SPREP), sintagma verbal (SV) e sintagma adverbial (SADV). Deste modo, as categorias sintagmáticas são projecções endocêntricas dos seus próprios núcleos, já que os segundos são da mesma natureza das primeiras.

Face ao exposto, e observando a estrutura geral do SN, verifica-se que ele representa uma categoria sintagmática que é projecção de um nome, podendo ser composto apenas por esse nome (exemplo [185a])<sup>84</sup> ou por um pronome (exemplo [185b]),<sup>85</sup> os quais são, simultaneamente, núcleo do referido SN. Contudo, este pode ainda ser realizado por um nome e por outros elementos opcionais, tais como determinantes (exemplo [185c]), quantificadores (exemplo [185d]), expressões quantitativas (exemplo [185e]) e modificadores, dividindo-se estes em SADJ's (exemplo [185f]), SPREP's (exemplo [185g]), sintagmas oracionais (exemplo [185h]) ou epítetos (exemplo [185i]) (Brito, 2003a:328-329):

[185] PE:

- a. <sup>SN</sup>[**Deus**] *perdoa os que se arrependem.*
- b. <sup>SN</sup>[**Ninguém**] *quis sobremesa.*
- c. <sup>SN</sup>[**Esses livros**] *estão à venda.*
- d. <sup>SN</sup>[**Muitos livros**] *ficaram por vender.*
- e. <sup>SN</sup>[**Uma porção de livros**] *ficou por vender.*
- f. *O Pedro é* <sup>SN</sup>[*um homem grande*].
- g. *O Pedro tem* <sup>SN</sup>[*muitas camisas de algodão*].
- h. *Perdi* <sup>SN</sup>[*o isqueiro que comprei ontem*].
- i. <sup>SN</sup>[*O João, aquele malandro*], *tem faltado às aulas.*

Os exemplos evidenciam que núcleo e complementos constituem a estrutura lexical da categoria SN, enquanto determinantes e quantificadores configuram a sua estrutura funcional. Por seu lado, os modificadores recobrem constituintes com variados valores semânticos, surgindo alguns mais ligados à estrutura lexical, enquanto outros se conectam mais à estrutura funcional. Em algumas realizações, o núcleo do SN é também aparentemente preenchido por especificadores (exemplo [186a]), mas, na verdade, está-se perante SN's em que o núcleo é uma categoria vazia, uma vez que tais especificadores se podem conjugar com outros especificadores e com diversos tipos de complementos (exemplo [186b]):

[186] PE:

- a. *Todos gostaram do espectáculo.*
- b. *Todos aqueles não gostaram do espectáculo.*

Sendo a forma do SN marcada por mecanismos de concordância, os exemplos patentes em [185] e [186] evidenciam também que o núcleo do SN determina a concordância dos especificadores, dos determinantes (artigos, demonstrativos e possessivos) e de alguns complementos (SADJ's e epítetos). Apesar de haver critérios que permitem a identificação dos nomes, deve ter-se em consideração que a concordância interna do SN é morfológica e semanticamente susceptível ao tipo de nome que pode servir de núcleo (nomes animados humanos, nomes animados não humanos, nomes não animados, nomes comuns, nomes próprios, nomes contáveis, nomes não contáveis ou massivos, nomes concretos e nomes abstractos). Estes aspectos determinam, então, as condições em que se estabelecem as relações de complementação a nível das estruturas lexical e funcional do SN. De facto, “*os nomes relacionáveis lexicalmente com outras categorias são quase sempre predicados semânticos, com lugares ou argumentos que recebem uma dada interpretação, que pode ser descrita através de relações temáticas*” (Brito, 2003a:331).

#### **2.6.4. Categorias funcionais e categorias lexicais**

A teoria linguística tem procurado estabelecer paralelismos entre a sintaxe das frases e a sintaxe dos SN's, já que este tipo de observações pode ajudar a entender quer aquilo que motiva os diferentes padrões de ordem das palavras nos SN's das diferentes línguas quer a relação estabelecida entre a sintaxe e a morfologia flexional dentro do SN (Montrul, 2004:12, apud Godinho, 2005:131). Tendo em conta que núcleo e complementos constituem a estrutura lexical da categoria SN, então os determinantes e as formas de quantificação (quantificadores, expressões quantitativas e expressões partitivas) formam a sua estrutura funcional, localizada à esquerda do nome. Assim sendo, na perspectiva da gramática generativa actual, o SN tradicional deixou de ser entendido como tal, para, de acordo com o princípio da endocentricidade, passar a ser visto como um SDET, a projecção máxima de um DET (Fig. 19). Esta hipótese postula, então, que os determinantes e formas de quantificação, que restringem a extensão do núcleo nominal, devem ser considerados o núcleo da categoria funcional SDET ou

sintagma quantificador (SQ), os quais são responsáveis pela selecção dos SN's como complementos. Desta forma, N é o núcleo da parte lexical ou categoria (Brito, 2003a:345-349, Fig. 19), cabendo ao SDET a função de atribuir valores referenciais do tipo [+definido] ou [-definido] ao SN complemento. Portanto, SDET e sintagma complementador (SCOMPL) estabelecem uma relação de paralelismo no SV.

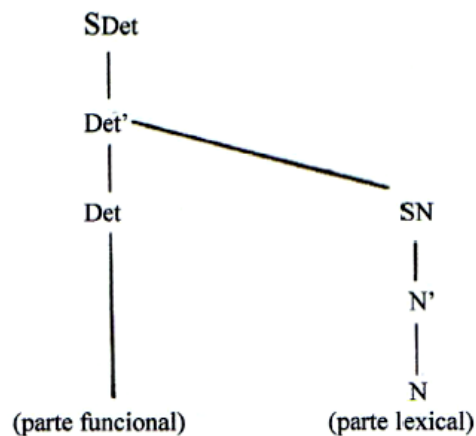


Fig. 19. Estrutura do SDET: parte funcional e parte lexical (Brito, 2003a:345, *apud* Godinho, 2005:132).

Na estrutura proposta, o núcleo da categoria SDET é preenchido por artigos ou demonstrativos. Assim, os determinantes (artigos e, por norma, possessivos, demonstrativos e quantificadores) precedem o nome núcleo do SN, enquanto os adjectivos, complementos e adjuntos aparecem, geralmente, depois deste (cf. Godinho, 2005:132). Ocorre, desta forma, uma relação entre posição linear e categoria gramatical, que influencia a marcação PL (Guy, 1981b:179), aspecto com o qual Scherre (1988, 1998a [1996]) entra em desacordo. Esta questão, à qual fizemos já referência no ponto 1.2 do presente trabalho, será por nós amplamente debatida no capítulo 5, especialmente no ponto 4.1.2.6, dedicado à observação do comportamento da variável *posição do item na cadeia do sintagma nominal* e seu cruzamento com outras variáveis independentes.

No que concerne aos artigos, precedem sempre os nomes e são de dois tipos: definidos e indefinidos. A distinção entre ambos é feita com recurso aos respectivos valores semânticos e, ao contrário dos segundos, os definidos nunca podem surgir isolados, quer por elipse do N (p.e. \**Eu comi o.* / *Eu comi um.*) quer devido ao movimento do N e do complemento (p.e. \**Manual de botânica, eu comprei o.* / *Manual de botânica, eu comprei um.*). Ainda relativamente aos definidos, podem surgir em

realizações com elipse nominal, caso sejam seguidos de um complemento (p.e. *Eu comprei os chinelos azuis e tu os amarelos.*), um modificador adjetival (p.e. *O Pedro comprou os verdes.*) ou uma oração relativa (p.e. *O João comprou os que são castanhos.*). Estes artigos surgem também em distribuição complementar relativamente aos demonstrativos, tornando agramaticais as frases que não respeitam este preceito (p.e. *\*Eu comprei o este jornal de hoje.*). Por sua vez, os indefinidos não acompanham nomes próprios (p.e. *\*Ontem, eu falei com um António.*) nem podem ser colocados depois de *todos* ou *ambos* (p.e. *\*Todos uns alunos foram à visita de estudo./ \*Ambos uns alunos faltaram às aulas hoje.*).

Os demonstrativos, ainda que não co-ocorram com definidos, podem realizar-se em construções de elipse nominal (p.e. *Prefiro estes aqui.*). Nas realizações de tipo déitico, os demonstrativos *este*, *esse* e *aquele* conjugam-se com localizadores para funcionarem como “filtros de vizinhança” (Lopes, 1972:141), mas apenas em distribuição complementar (p.e. *\*Este aqui livro é do Pedro.*). Os demonstrativos podem ainda possuir valor anafórico, remetendo para o discurso anterior (p.e. *Eu comprei livros de matemática e botânica. Aqueles foram mais caros do que estes.*) e ser realizados pré (exemplo [185c]) ou pós-nominalmente com exclamativas (exemplo [187a]) ou relativas apositivas (exemplo [187b]), mas sempre em distribuição complementar com os artigos, sob pena de a construção se tornar agramatical (exemplo [187c]):

[187] PE:

- a. *Que calor este!*
- b. *Comprei um carro ontem, carro esse que está na minha garagem.*
- c. *\*Os esses livros estão à venda.*

Ainda assim, a distribuição frásica dos demonstrativos justifica a aplicação da estrutura frásica patente na Fig. 19 (Brito, 2003a:349).

Relativamente aos possessivos, exprimem valores temáticos e de determinação, possuindo marcas de género e número denunciadoras de uma dupla natureza lexical-funcional. Portanto, podem co-ocorrer na cadeia pré-nominal com artigos definidos (exemplo [188a]), demonstrativos (exemplo [188b]) e quantificadores (exemplo [188c]). Os possessivos podem pospor-se ao nome, se este não for antecedido de artigo (exemplo [188d]) ou se realizar com artigo indefinido (exemplo [188e]). A localização pós-



nominal pode ocorrer com cardinais (exemplo [188f]) ou estruturas não argumentais, envolvendo vocativos (exemplo [188g]) ou posições predicativas (exemplo [188h]):

[188] PE:

- a. ***O meu irmão*** chega amanhã a Macau.
- b. ***Aquela tua amiga*** é muito simpática.
- c. ***Estes meus muito amigos*** partem hoje para Lisboa.
- d. Estava à espera de ***notícias tuas*** há mais de uma semana.
- e. Recebi ***uma carta tua***, que me deixou muito contente.
- f. Hoje vi dois ***empregados seus***.
- g. ***Virgem nossa***, que estupidez!
- h. ***Aquele ali*** é um ***empregado teu***?

Nos exemplos até agora exibidos, temos focalizado atenção nas estruturas em que o SDET é constituído por uma sequência de palavras em torno de um nome. Apesar de as sequências referencialmente autónomas (p.e. os pronomes pessoais ou os demonstrativos *isto, isso, aquilo*) poderem constituir, por si sós, SDET's (p.e. *Vocês fizeram serão?* / *Isto tem muito que se lhe diga.*), abster-nos-emos de as observar e comentar, já que caem fora do âmbito do presente estudo. Pelos mesmos motivos, não observaremos as elipses nominais com determinantes nem as realizações com nomes próprios.

Relativamente à quantificação dos nomes, pode ser expressa com recurso aos plurais simples (p.e. *Hoje comprei chocolates para ti.*), aos quantificadores propriamente ditos (p.e. *Hoje comprei muitos livros na Feira dos Livros.*) ou às expressões quantitativas (p.e. *Hoje comprei uma data de livros na Feira dos Livros.*). Sobre os tipos de elementos abrangidos pelos quantificadores e expressões quantitativas, bem como as suas projecções endocêntricas em estruturas sintácticas, remetemos as explicações pertinentes para Brito (2003a:356-365), que abrangem também as expressões partitivas (p.e. *Alguns dos livros que comprei, são raríssimos.*).

No que aos modificadores nominais (restritivos ou apositivos) diz respeito, centraremos a nossa atenção nos adjectivos. Estes podem ocorrer antes (exemplo [189a]) ou depois do nome (exemplo [185f]), mas os posicionamentos pré-nominais contemplam situações bastante restritas. Contudo, como se observou no ponto 2.6.1 do presente trabalho, há adjectivos que possuem um traço formal rígido, pelo que só podem ser realizados antes (exemplos [189b] e [189c]) ou depois do nome, (exemplos [189d] e [189e]), consoante o caso. Os que podem ocorrer em posições indistintas, mudam

normalmente de valor semântico, conquanto se encontrem antes ou depois do nome. Nestes casos, a posição pós-nominal está associada a significados mais objectivos ou denotativos (exemplo [185f]), enquanto a localização pré-nominal surge mais ligada a interpretações subjectivas ou conotativas (exemplo [189a]). Não obstante, há adjectivos cuja alteração de posição relativamente ao nome não acarreta mudança de valor semântico (exemplos [176a] e [176b]), estando o posicionamento dependente de uma mera questão estilística ou da intenção de focalizar a informação.

[189] PE:

- a. *O Pedro é um **grande** homem.*
- b. *Aquilo não passou de um **mero** mal-entendido.*
- c. *\*Aquilo não passou de um mal-entendido **mero**.*
- d. *O vestido **azul** da Maria é muito bonito.*
- e. *\*O **azul** vestido da Maria é muito bonito.*

Note-se também que há nomes não relacionáveis lexicalmente com outras categorias e que podem surgir seguidos de SPREP's (os complementos determinativos). Estes, por seu lado, podem restringir o significado do SN (p.e. *O manual de botânica.*), tal como acontece nas ocorrências com relativas restritivas (p.e. *O manual que trata de botânica.*) ou com determinantes (p.e. *Esse manual.*). Outros SPREP's, contudo, podem exprimir relações semânticas mais ténues, a ponto de originarem realizações ambíguas (p.e. *Gostei muito do quadro do Manuel da Glória.*).

#### 2.6.4.1. Aquisição do sintagma determinante

Godinho (2005, 132-138) refere pormenorizadamente o modo como ocorre a aquisição da categoria funcional SDET. Dado que esta domina o SN, a projecção máxima do nome, mas que não contém o determinante, é também uma categoria que descreve uma expressão nominal referencial, aparecendo como categoria funcional ligada à categoria lexical SN. Desta forma, o N pode ocorrer sem DET, originando SNN's, mas que não invalidam a hipótese de existir projecção do SDET (Baxter. Lucchesi & Guimarães, 1997:8, nota 14, *apud* Godinho, 2005:132). Meisel (1994a:92) constatou que, na fase FL0 da aquisição bilingue, não ocorrem categorias funcionais. Estas acabam apenas por emergir faseadamente no tempo, para determinarem o desenvolvimento das categorias gramaticais associadas à flexão (concordância, tempo e finitude). Por

consequente, o SDET, apesar de ser foneticamente vazio, continua a encerrar em si as propriedades referenciais de especificação do SN (Godinho, 2005:132). Note-se que, sem o primeiro, o segundo estaria também impedido de actuar como um argumento do verbo da frase, um aspecto que vai determinar que o núcleo DET detenha a função de subordinador, ou seja, uma função semelhante à desempenhada pelos complementadores, enquanto núcleos das frases encaixadas (Godinho, 2005:133).

Como a flexão é um processamento sintáctico, não pode ocorrer dentro do SN, como o comprova o facto de, empiricamente, acontecer marcação variável de género e número no núcleo do SN (Baxter, Lucchesi & Guimarães, 1997:9, *apud* Godinho, 2005:133, nota 173). De facto, é na categoria NUM, pertencente ao nó funcional SNUM, (categoria intermédia entre o SN e o SDET), que estão comportados os traços de número [+/- plural], a fim de serem mapeados e resgatados pelo SDET. (Picallo, 1991; Carstens, 1991; Bernstein, 1993; Cinque, 1994; Koehn, 1994; Müller, 1994a; Montrul, 2004; Godinho, 2005) (Fig. 21):

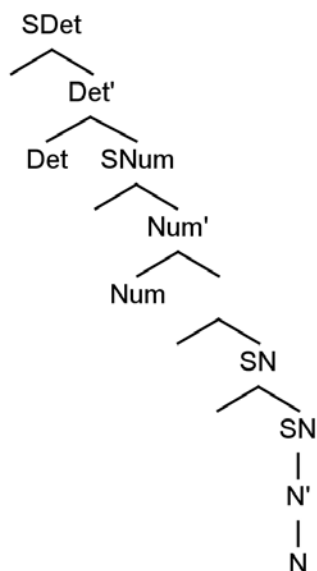


Fig. 20. Estrutura do SDET: nó funcional SNUM.

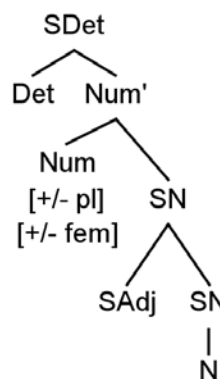


Fig. 21. Representação funcional da estrutura do SN (Montrul, 2004:35, *apud* Godinho, 2005:134).

No português, os traços funcionais de número atraem os nomes, que sobem de N para NUM (cf. Fig. 22), originando a ordem frásica típica do português (N > ADJ) (Bruhn de Garavito & White:2002:154, *apud* Godinho, 2005:134). Ainda a ter em consideração é o pressuposto de que a categoria SNUM contém, igualmente, a propriedade de

concordância em género (Ferrari-Neto, Corrêa & Augusto, 2005:1077), assumindo, assim, as funções de categoria CONC (ou AGREE) (Cerqueira, 1994; Godinho, 2005:134).

As aquisições de L1 e L2 exibem aspectos comuns e são orientadas por operações particulares. No caso concreto da aquisição do SDET na L1, as categorias funcionais apenas detêm função gramatical, subordinando a si as categorias lexicais e circunscrevendo os parâmetros de variação aos seus traços (cf. Godinho, 2005:135). No caso concreto dos falantes de línguas do grupo níger-congo atlântico ou do santomense em aquisição do PtgL2, eles não só terão de adquirir os aspectos morfossintácticos da LA (aprender que os nomes da LA não são marcados com afixos temáticos) mas também perceber que a marcação de género está sujeita a marcação PL e concordância em género e número, normalmente com recurso a sufixos que se dissociam pelo nome, determinantes e adjectivos (cf. Godinho, 2005:136).

A aquisição precoce da categoria funcional SDET no PtgL1 revela variação na concordância em género e número nas suas fases iniciais (Faria, Freitas & Miguel, 1998, 2001; Name & Corrêa, 2003, Costa & Santos, 2004 [2003]; Corrêa, Augusto & Ferrari-Neto, 2005; Ferrari-Neto, Corrêa & Augusto, 2005; Castro & Ferrari-Neto, 2007), uma situação ocorre também na aquisição bilingue (Barreña, 1997; Lléo, 2001; Meisel, 1994a). As etapas de aquisição são discriminadas por Godinho (2005:137), iniciando-se com a produção de nomes, categorias que exibem também os traços de número em primeiro lugar (Faria, Freitas & Miguel, 2001:55, *apud* Godinho, 2005:137). Por volta dos catorze meses, os pequenos falantes produzem também sílabas pré-nominais, entendidas como proto-determinantes ou proto-morfemas (*fillers*) (Faria, Freitas & Miguel, 1998; Name & Corrêa, 2003; López-Ornat, 1997; Lléo, 1997, 1998, 2001), visto assumirem a função de marcadores de lugares morfológicos. A partir dos dezoito meses, as crianças começam a elaborar SN's compostos por um elemento vocálico, do tipo DET, e um N marcado em género. No final desta etapa da aquisição, inicia-se a inserção correcta dos adjectivos na ordem frásica, cuja concordância em género com o N é feita quase sem variação, especialmente quando este é do tipo inanimado. O uso de flexões e concordância revela que se iniciou a fase da aquisição gramatical. Por outro lado, a realização dos diferentes tipos de DET nas fases precoces da aquisição da L1 fornece indicativos importantes não só para a discussão acerca da natureza das gramáticas

precoces mas também para ajudar a determinar as etapas da aquisição em que as categorias funcionais são disponibilizadas (Godinho, 2005:137).

A ausência de SDET's na etapa FL0 da aquisição tem sido apontada por diversos estudiosos (Koehn, 1994; Müller, 1994a; López-Ornat, 1997), pelo que, na transição desta fase para a fase gramatical, “*as crianças fazem corresponder ‘pistas’ fonoprosódicas a funções pragmático-semânticas*” (Godinho, 2005:138). Portanto, a criança não consegue discriminar determinados sons, entendendo formas distintas (DET+N) como sequências fonoprosódicas únicas (Koehn, 1994:49; DeGraff, 1999b:482; Faria Freitas & Miguel, 2001:53). A finalidade pragmático-semântica destas realizações permite que a criança percepcione as regularidades linguísticas, extraia as suas representações morfológicas e produza, depois, DET's que concordam em género e número com o N (cf. Godinho, 2005:138).

Na aquisição do espanhol L1, a importância da fonologia na aquisição morfossintáctica é evidenciada pelas vogais pré-nominais proto-morfémicas, algumas das quais apresentando regularidade linguística (López-Ornat, 1997, *apud* Godinho, 2005:138). Deste modo, a elaboração gramatical segue um trajecto que transita da produção fonoprosódica para a realização sintáctica. Estes aspectos vêm contrariar os pressupostos de uma observação da aquisição da linguagem numa perspectiva meramente sintáctica (Montrul, 2004:42, *apud* Godinho, 2005:138), relevando a importância da fonologia e da prosódia na aquisição.

No caso concreto do PEL1, Faria, Freitas & Miguel (1998, *apud* Godinho, 2005:138-139) apontam a realização de proto-morfemas (*fillers*) na posição do DET, confirmando que as categorias funcionais estão presentes desde as etapas primordiais da aquisição. Para além destes proto-determinantes, Faria, Freitas & Miguel (1998, 2001, *apud* Godinho, 2005:139, notas 185 e 186) mencionam ainda a inserção de morfemas de género e número nos nomes, sob a forma de fricativa morfológica em coda de sílaba, que é realizada anteriormente às outras fricativas em coda de sílaba.<sup>86</sup> Como se verifica, tudo aponta para que a categoria SDET esteja disponibilizada desde a fase FL0 da aquisição, ainda que não haja controlo sobre as formas morfofonológicas dos elementos que a compõem. Nesta conformidade os proto-morfemas atestam que as posições estruturais já se encontram contempladas, embora as operações de mapeamento de

traços só venham a acontecer mais tardiamente (Faria, Freitas & Miguel, 1998:119, *apud* Godinho, 2005:139), por norma a partir dos 2;0 anos, isto é, quando a maturação neurológica da criança lhe disponibiliza a GU (Meisel, 1994a:127), passando a sua gramática a operar a partir desta fase dentro da Hipótese Continuista da aquisição (Pinker, 1984; Hyams, 1987; Kato, 1995; Rice & Wexler, 1996). Estes aspectos são confirmados, nas mesmas pesquisas, com o facto de a gramática das crianças que adquirem o PE se apresentar mais próxima da gramática dos adultos do que a gramática de crianças que adquirem línguas germânicas ou outras línguas românicas (Godinho, 2005:139).

Também em estudos sobre aquisição de artigos no espanhol e no alemão é apontado o aparecimento de proto-morfemas na posição do DET, preenchendo e assumindo, posteriormente, a posição sintáctica destas categorias (Lléo, 1998; Montrul, 2004, *apud* Godinho, 2005: 139-140), o que sugere continuidade.

Os estudos de Name & Corrêa (2003), Ferrari-Neto (2003), Corrêa, Augusto & Ferrari-Neto (2005) e Ferrari-Neto, Corrêa & Augusto (2005), realizados com crianças em processo de aquisição do PB, evidenciam que elas são, igualmente, sensíveis aos determinantes, ainda que estes não apresentem um padrão fónico regular (Godinho, 2005:140). As conclusões dos estudos não confirmam, desta forma, os pressupostos de que o uso de determinantes não é frequente na produção inicial das crianças (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993-1994; Meisel, 1994a, 1997). De facto, as crianças com cerca de catorze meses de idade conseguem já identificar os determinantes como uma classe fechada e mapeá-los com o DET, evidenciando que, nesta fase da aquisição do PB, os requisitos para identificação do género, número e pessoa, se encontram já contemplados. (Name & Corrêa, 2003:25, *apud* Godinho, 2005:140).

Estes pressupostos são compartilhados por Lléo (1997, 1998, 2001), que entende a projecção do SDET numa fase mais precoce como um requisito solicitado pelo contexto, configurando morfemas funcionais relacionados com os artigos e confirmando a hipótese de maturação. Por conseguinte, o acesso a estas categorias não é inato, sendo a aquisição da categoria SDET em PtgL1 governada por aspectos fonológicos e prosódicos. Daí as crianças portuguesas adquirem esta categoria antes das crianças em

aquisição do alemão ou inglês L1, já que estas estão submetidas a um *input* fonológico e prosódico distinto do português (Godinho, 2005: 141-142).

#### **2.6.4.2. Aquisição de género e número**

A ultrageneralização das regras de concordância de género nas fases primordiais da aquisição da L1 evidencia que a aplicação da regra de concordância entre DET e N já se encontra disponibilizada nestas, apesar de não ser aplicada correctamente (Godinho, 2005:142). Ao procurar determinar o momento preciso em que as crianças adquirem o mecanismo sintáctico para a concordância do género no PB, Name & Corrêa (2003) constataram que a variação morfofonológica ocorre na categoria DET. Esta particularidade evidencia que as crianças possuem um dispositivo inato para processar as propriedades formais desta categoria (Godinho, 2005:142), aspecto que seria confirmado por Ferrari-Neto, Corrêa & Augusto (2005:1077) em pesquisa sobre o modo como as crianças com cerca de dois anos de idade percebem a informação morfofonológica relativa ao número e processam a concordância de número no âmbito do SDET no PBL1. No estudo, os autores concluem que tanto o género como o número são processados no âmbito do SDET, uma vez que é *“a informação de género expressa no DET e o processamento da concordância de género expressa no SDET que possibilitam à criança identificar o género de palavras novas (nomes com traço semântico inanimados, nos quais o género é intrínseco, sem motivação semântica, de um ponto de vista sincrónico)”*. Como tal, os autores entendem que o traço de género pode ser mapeado a nível do DET0, uma situação idêntica à do mapeamento de traço de número (Longobardi, 1994:621). Contudo, essa não é a visão de Lucchesi (2000a), para o falar de HEL-Ba, e Figueiredo (2009a), para o PA, que entendem que a inserção da marca de género ocorre no nome núcleo do SN destas variedades de português, expandindo-se a partir deste para se fixar nos elementos à sua esquerda, primeiramente, e só depois nos itens à sua direita, embora de forma marginal. Também Godinho (2008, 2009), em estudos sobre aquisição da concordância de género e número do PtgLE por adultos falantes de cantonense L1, em três momentos distintos, refere que os nomes núcleo do SN são os elementos que exibem maior índice de marcas, sendo responsáveis pela fixação dos traços da flexão morfológica de género no primeiro momento da aquisição. Posteriormente, os traços

disseminam-se pelos outros itens, adjacentes ao núcleo (cf. Lucchesi, 2000a; Figueiredo, 2009a). Segundo a autora, estes aspectos permitem predizer que a concordância se vai espalhando pelos diferentes elementos do SN, fazendo prever que a aquisição da concordância neste é um sistema que sofre maturação gramatical. Desta forma, os princípios da GU actuam a partir dos primeiros momentos da aquisição, a que se seguem, mais tardiamente e de modo descontínuo, outros aspectos operativos (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994; Meisel, 1994a, 1997), sendo o desenvolvimento sintáctico conduzido pela aprendizagem de novos itens lexicais e morfológicos (Montrul, 2004:96). Este aspecto implica que, de certa maneira, as novas estruturas só são aprendidas caso as etapas de aquisição prévia estejam completadas (Pienemann, 1998, 2005b), conforme o confirmam, aliás, os momentos dois e três da aquisição observados por Godinho (2008, 2009). Por outro lado, os achados permitem também concluir que os adultos falantes de cantonense L1 em aquisição do PtgLE cometem erros de concordância em virtude de os traços parametrizados para aquisição do género e número deixarem de estar disponibilizados depois do período crítico do desenvolvimento.

Para entendimento de aspectos concernentes à concordância de número, Godinho (2005, 143-144) aponta a importância dos estudos de Marrero & Aguirre (2003), Pérez-Pereira (1989) e Bruhn de Garavito (1994). No primeiro são abordadas as etapas de desenvolvimento do PL no espanhol, em que as crianças começam por não fazer a distinção entre formas singular/plural, para, posteriormente, aplicarem o PL unicamente na classe DET e, depois, o inserirem nos outros elementos do SDET, ainda que com variação na aplicação da regra de concordância. Para Montrul (2004, *apud* Godinho, 2005:144), esta aquisição tardia dos alomorfes de PL é determinada por factores fonológicos, enquanto a aquisição do género é mais precoce, em virtude de esta categoria ser motivada por factores meramente linguísticos. Por seu lado, Pérez-Pereira (1989, *apud* Godinho, 2005:143) debruçou-se sobre a aquisição dos alomorfes /s/ e /es/ na mesma língua, constatando dois aspectos: (i) ultrageneralização do primeiro alomorfe, sendo também aplicado nas palavras que requerem o uso da vogal epentética, sem que esta seja realizada; (ii) tendência para manter no singular os itens terminados em -s.



Quanto ao estudo de Bruhn de Garavito (1994), analisa a forma como aprendentes de L1 francesa adquirem o PL no espanhol L2. Os estádios de aquisição são idênticos aos apontados por Pérez-Pereira (1989), já que os falantes começam por não realizar a marcação de PL, adquirindo depois o alomorfe de PL /s/ e, apenas posteriormente, o marcador de pluralização /es/. A conclusão é que “*a aquisição tardia do alomorfe /-es/ não é um problema morfológico, mas sim de natureza fonológica e diz respeito à aquisição da epêntese e das regras de silabação do espanhol*” (Godinho, 2005:144).

Ferrari-Neto, Corrêa & Augusto (2005), no seu estudo sobre a forma como as crianças com cerca de dois anos de idade percebem a informação morfofonológica relativa ao número e processam a sua concordância no âmbito do SDET no PB, levaram em conta os princípios minimalistas de que as relações sintácticas entre elementos seleccionados do léxico se realizam por meio do pareamento de traços formais mais e menos interpretáveis (Chomsky, 1996 [1995], 2001b). Portanto, é esse pareamento que permite ao dispositivo computacional universal operar sobre traços do léxico de origem semântica. A existência de traços não-interpretáveis correlatos a traços semanticamente interpretáveis (p.e. o número) permite que o número conceitual encontre uma expressão gramatical, ou seja, tome parte em operações sintácticas (*merge* e AGREE). Neste contexto, os autores entendem que a concordância é o pareamento de traços interpretáveis com traços não-interpretáveis, numa configuração de C-comando. No processo de pareamento, os traços não-interpretáveis são valorados ou eliminados, tendo o resultado dessa valoração uma expressão morfológica, naquilo que tradicionalmente se designa como concordância de número, género e pessoa, em línguas morfológicamente mais elaboradas, como as românicas. No caso específico da concordância de número, o traço interpretável encontra-se no nome (Chomsky, 1996 [1995]) ou numa projecção funcional independente (Ritter, 1991:61).

Em ambas as situações, o número interpretável tende a expressar-se numa desinência flexional do nome, ganhando visibilidade na *interface* fonética. Quanto aos restantes elementos do SDET (determinantes, possessivos, adjectivos, etc.), apresentam traços não-interpretáveis de número, estabelecendo, desta forma, uma relação de concordância com o nome ou com a projecção funcional NUM.<sup>87</sup> Na sequência de tal relação, o SDET assume o traço de número definido no nome (ou no SNUM). Como em algumas línguas

os elementos portadores dos traços não-interpretáveis adquirem um afixo flexional correspondente ao valor do traço de número, aquando da sua codificação morfológica, a informação relativa ao número nesses elementos torna-se visível na *interface* fonética. Deste modo, co-existem dois sistemas no PB: (i) no modelo padronizado, tanto a expressão morfológica do traço interpretável de número na desinência do nome como a expressão morfológica do traço valorado de número no determinante, e demais elementos que entram em concordância com o nome, encontram-se visíveis na *interface* fonética (p.e. *A-s casa-s branca-s*); (ii) no modelo não-padrão, a informação de número tende a ser visível na *interface* fonética apenas no DET (ou no elemento mais à esquerda no SDET), originando concordância variável (p.e. *A-s casa branca*). Por conseguinte, quando a expressão do número é feita apenas no DET, a identificação dessa informação, por parte da criança, durante os enunciados da língua em aquisição, envolve o processamento da concordância no SDET.

Ferrari-Neto, Corrêa & Augusto (2005:1081) chamam ainda a atenção para o facto de o fonema /s/ interferir nos estímulos testados. Todavia, o estudo não clarifica se este efeito pode ser tomado como indicativo de uma fase de desenvolvimento linguístico na qual a criança sofre interferência desse tipo de condicionamento silábico, e que a leva a processar singulares quando deveriam ser inseridos os morfemas de número PL. Ainda assim, os resultados do estudo permitem verificar que as crianças com cerca de dois anos de idade estão sensíveis à expressão gramatical de número no PB, visto que, nesta fase, passam a processar diferentemente as condições gramatical e não-gramatical (cf. Meisel, 1994a:127). A não-diferença entre as condições gramatical padrão e não-padrão indica ainda que a informação de número é extraída da categoria DET, o que implica processamento da concordância no SDET, consoante o modelo de língua a adquirir. Face a estes resultados, os autores atribuem o processamento da concordância a um dispositivo computacional universal, que se encontra disponível aquando dos primeiros contactos da criança com a língua, sendo esta, simultaneamente, agente e paciente da identificação dos aspectos específicos no idioma da comunidade.

Mais recentemente, Castro & Ferrari-Neto (2007) procuram observar se as crianças em aquisição do PEL1 e do PBL1 processam a concordância de número de forma idêntica, isto é no âmbito do SDET. Para o estudo, os autores tomaram em consideração

três perspectivas com assunções diferentes acerca da localização do traço interpretável de número no SDET: (i) em termos minimalistas, o número interpretável situa-se em DET e o não-interpretável em N (Magalhães, 2004); (ii) a expressão morfológica de número expressa-se na categoria funcional SNUM e operação de AGREE (Brito, 1996); (iii) na perspectiva da MD, o morfema de PL *-s* é do tipo *singleton* no PB e do tipo dissociado no PE (Costa & Silva, 2006b). Note-se que “*esta última proposta acentua o carácter de micro-variação paramétrica relativamente à expressão morfológica de número entre duas variedades, assumindo que o D é a posição relevante para a interpretação semântica (em que se inclui a referência de pluralidade) de um DP*” (Castro & Ferrari-Neto, 2007:67). Assim, é esperado que a referência PL se recupere a partir de uma marca morfofonológica unicamente em DET.

Lembre-se que, para o PB, Ferrari-Neto, Corrêa & Augusto (2005) verificaram que as crianças de cerca de dois anos percebem informação morfofonológica relativa a número a partir da categoria DET, processando a concordância de número no âmbito do SDET, numa fase em que, não necessariamente, a morfologia de número se encontra estabelecida na produção (Ferrari-Neto, 2003; Simões, 2004). Dos mesmos resultados se inferiu que a posição estrutural DET já está activa nesta fase inicial do processo de aquisição da linguagem em PB e que a mesma desempenha um papel fundamental na identificação de elementos de uma classe aberta (Höhle *et alii*, 2002:108) e, conseqüentemente, na aquisição de novos nomes (Corrêa, 2001, 2006).

No que diz respeito ao PE, alguns estudos baseados em dados de produção apontam também para a hipótese de as crianças serem sensíveis tanto à posição estrutural sintáctica como à presença de marcas de acordo morfológico de número, uma vez que produzem proto-determinantes em estágios precoces do processo de aquisição da língua (Freitas & Miguel, 1998; Faria, Freitas & Miguel, 1998, 2001; Soares, 1998; Costa & Santos, 2004 [2003]; Castro, 2007; Castro & Ferrari-Neto, 2007). Todavia, Faria, Freitas & Miguel (2001) e Costa & Santos (2004 [2003]) argumentam que os dados disponíveis não concedem evidência convincente para que o proto-determinante não seja visto como um proto-morfema prosódico. Desta forma, o estudo de Castro & Ferrari-Neto (2007) visa testar qual é a posição relevante do morfema para a interpretação da pluralidade por

crianças em processo de aquisição do português, confrontando as duas variantes, PB e PE: (i) se nas duas posições (DET e N); (ii) ou se apenas em uma (DET ou N).

No caso das crianças em aquisição do PE, os resultados apontam para um efeito significativo de *expressão morfológica do número*, com uma quantidade maior de respostas PL para as condições elicitadoras de resposta PL. Contrariamente, e tal como fora revelado no estudo de Ferrari-Neto, Corrêa & Augusto (2005), as crianças em processo de aquisição do PB sugeriram uma interferência da marca –s na marcação PL nos nomes terminados em /s/ no singular. Desta forma, as crianças portuguesas tratam de forma indiferenciada as duas condições do conjunto de estímulos elicitadores de PL, assim como as duas condições do conjunto de estímulos não gramaticais (para PL). Opostamente, as crianças brasileiras entendem os dois estímulos de forma diferenciada, podendo inferir-se que a posição relevante para a percepção da marca de pluralidade é DET, e não N. Daí resulta que as crianças portuguesas identificam a informação morfológica relativa ao número da sua variedade, como identificam também a posição DET como sendo a do morfema de PL para a interpretação da pluralidade. Desta forma, a *performance* destas crianças está mais próxima da gramática-alvo, levando os autores do estudo a relacionarem a constatação com o facto de o sistema de marcação de número em PE ser mais estável. Por outro lado, como as crianças brasileiras são confrontadas com um *input* que obedece a padrões inconsistentes, revelador de várias gramáticas em competição, apresentam um sistema morfológico de marcação de número no SDET que se estabiliza mais tardiamente (Castro & Ferrari-Neto, 2007:74-75).

A comparação dos resultados obtidos para o PE e o PB vem também corroborar os pressupostos de Costa & Silva (2006b), confirmando-se não só que DET é a posição relevante para a interpretação da referência PL de um SDET mas também que a marca morfofonológica de PL no nome é redundante (Castro & Ferrari-Neto, 2007:75), uma vez que também as crianças portuguesas recuperam a referência PL a partir de uma marca morfofonológica apenas em DET, ainda que essa não seja a expressão gramatical da sua variedade. O exposto permite concluir que as crianças com cerca de dois anos de idade são sensíveis à expressão morfológica de número da sua língua, tanto em PE como em PB, e reconhecem que a informação crucial de número é extraída de determinantes nas duas variedades. Assim, ao que tudo parece indicar, a categoria DET estará activa no

processo de aquisição desde uma fase bastante precoce, podendo as crianças valer-se, exclusivamente, da informação morfofonológica relativa ao número para atribuição de referência PL de um SDET flexionado em número, sendo a informação de número no nome entendida como redundante (Kiparsky, 1972:195; Castro & Ferrari-Neto, 2007:75). Assim, o que conta é a expressão da pluralidade em DET, independentemente da variedade de português. Este aspecto torna ainda feliz a hipótese de Costa & Silva (2006b) acerca da variação na expressão morfofonológica de PL em PE e PB.

Numa outra perspectiva, o facto de as crianças colocarem correctamente o adjectivo, desde o início da aquisição, argumenta, igualmente, a favor de elas conhecerem precocemente a sintaxe do SDET, apesar de realizarem variação no uso da concordância. Esta particularidade sugere que o SNUM já está activo desde o início da aquisição, uma vez que a operação de elevação do N é aplicada de forma correcta. Paralelamente, a questão da aquisição do parâmetro *N-drop*, comum no espanhol e no português, vem também reforçar a hipótese acerca do conhecimento precoce da sintaxe do SDET. Estabelecendo uma relação entre estes dois aspectos, Snyder, Senghas & Inman (2001:172) concluem que há relação entre concordância e *N-drop*.

Ao que tudo indica, as crianças detêm os rudimentos do género e número desde a fase mais precoce das aquisição, mas apenas dominam essas categorias mais tardiamente. Este aspecto ficar-se-á a dever às distintas características morfológicas e prosódicas das diferentes línguas, como o comprova o facto de os determinantes surgirem primeiramente no espanhol e no português e, só depois, no francês, alemão e inglês (Godinho, 2005:145). Ainda assim, é preciso ter em conta que a aquisição de género ocorre primeiro do que a de número (Müller, 1994a; Franceschina, 2002, 2005; Corrêa, Name & Ferrari-Netto, 2004; Montrul, 2004; White *et alii*, 2004) um aspecto que Lucchesi (2000a) sustenta, ao analisar sintagmaticamente a inserção de marcas de género no dialecto de HEL-Ba. Deste modo, a captação da regra da concordância de número, por ser funcional, implica, quase necessariamente, a aquisição da regra da concordância de género, que é lexical. Paralelamente, os resultados do nosso estudo (ponto 3.6.7) vêm mostrar que, no plano mórfico, o oposto não constitui condição necessária, ou seja, que a aquisição da concordância de género não implica, necessariamente, a posterior aquisição da concordância de número em SN's de estrutura

reduzida. Koehn (1994:34), em estudos sobre aquisição bilingue das línguas alemã e francesa como L1's, apresenta duas abordagens que podem dar conta do fenómeno da aquisição de morfologia de género e número no SN:

- (i) Item e Processo, que considera que itens e afixos estão separados do radical. De acordo com este princípio, ocorre um processo que une ambos os elementos e faz com que a aquisição das marcas morfológicas seja entendida como um aprendizado de formas (alomorfes) e regras;
- (ii) Modelo de Esquemas (*Schema Model*), que defende que não existem apenas relações entre formas básicas e derivadas em termos de regras de afixação, mas também interligações a nível fonológico, que podem representar generalizações. Deste modo, os afixos não serão armazenados independentemente dos radicais com os quais podem ocorrer, levando a que a estrutura interna das formas morfológicamente complexas seja reconhecida por comparação com outras formas, com recurso a conexões lexicais.

Para Koehn (1994:35), a aquisição de marcas morfológicas é explicada de acordo com o Modelo de Esquemas, estando a gramática sujeita a um processo de maturação (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994; Meisel, 1994a, 1997) e que passa pelas seguintes etapas: (i) o desenvolvimento do conceito semântico subjacente para número (um ou mais do que um); (ii) o reconhecimento de que género e número são sistematicamente codificados em categorias sintáticas específicas; (iii) a aquisição das realizações morfológicas apropriadas desses traços (e de fenómenos de concordância com eles relacionados).

Assim sendo, o falante em processo de aquisição lida inicialmente com a palavra como um todo (Koehn, 1994:49; DeGraff, 1999b:482; Faria Freitas & Miguel, 2001:53), pelo que a aplicação da pluralização não depende da aplicação da marca de género, prevalecendo o Modelo de Esquemas como base para novas formações, o que leva a que a aquisição de marcas se processe com recurso à indução. Por conseguinte, se se considerar que o ponto de partida para a ASL é a gramática da L1 e que as propriedades computacionais da L1 se transferem em bloco para as fases iniciais da ASL, excepto as matrizes fonéticas e os itens léxico-morfológicos (Schwartz & Sprouse, 1996; Sprouse

& Vance, 1999; Siegel, 2006; Sprouse, 2006), a aquisição das marcas dar-se-á por comparação e analogia, acabando as estruturas sintagmáticas do PA por reflectirem um sistema de pluralização idêntico ao do santomense e, conseqüentemente, dos ancestrais substratos deste, isto é, das línguas do grupo níger-congo atlântico. Paralelamente, como as distinções sexuais não são importantes para estas línguas, a marcação de género pode limitar-se ao núcleo do SN (ponto 1.4.2 do presente trabalho), não ocorrendo concordância em género em virtude de os outros constituintes da cadeia sintagmática permanecerem inalterados. Conclui-se, então que a correlação entre concordância de género e concordância de número não é categórica a nível atomístico, indicando não só que “*o encadeamento/aquisicional género => número se verifica de forma mais efectiva no plano sintagmático do que no mórfico*” (Lucchesi, 2000a:265) mas também que a não-concordância de natureza criouliizante tende a ser simultânea. Como simultânea tende a ser também a aplicação dos mecanismos de natureza descriuolizante, confirmada pelos dados de Lucchesi (2000a) para a fala de HEL-Ba, indiciando a intervenção do Princípio da Coesão Estrutural (Haiman, 1983:782) nesta.

Conforme observado por Godinho (2005:145), os fenómenos a que fizemos referência na aquisição da L1 estabelecem paralelismos com o processo de ASL, especialmente no que concerne à aquisição do PtgL2. A diferença reside no facto de, normalmente, a variação desaparecer bastante cedo na aquisição de L1, mas prolongar-se na ASL, chegando a fossilizar-se na interlíngua. Tal ficará a dever-se à incapacidade de aquisição dos traços virtuais não especificados das categorias funcionais (FFFH), já que estas deixam de estar disponibilizadas depois do período crítico do desenvolvimento.

### **2.6.5. Aquisição do sintagma determinante em L2**

A variação registada no processo de ASL é importante para se observar a problemática em torno da refixação dos parâmetros e para análise de aspectos metafísicos relacionados com a aquisição das categorias funcionais na L2 (Godinho, 2005:145). Partindo da aceitação, em linguística, “*que a morfologia explícita conduz à aquisição da sintaxe*” (Montrul, 2004:71, *apud* Godinho, 2005:146), passaremos a analisar a questão dos dados que poderão fornecer evidência sobre a forma como se dá a

aquisição das categorias funcionais da LA, tendo em conta que esta possui traços que não existem na L1 dos aprendentes. Considerando que a aquisição das categorias funcionais pressupõe a aquisição quer de traços quer de formas morfofonológicas, é lícito questionar se os falantes de uma L2 são capazes de adquirir plenamente os traços que não existem na sua L1 e, conseqüentemente, se a variação registada resulta da impossibilidade de adquirir as categorias funcionais marcadas por tais traços (Godinho, 2005:146).

Tendo em conta os pressupostos sobre ASL que avançámos nos pontos 2.3 e 2.4.1.1 do presente trabalho, passemos a observar as diferenças registadas a nível de traços no NUM de algumas línguas que possuem determinantes e uma categoria funcional SDET. Em inglês, a ocorrência de movimento não-visível do N não só determina a ordem ADJ > N como origina concordância fraca em género e número entre determinantes, nomes e adjectivos (Godinho, 2005:147). Deste modo, o SDET em inglês vai apresentar diferenças substanciais em relação ao SDET em espanhol ou português, sendo a mais relevante registada a nível da força dos traços de NUM, mais fracos na primeira língua e mais forte nas outras duas (Carstens, 1991; Ritter, 1991; Valois, 1991; Bernstein, 1993; Montrul, 2004).

Normalmente, os falantes em processo de ASL são foneticamente mais elaborados do que as crianças em situação de aquisição da L1. Este aspecto leva os primeiros a não produzirem proto-determinantes, sendo esta uma das diferenças mais sintomáticas a nível das duas aquisições (Godinho, 2005:147). Contudo, e exceptuando esta particularidade, alguns autores entendem que, em tudo o mais, os dois processos são bastante semelhantes. Produções de aprendentes de espanhol L2 apresentam realização de *N-drop* em contextos idênticos aos da L1 (SADJ, SPREP e SCOMPL), implicando que, *“independentemente de não haver concordância e a omissão de determinantes ocorrer na fase em que ocorrem os proto-determinantes, os restantes aspectos têm um desenvolvimento semelhante nos dois tipos de aquisição”* (Godinho, 2005:147).

A variação registada a nível da concordância em género e número na ASL tem sido constatada por alguns autores (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashes, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003) e entendida como dificuldade na aquisição dos traços que não existem na L1, já que as categorias deixam de estar



disponibilizadas depois do período crítico do desenvolvimento (FFFH – Godinho, 2005:147). Este aspecto é evidente nos estudos sobre aquisição do género no espanhol L2 por parte de falantes de francês L1 (dois idiomas detentores de morfologia de concordância), não se tendo detectado variação significativa na aquisição dos traços do SDET (Bruhn de Garavito & White, 2002:167, *apud* Godinho, 2005:148). Contudo, no que diz respeito à questão da concordância de género, constatou-se uma ultrageneralização da forma masculina em palavras que requerem o género feminino, uma particularidade já notada em outros estudos (Tarone, Frauenfelder & Selinker, 1976; Cain, Weber-Olsen & Smith, 1987; Finneman, 1992; Dewaele & Véronique, 2001).

Dos estudos referidos interessa ainda destacar duas particularidades idênticas em aquisição de L1 e ASL: *“a concordância em género é primeiro estabelecida com o determinante e depois com o adjectivo. O número de erros também é menor com determinantes do que com outros modificadores”* (Godinho, 2005:148). Este segundo aspecto foi também observado em outros estudos sobre aquisição do género (Lucchesi, 2000a; Figueiredo, 2009a), fazendo pressupor que o género masculino parece encerrar o valor de *“default”*, daí a opção pela sua aplicação, quando não existe a certeza sobre o género do nome. Contudo, este detalhe não seria confirmado por McCowen & Alvord (2006), cujo estudo não evidencia diferenças significativas na realização das formas masculina e feminina, um pressuposto vem ao encontro dos achados de Lucchesi (2000a) sobre a aquisição do género pelos falantes de HEL-Ba. De facto, os falantes submetidos a observação exibiram um elevado grau de variação afectando ambas as formas, sendo algumas delas regularizadas posteriormente. Segundo McCowen & Alvord (2006:168), a tendência para o uso da forma masculina só acontece quando o falante não tem a certeza acerca do género do nome. Deste modo, este tipo de variação não pode ser entendido como uma generalização, visto registar-se, de forma esporádica, a nível individual. Por outro lado, a regularização da variação evidencia que os traços podem ser adquiridos numa L2 depois de ultrapassado o período crítico da aquisição, conforme preconizado pelos defensores da FTFA (Cook, 1994; White, 1989, 2003; Bruhn de Garavito & White, 2002; Montrul, 2004; Cook & Newson, 2007 [1988]). Por conseguinte, as dificuldades reveladas na aquisição não são da responsabilidade da L1,

resultando a variação no género do facto de os falantes em processo de ASL não conseguirem fazer a correspondência do traço de género à sua forma morfofonológica (Bruhn de Garavito & White, 2002, *apud* Godinho, 2005:149).

Estudos sobre aquisição do género em espanhol L2 por falantes de inglês L1 (língua que não detém género) e francês L1 (idioma com género) mostram que os primeiros apresentam maior dificuldade na aquisição deste traço. (White *et alii*, 2004, *apud* Godinho, 2005:149). No que concerne à aquisição do número, acontece nivelamento final entre ambos os tipos de falantes, apesar de os ingleses apresentarem um grau de variação ligeiramente mais elevado nas fases iniciais da aquisição, mas que é regularizada posteriormente. Por seu lado, a aquisição de concordância do masculino é a mais problemática e, novamente, são os franceses a apresentar a melhor *performance*. Estas observações permitem concluir não só que a L1 interfere no estado inicial da ASL (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006) mas também que os erros iniciais de morfologia flexional podem ser ultrapassados com a aquisição da morfologia flexional e itens flexionais, refixando-se, assim, os parâmetros. Todavia, o estudo acaba concluindo que, no geral, a interferência da L1 não é relevante na ASL, ao contrário do que sucede com o grau de proficiência, que é determinante neste tipo de aquisição (White *et alii*, 2004, *apud* Godinho, 2005:150).

Os aspectos expostos permitem perceber que o espanhol e o português têm projecções de SDET desde a fase inicial da aquisição e também que, apesar de ocorrer variação na concordância em género e número durante determinados estádios de aquisição da L1 e ASL, ela tem tendência a regularizar-se, fazendo pressupor acesso total à GU em ASL (Cook, 1994; White, 1989, 2003; Bruhn de Garavito & White, 2002; Montrul, 2004; Cook & Newson, 2007 [1988]). Por seu lado, outros autores (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003) propõem que a gramática da interlíngua se restrinja às propriedades formais da L1, o que impossibilita a aquisição de novas propriedades, em virtude de elas não estarem instanciadas nas L1's. Como tal, a aquisição dos traços não será problemática se as duas línguas os possuírem, uma vez “*estão activados na representação da L1 e, por isso, disponíveis na gramática da IL*” (Godinho, 2005:150).

Os estudos abordados evidenciam igualmente que, apesar de se registar variação nas etapas iniciais da aquisição, as categorias sintácticas já se encontram universalmente presentes desde a FL0, manifestando-se, precocemente, através da produção de proto-determinantes. No entanto, ainda que os determinantes sejam adquiridos bastante cedo, a concordância em género e número a nível do SDET só acontece posteriormente em ambas as aquisições, podendo persistir mais tempo na ASL, a ponto de se fossilizar em alguns falantes (cf. Godinho, 2005:151). Deste modo, a CPL-var no SN do PA poderá ser estar intimamente ligada ao facto de o santomense e as línguas do grupo níger-congo atlântico, a exemplo do que Godinho (2005:385) constatou relativamente aos falantes de cantonenseL1 em aquisição de PtgL2/LE, não possuírem traços de género e número.

#### **2.6.6. Aquisição da concordância de PL no sintagma determinante**

A aquisição da concordância de PL no SDET pressupõe a detecção da categoria número e a realização dos processos morfossintácticos inerentes à mesma, levando as regras da formação das palavras (*Word Formation Rules*) a actuarem em duas direcções (Aronoff, 1976; Spencer, 1991; Cerqueira, 1994:119):<sup>88</sup>

*“(a) As regras são responsáveis por garantir a formação das palavras da língua «estabelecendo as operações que produzem os resultados desejados»;*

*(b) As regras operam em sentido inverso, não para formar palavras mas para analisar as suas partes constituintes” (Godinho, 2005:152).*

Estes aspectos implicam que as flexões só poderão ser alojadas se a palavra for analisada em função dos traços morfológicos das diferentes categorias morfossintácticas accionadas. Deste modo, Faria, Freitas & Miguel (2001:51-54, *apud* Godinho, 2005:153-154 e nota 204), ao debruçarem-se sobre as projecções funcionais do SN do PE, verificaram que o morfema PL realizado com a fricativa /s/, em final de palavra, acontece, geralmente, em sílabas não acentuadas, em virtude de o acento principal em PE gerar, por norma, palavras paroxítonas. Por outro lado, como as sílabas tónicas protegem a informação fonológica dos processos fonéticos de apagamento, (Delgado-Martins, 1996:319), as fricativas morfológicas /s/, marca de PL, deveriam ser os últimos itens a emergir, visto ocorrerem em posição final, ou seja, em sílabas átonas. Todavia, e

conforme apontado, listado e comentado por Godinho (2005:153-155), a propósito do momento em que sobrevêm as fricativas que determinam a pluralização, a *saliência fónica* não é relevante para a inserção das mesmas. Face a este achado, que corrobora o postulado de Cerqueira (1994) acerca da disponibilidade inata de um dispositivo para recepção e descodificação do *input* fonológico, verifica-se que “*a aquisição «precoce» das fricativas em codas morfológicas, em posição final de palavra, em sílabas não acentuadas, deve-se ao seu estatuto morfológico no sistema da língua (português europeu) e à posição (final) que ocupam na palavra, e que as crianças são capazes de as isolar e interpretá-las como um constituinte separado – um morfema – ao nível da palavra*” (Godinho, 2005:154-155). Assim, apesar de as crianças revelarem incapacidade de isolar e interpretar separadamente os morfemas a nível das palavras nas fases precoces da aquisição bilingue (cf. projecto DUFDE, Koehn, 1994:49) e ASL (cf. DeGraff, 1999b:482), conseguem perceber as propriedades semânticas e morfossintáticas dos elementos lexicais.

As realizações que consubstanciam o nosso *corpus*, a exemplo do que sucede com os dados produzidos por falantes de PtgL1 (Guy, 1981a; Cerqueira, 1994; Scherre, 1988; Lopes, 2001; P. Andrade, 2003; Baxter, 2004, 2009; Jon-And, 2008, 2009) e PtgL2/LE (Godinho, 2005), não só espelham realizações decorrentes do contacto entre línguas mas evidenciam também que, a exemplo do que sucede com as produções registadas na aquisição de L1 e bilingue, a marcação de PL pode revelar variação, não incidindo em todos os constituintes do SN (ou do SDET) que a requerem. Não obstante, o facto de ocorrerem marcas de PL, mesmo que esporádicas, faz pressupor que os falantes em questão identificam a fricativa /s/ em final de palavra como um marcador do valor PL. Assim, estamos perante dois padrões de marcação, semelhantes aos identificados por Godinho (2005:156) nas produções de falantes do cantonense L1 em aquisição do PtgL2/LE: (i) o sistema de marcação PL respeita as regras de aplicação da flexão em número do português e esta é inserida em todos os elementos da frase que a requerem (exemplo [190]); (ii) o sistema de marcação PL não respeita as regras de aplicação da flexão em número do português e esta não é inserida na totalidade dos elementos da frase que a requerem (exemplo [191]):

[190] PA:  
*os homens holandeses vieram* [CASTEH1]

[191] PA:  
*As quatro roda pa cima* [OSVALH1]

Considerando que os quantificadores, determinantes e modificadores adjetivos concordam com o núcleo sintagmático, então estes elementos deverão ser portadores da marca morfológica PL patente no mesmo (exemplo [190]). Todavia, o exemplo [191] contraria este entendimento, evidenciando que este tipo de traço configura uma categoria vazia no núcleo. Deste modo, e de acordo com a Hipótese do SDET (Picallo, 1991; Carstens, 1991; Ritter, 1991; Cerqueira, 1994; Valois, 1991; Bernstein, 1993; Cinque, 1994; Koehn, 1994; Müller, 1994a; Montrul, 2004), que postula a existência de categorias funcionais entre o SDET e o SN, então é lícito inferir que o traço de número [+/- plural] do núcleo nominal é mapeado em NUM na sintaxe (visível ou invisível), consoante a sua força morfológica, a qual pode determinar a elevação do N, a partir do seu posicionamento original (Cerqueira, 1994:121, *apud* Godinho, 2005:157).

Segundo o Princípio da Projecção Alargado da Teoria X-barras (Rizzi, 1982; Jaeggli, 1984; Burzio, 1986; Jaeggli & Safir, 1989; Grimshaw, 2005 [1991]), N, um item lexical, e DET, um elemento funcional, compartilham os mesmos traços de categoria sintáctica [-V, +N], apesar de, funcionalmente, pertencerem a classes distintas. A distinção entre ambos, a nível de traços funcionais, é justificada por em termos de desdobramento (Cerqueira, 1994:121-125; Godinho, 2005:157-165), ou seja, é assegurada por um traço funcional, no caso o traço {F}, que se vai desdobrar em dois: {F0}, para as categorias lexicais; e {F1}, para as categorias funcionais. Assim, o valor {F} de um nó não consubstancia parte da sua análise categorial, visto não possuir o mesmo valor dos traços categoriais na formação de projecção alargada nem ser um traço binário que estabeleça interacção com outros traços categoriais (Godinho, 2005:157-158). Desta forma, as categorias funcionais são essencialmente relacionais, já que estabelecem relações com categorias lexicais, delas se constituindo a sua parte funcional, como sucede com o SDET, relativamente ao N. (Cerqueira, 1994:121; Godinho, 2005:158).

Tendo em conta a questão do desdobramento do traço funcional {F}, proposta por Cerqueira (1994:121), as categorias derivadas de DET e N, listadas por Godinho (2005:158), serão as seguintes:

(1)	DET	[-V, +N]	{F1}	L0
	DET'	[-V, +N]	{F1}	L1
	SDET	[-V, +N]	{F1}	L2
(2)	N	[-V, +N]	{F0}	L0
	N'	[-V, +N]	{F0}	L1
	SN	[-V, +N]	{F0}	L2

Por seu lado, as projecções não-alargadas (ou de núcleo perfeito) são estabelecidas de acordo com os seguintes parâmetros (Godinho, 2005:158-159):

#### PROJECCÃO DE NÚCLEO NÃO-ALARGADO

$x$  é núcleo perfeito de  $y$ , e  $y$  é uma projecção perfeita de  $x$ , caso:

- (i)  $y$  domine  $x$ ;
- (ii)  $y$  e  $x$  compartilhem de todos os traços categoriais [-V, +N];
- (iii) todos os nós entre  $x$  e  $y$  tenham os mesmos traços categoriais;
- (iv) o valor de S de  $y$  seja igual ao valor de S de  $x$ .

Segundo estes princípios, SDET é projecção de DET e DET', visto compartilhar traços categoriais e flexionais com estes. Porém, SDET não é projecção de N, N' ou SN, visto não compartilharem traços flexionais. Por seu lado, SN é projecção de N' e N.

Quanto às projecções alargadas, são definidas de acordo com os seguintes princípios (Godinho, 2005:159):

#### PROJECCÃO DE NÚCLEO ALARGADO E PROJECCÃO ALARGADA

$x$  é núcleo alargado de  $y$ , e  $y$  é uma projecção alargada de  $x$ , caso:

- (v)  $y$  domine  $x$ ;
- (vi)  $y$  e  $x$  compartilhem de todos os traços categoriais;
- (vii) todos os nós entre  $x$  e  $y$  tenham os mesmos traços categoriais;
- (viii) o valor de S de  $y$  seja maior que o valor de S de  $x$ , onde  $n$  actue entre  $x$  e  $y$ , se  $y$  dominar  $x$  e  $n$ . Por outro lado,  $n$  domine  $x$ , mas  $x$  não domine  $y$ .

De acordo com estes pressupostos, é possível combinar projecções de núcleo perfeito com outras maiores, o que leva o SDET a deter cinco núcleos alargados: DET', DET, SN, N' e N. Simultaneamente, SDET é projecção alargada de DET, DET', SDET, N, N' e SN, o que determina que um sintagma seja, ao mesmo tempo, COMPL de um núcleo e parte da mesma projecção alargada, caso observe os requisitos exigidos pelo Princípio da Projecção Alargada (Cerqueira, 1994:122; Godinho, 2005:159).

Se dos núcleos de uma projecção alargada é projectada uma dada informação, todos os seus traços deverão concordar com a projecção SN > SDET, manifestando-se caso a morfologia registe os seus valores (Grimshaw, 2005 [1991]:114). Todavia, nos casos em que não ocorre concordância entre DET e N a manifestação não é visível, já que esta se torna também alargada, devido à projecção (Godinho, 2005:159-160).

Além destas determinações, a Teoria da Projecção Alargada garante igualmente que a projecção COMPL de um núcleo seja parte integrante dessa projecção, visto compartilharem traços funcionais e categoriais. Portanto, se número é traço de N, sê-lo-á igualmente de DET, o que leva a projecção de SDET e DET a incluírem traços do núcleo alargado N e SN. Perante tal, a concordância é um fenómeno derivado que assegura a correcta elaboração das estruturas nominais (Cerqueira, 1994:123; Godinho, 2005:160).

Sendo a Teoria da Projecção Alargada uma das componentes da Faculdade da Linguagem, o pareamento do traço morfológico de número entre todos os elementos do SN, que permite elaborar a concordância visível, só será possível caso o falante receba a informação acerca do traço *-s* (Godinho, 2005:160). Contudo, tal pode não suceder, conforme exemplos [190] e [191]. Nestes casos, a não-realização fonética da marca PL licenciada por {s} obedecerá a uma operação de identificação semântico-funcional, idêntica à que afecta a categoria sintáctica *pro* Cerqueira (1994:125, *apud* Godinho, 2005:161). De facto, a ocorrência de sujeitos pronominais nulos em línguas como o português depende da elaborada morfologia do seu sistema flexional de concordância, o qual permite a recuperação dos traços gramaticais do sujeito a partir do conteúdo morfológico das terminações verbais (Raposo, 1992:477-478). Deste modo, a projecção alargada vai não apenas assegurar os licenciamentos formal e funcional mas também permitir que se formulem os pressupostos que determinam o licenciamento de {s} nulo, os quais, por seu lado, permitem que se compreenda o porquê de não acontecer

licenciamento de {s} nulo quando o SN é configurado unicamente por N (Godinho, 2005:161):

LICENCIAMENTO DE {S} NULO

{s} pode ser nulo, caso:

- (ix) seja parte integrante de uma projecção alargada;
- (x) a sua presença seja notificada por algum item da projecção.

Levando em consideração quer os princípios do licenciamento de {s} nulo quer o facto de a concordância ser um fenómeno derivado, Godinho (2005:161) estabelece os considerandos finais sobre as condições de base que determinam a segunda: (i) advir da relação entre SPEC e núcleo; (ii) resultar de projecção. No segundo caso, caberá ao SDET, enquanto projecção máxima de N e SN, projectar os traços destes dois núcleos alargados, permitindo que eles estabeleçam o pareamento de traços entre si. Por sua vez, a marca PL em DET vai permitir que a mesma não seja visível em N, ou seja, vai tornar possível o licenciamento de {s} nulo neste.

Tendo em conta tais considerandos, determinadas realizações dos falantes de Almojarife (exemplo [192]) permitem constatar que as mesmas encerram representações sintácticas distintas, identificadas consoante o contexto em que são produzidas:

[192] PA:

*lodava as conta com el poque a... a roça piquena.* [CELESH3]

Nesta realização, o SN “*a roça piquena*” permite duas configurações estruturais (cf. Godinho, 2005:162):

- (i) SDET composto pelo artigo “a”, o nome “roça” e o adjectivo pós-nominal “piquena” (pequena);
- (ii) Uma construção predicativa com apagamento de cópula, “*a roça era piquena*”, na qual o adjectivo é parte integrante do SV.

Partindo do princípio que, na aquisição de L1 e ASL, N e DET+N são produzidos numa mesma fase (Montrul, 2004:42-43; Godinho, 2005:161-162), é lógico inferir que



os falantes almorarifanos dominem a categoria funcional SDET, apesar de nem sempre a realizarem. Este entendimento poderá ser extensivo às questões que determinam a ausência visível, na sintaxe, da marcação PL em todos os itens do SDET (exemplo [191]). Assim, o facto de os falantes de Almorarife não marcarem o número em todos os elementos do SN não implica, necessariamente, que estejam a violar princípios gramaticais, uma vez que essas sequências se assemelham às que são realizadas na FLf de falantes nativos da LA (cf. Godinho, 2005:162). Para a autora, os pressupostos sobre a elaboração da concordância PL, por um lado, e acerca da forma como se manifesta a ausência de marcas PL, por outro lado, argumentam a favor da proposta que o ponto de partida para a ASL é a gramática da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), pelo que a variação resulta de um problema na correspondência de traços, e não da impossibilidade de aquisição destes, seja por falha no restabelecimento dos parâmetros (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003) seja pela ausência dos mesmos traços na L1 (Leung, 2003:206, *apud* Godinho, 2005:162). Portanto, com a aquisição intermédia de morfologia flexional relevante e de itens lexicais, é possível refixar os parâmetros e atingir um estágio final de aquisição.

Não obstante, segundo Hawkins (1993, 1998, 2001), os traços [–interpretáveis] não estão presentes no processo de aquisição, em virtude de serem os mais difíceis de adquirir. Por seu lado, Franceschina (2002, *apud* Godinho, 2005:163) sustenta que a aquisição dos traços [–interpretáveis] de número e género não oferece dificuldades na ASL, caso os falantes os possuam nas suas L1's. Como este tipo de traços é opcional nas línguas, os falantes nativos daquelas que os não possuem experimentam dificuldades na sua aquisição, quando em processo de ASL's nas quais eles se encontram instanciados. É o que acontece com os falantes da generalidade dos idiomas do grupo níger-congo atlântico ou do santomense, submetidos à aquisição do Ptg, uma língua parametrizada, já que as suas L1's não detêm concordância do tipo sufixal em número e género. Deste modo, é lícito pressupor que a CPL-var registada no SN dos informantes do PA poderá ser motivada pelo facto de o traço de número não estar instanciado nas L1's dos seus antepassados (cf. Godinho, 2005:164). Estudos sobre a concordância PL no SN do PB (Braga & Scherre, 1976; Braga, 1977; Scherre, 1978; Ponte, 1979; Nina, 1980; Guy,

1981a; Scherre, 1988; Lopes, 2001; P. Andrade, 2003; Tieppo, 2003; Baxter, 2009) e de variedades africanas de português (Moreno & Tuzine, 1998; Baxter, 2004, 2009; Figueiredo, 2008, 2009a, 2009b; Jon-And, 2008, 2009) têm revelado que a presença ou ausência de marcas não é aleatória (Tarallo, 1986:8), já que os contextos linguísticos e/ou extralinguísticos que as motivam permitem que se prevejam as tendências da marcação (ponto 2.5 do presente trabalho). Alguns destes trabalhos sobre o PB, como os de Guy, (1981a; *apud* Godinho, 2005:164), Lopes (2001, *apud* Godinho, 2005:164), P. Andrade (2003) ou Baxter (2009), apontam a hipótese de a variação resultar de um processo de transmissão linguística irregular, que afectou a aquisição da língua portuguesa por parte de ameríndios e escravos africanos deslocados geográfica e culturalmente pelos colonizadores. Quantos às variedades africanas, exceptuando o trabalho de Moreno & Tuzine (1998), que não aborda a questão da aquisição, todos os outros referem que a reestruturação terá sido condicionada por factores que ancoram na aquisição do PtgL2 e PtgL1 nativizado, em situação de contacto que envolveu influências das línguas nativas do grupo níger-congo atlântico ou línguas crioulas. Alguns destes trabalhos têm também corroborado a hipótese levantada por Guy (1981a:3001-302) de que o ancestral substrato africano será a potencial base que origina a CPL-var no PB, cuja marcação, tal como nas variedades referidas, revela tendência a ser inserida no elemento imediatamente à esquerda do núcleo do SN.

### **2.7. Aquisição das categorias funcionais da L2: os casos do santomense e do PA**

Os radicais dos nomes não possuem flexão em género e número no santomense nem nos seus substratos. Nestes, os classificadores nominais, apensos inicialmente, regem a concordância PL. Nas línguas bantu, por exemplo, todos os elementos que compõem o SN e se relacionam com o nome recebem marcação temática, mantendo-se invariáveis os radicais dos constituintes da cadeia sintagmática. Estes aspectos determinam que os falantes destas L1's, em situação de aquisição do PtgL2, tenham que fixar novas propriedades sintácticas na interlíngua, a fim de elaborarem enunciados com acordo no SN, já que as regras da LA determinam que a concordância no PE se processe com recurso a morfemas sufixais e dissociados ao longo da estrutura sintagmática.

A possibilidade de ocorrer conflito na aquisição dos traços que determinam a marcação em género e número na interlíngua leva-nos a observar como se processa a concordância no SN do PE, semelhante à que ocorre no domínio da frase. Sem nos determos muito em pressupostos generativistas, interessará reter que a concordância ao nível do SN dependente, entre outras, das seguintes condições: (i) relações entre SPEC e núcleo; (ii) categorias funcionais relacionadas com a concordância; (iii) movimento do núcleo internamente ao SN; (iv) mecanismos pronominais. A partir da Teoria X-Barra, obtivemos os diagramas das Figs. 19 e 20, que determinam que a categoria funcional NUM, localizada entre o DET e o SN, é aquela onde se situam os traços de número (Picallo, 1991; Carstens, 1991; Ritter, 1991; Valois, 1991; Bernstein, 1993; Cinque, 1994; Montrul, 2004). Deste modo, se tivermos em conta o posicionamento rígido de determinados adjectivos do português, como é o caso dos adjectivos relacionais ou que digam respeito a formas e cores (exemplos [193a] e [193b]), o nome movimenta-se para permitir a concordância, podendo este movimento traduzir uma propriedade económica da língua, à qual os falantes recorrem para evitarem a inserção de mais elementos lexicais na estrutura frásica. Contudo, como a posição DET se encontra preenchida, é necessário criar mais derivações nas árvores sintagmáticas para se inserir o nome. Deste modo, para o exemplo [193a] teríamos, então, a configuração da Fig. 22:

- [193] PTG:  
 a. *A bola vermelha.*  
 b. *\*A vermelha bola.*

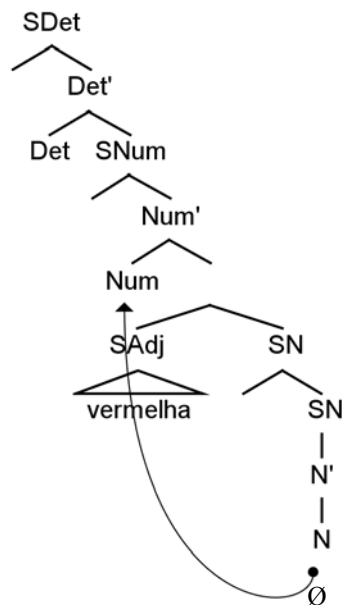


Fig. 22. Movimento de N para NUM.

Como o ADJ detém uma posição fixa, é o N que sofre elevação, transitando da área dos itens lexicais para a zona dos elementos funcionais, ou seja, eleva-se da posição de N para a de NUM, assumindo a função de núcleo do SN. A posição de especificador vai então torná-lo morfologicamente visível e, ao entrar em relação próxima com os outros itens do SN, permite que os seus traços [+interpretáveis] de número procedam ao mapeamento dos correspondentes traços [-interpretáveis], originando a concordância sintagmática. No entanto, convém referir que o género possui um traço bastante lexical, revelando-se indiferente na especificação da entidade, enquanto o número possui um traço mais funcional, tendo importância para a construção da referência. Assim sendo, a aquisição de género ocorre antes da aquisição de número no plano da concordância dentro do SN, (Müller, 1994a; Franceschina, 2002, 2005; Corrêa, Name & Ferrari-Netto, 2004; Montrul, 2004; White *et alii*, 2004).<sup>89</sup>

Como as gramáticas do santomense e seus substratos do grupo níger-congo atlântico não apresentam sistema de marcação idêntico ao do PE, a categoria N, ao elevar-se para a posição nuclear, não vai exibir morfema de PL, originando a concordância a nível meramente temático. Tendo ainda em conta que a competência em L2 é igualmente condicionada pela ecologia das línguas, deve dar-se especial atenção ao contexto em que esta ocorre: endolingue ou exolingue. O santomense, substrato do PA, desenvolveu-se em local onde um grupo dominante (os colonos portugueses), que veio de fora com o objectivo de formar uma sociedade nova, estabeleceu contacto com comunidades geográfica e culturalmente deslocadas por esse mesmo grupo dominante. Inicialmente, a população deslocada, constituída por poucos elementos, era submetida a um *input* regular da LA, pelo que apenas fazia uso das suas línguas de origem em situações restritas, isto é, em ambiente familiar. Contudo, esta população variada sobrepôs-se em número ao do grupo dominante em determinado momento da constituição histórica da nova sociedade. Assim, apesar de associada cultural e economicamente a este, a maioria dos membros da sociedade dominada, face a um contacto cada vez mais distanciado em relação aos detentores da LA, via os estímulos linguísticos desta diminuírem acentuadamente. Estes aspectos determinaram, então, que o uso das L1's se mantivesse, ao mesmo tempo que se desenvolvia o santomense. Paralelamente, como o sistema das línguas do grupo níger-congo atlântico faladas por eles não possui a categoria DET,

impossibilitava-os de fazerem a correspondência das características formais às formas morfofonológicas (cf. DeGraff, 1999b; Prévost & White, 1999 – MSIH). Deste modo, o DET surgia fonologicamente interpretado como um afixo semântico apenas no início do N (Baxter, 2004:120),<sup>90</sup> e não como uma categoria funcional que determina a construção a nível sintáctico, originando a instanciação de novos/errados parâmetros (Gonçalves, 2004:235) na L2, daí resultando itens gramaticais relexificados a partir de estruturas do substrato e lexemas derivados do superstrato. Posteriormente, a instanciação em cadeia dos novos/errados parâmetros nos DLP's determinou uma situação de transmissão linguística irregular e inviabilizou a possibilidade de um reajustamento paramétrico, fazendo com que os novos/errados parâmetros fossem tomados como modelo para uma nova reinstanciação paramétrica do PAL1. Consumava-se, deste modo, a variação no sentido da mudança que, por um lado, acentua a distância tipológica entre o PE e a capacidade gramatical (ou língua-I) atingida pelos falantes do PA e, por outro lado, faz com que alguns traços do santomense e das ancestrais línguas do substrato níger-congo atlântico sejam retidos no PtgL2 e sobrevivam na gramática do PAL1. Em sequência de tal, algumas propriedades gramaticais do PE perderam-se também no *continuum* de aquisição, originando a *especificidade gramatical* do PA.

Assim sendo, as variações que se detectam no DET do PA terão origem em dificuldades de percepção do tipo fonológico, visando a construção mental de concordância temática, e não tanto do tipo estrutural, já que estas efectuam a concordância a nível sintáctico. Porém, como a aprendizagem da L2 pode acontecer por indução (Selinker, 1972:216), é também de considerar que, a nível mais avançado da aquisição do PA, sobretudo em falantes escolarizados, ocorra a percepção de que se está perante itens lexicais distintos, passando a concordância a operar sintacticamente. Como o valor semântico expresso pelos afixos flexionáveis não é suficientemente robusto para determinar a sua sobrevivência nos estágios mais avançados do desenvolvimento da interlíngua (Plag, 2008a:125), ocorre, posteriormente, uma refixação paramétrica e conseqüente substituição do sistema de pluralização do santomense e dos seus substratos níger-congo atlânticos pelo sistema de pluralização sufixal do PE, levado a cabo pelos falantes bilingues e mais escolarizados de Almoxarife.

## 2.8. A refixação dos parâmetros

No ponto 2.4.1.1 do presente trabalho, referimos que a corrente linguística que defende o acesso indirecto à GU entende que o ponto de partida para a ASL é a gramática da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006). Por conseguinte, as propriedades computacionais da L1 transferem-se em bloco durante as fases iniciais da ASL, com excepção das matrizes fonéticas e dos itens léxico-morfológicos. Nesta conformidade, a construção gramatical é orientada pela GU, que mantém sempre os seus princípios activos, fazendo com que, no processo de ASL, os aprendentes adquiram valores paramétricos a partir dos valores de parametrização da L1 e cometam erros de morfologia flexional.

O Programa Minimalista (Chomsky, 1996 [1995]) advoga que os *traços-phi* não-interpretáveis, entre os quais se encontram os traços de género e número, têm como finalidade activarem as operações sintácticas. Os traços não verificados são valorados e, caso necessário, apagados através da concordância com traços valorados a partir de outro núcleo da estrutura frásica. Deste modo, os *traços -phi* [*ugen*] e [*unum*] vão mapear o conjunto de traços valorados para o mesmo valor binário do traço em questão, que terá de partir de um elemento nuclear, no caso a categoria N, projectada em NUM. No PE, acontece uma relação de especificador em que o núcleo é morfologicamente visível, permitindo que os elementos em questão possam copiar ou compor o seu valor a partir deste núcleo, sendo verificados, posteriormente, por forma a produzir-se a concordância em todos os elementos de núcleo e seus complementos. Os morfemas de género e número são, então, do tipo dissociado, espalhando-se por todos os contextos do mesmo caso dentro do SN. Finalizadas as operações sintácticas, os elementos são remetidos para a FL, com radicais-terminais para um determinado valor. Como temos vindo a referir, no caso do género, ele vai deter características bastante lexicais (revela-se indiferente para especificar a entidade), enquanto o número é mais funcional, construindo a referência.

Sem assumirmos que a variação é provocada por mecanismos de transferência directa da L1, não poderemos, todavia, deixar de levar em consideração que os aprendentes adultos de uma L2 são proficientes, à partida, na sua L1. E no estado inicial da ASL, tal pode determinar que os aprendentes adquiram valores paramétricos a partir dos valores

de parametrização da L1. Como o PE é uma língua que detém valor positivo para ambos os *traços-phi* de género e número ao nível da sintaxe, este aspecto pode ser determinante na fase inicial da aquisição do PtgL2 quer por falantes do santomense quer por falantes dos seus substratos níger-congo atlânticos, já que estes aprendentes detêm L1's caracterizadas por configuração negativa para ambos os *traços-phi* referidos. De facto, tais idiomas possuem um reduzido sistema flexional, pouco elaborado em traços virtuais não especificados (as categorias funcionais). Portanto, não só estão privados de flexão sufixal de género e número como também marcam estas categorias de forma distinta do PE, recorrendo a classificadores pré-nominais que configuram morfemas do tipo *singleton*. Paralelamente, e no caso específico do género, este pode ser assinalado na morfossintaxe dos nomes com o traço semântico [+animado], já que, nas mencionadas línguas, os falantes expressam também o género recorrendo, por vezes, a nomes adjetivados colocados depois do nome, como acontece com os termos *mulher* ou *fêmea* (género feminino) e *homem* ou *macho* (género masculino) (ponto 1.4.2 do presente trabalho).

Como se depreende, esta opcionalidade de fixação de valores poderá levantar problemas na aquisição do PtgL2, levando à instanciação de novos/errados parâmetros (Gonçalves, 2004:235), aspecto este que seria minorado ou eliminado, caso os falantes partissem para a aquisição com valores positivos (cf. Franceschina, 2002:100), à semelhança do que acontece na LA. Por outro lado, e embora o modelo de acesso indirecto à GU determine também que, com a aquisição intermédia de morfologia flexional relevante e de itens lexicais se consegue ultrapassar os erros de morfologia flexional característicos das fases iniciais da aquisição (Schwartz & Sprouse, 1996; Sprouse, 2006), ou seja, se refixem os parâmetros, a verdade é que a FFFH entende que as categorias funcionais, nas quais se encontra incluído o número, deixam de estar disponibilizadas depois do período crítico do desenvolvimento, impedindo a refixação para tais categorias. Em consequência deste aspecto, acontece a ausência em cadeia das referidas categorias funcionais no *input*, o que impossibilita o posterior reajustamento paramétrico de acordo com os modelos da LA, fossilizando-se determinadas estruturas transferidas das ancestrais L1's, e que são responsáveis pela variação registada no PA. Portanto, será importante que se observem também quer as questões inerentes à

concordância variável ao nível do SN do PA quer as condições que poderão determinar as fossilizações e, conseqüentemente, consumarem a mudança da variação.

## 2.9. Concordância variável ao nível do SN

No que diz respeito à concordância variável, o PB tem-se revelado como um caso paradigmático para o debate acerca da visão completamente configuracional da concordância. Ao contrário do PE, que apresenta um padrão uniformizado de concordância verbal para a relação sujeito/verbo e que estabelece paralelos com o da concordância interna ao SDET, o PB exhibe dois padrões de concordância verbal e apenas um para a concordância interna ao SDET, inviabilizando um paralelo de acordo com o que sucede no PE. Nos casos de concordância verbal, verificam-se, então, três situações-base no PE: (i) realizações com ordem frásica padrão (SVO) exibindo concordância entre todos os constituintes frásicos (exemplo [194a]): (ii) aplicação de concordância quando ocorre inversão do sujeito na ordem da frase (exemplo [194b]); (iii) uso de concordância quando se levam a cabo realizações com verbos inacusativos (exemplo [194d]).

[194] PE:

- a. *Os carros são bonitos.*
- b. *Comeram, os meninos, o doce.*
- c. *\*Comeu, os meninos, o doce.*
- d. *Ontem chegaram dez navios ao porto de Lisboa.*
- e. *\*Ontem chegou dez navios ao porto de Lisboa.*

No que diz respeito à concordância interna ao SDET, o PE não levanta problemas em termos descritivos, já que apresenta concordância entre todos os elementos frásicos que a exigem:

[195] PE:

- a. *Os carros.*
- b. *Os carros bonitos.*
- c. *Os primeiros carros.*
- d. *Os meus carros.*

No caso do PA, a similaridade com realizações do PB é bastante evidente, sobretudo no que respeita à concordância interna ao SDET. Assim sendo, os pressupostos avançados para o PB são igualmente válidos para a fala de Almojarife. Concretamente



sobre a concordância verbal, o PA, a exemplo do PB, exhibe variação na relação sujeito/verbo, que pode ser de dois tipos. Desta forma, com a ordem frásica padrão (SVO), a variação pode realizar-se das seguintes maneiras: (i) ocorrendo concordância entre verbo e apenas um dos elementos do SN sujeito, por norma o item pré-nuclear (o determinante), isto é, aquele que revela tendência para receber a marca de PL (exemplo [196]); (ii) ausência de concordância entre verbo e elemento pré-nuclear (exemplo [197]).

[196] PA:  
*Os dirigente ão são homens assim* [CASTEH1]  
“Os dirigentes não são homens assim”

[197] PA:  
*Depois as piquena ão tá respondê* [OSVALH1]  
“Depois, as pequenas não responderam”

Por outro lado, acontecendo inversão do sujeito na ordem da frase, isto é, construções do tipo VS,<sup>91</sup> ocorre quase sempre falta de concordância entre estes dois elementos (exemplo [198]):

[198] PA:  
*Então táqui os home que veio preso* [MANOH3]  
“Então estavam aqui os homens que vieram presos”

Deste modo, temos dois tipos de concordância variável para o sintagma verbal, o que não acontece com a concordância variável interna ao SDET, que é idêntica nas construções em ambas as realizações SV e VS. Ainda assim, a marca de PL pode incidir sobre outros itens do SN, que não o elemento pré-nuclear. Não obstante, como a marcação é quase categórica no item pré-nuclear em todas as gerações, é consensual que se tenha esta como o modelo para efeitos de análise, isto é, se admita como paradigmática a concordância interna ao SDET apontada nas ocorrências do tipo da registada no exemplo [191]. De facto, o que sucede com as marcações que divergem desta é que, por norma, denotam gramáticas em competição no mesmo indivíduo, fruto de causas várias, que podem ser de carácter linguístico e/ou social. De facto, se olharmos para os exemplos que se listam seguidamente, constatamos que a variação é toda realizada pelo mesmo falante, e pode ir desde a marcação PL que incide apenas no

determinante (exemplo [199a]) até à marcação que afecta ambos os itens (exemplo [199c]), passando ainda pela marcação registada somente no nome (exemplo [199b]). Assim, nunca é de descartar a possibilidade de estas gramáticas em competição poderem dificultar uma análise em termos de se estabelecer uma visão configuracional uniforme da concordância interna ao SDET, tanto no PB como no PA:

- [199] PA:
- a. *Os dirigente ão são homens assim* [CASTEH1]  
“os dirigentes não são homens assim”
  - b. *Eis a razões que eu já ão stô cá.* [CASTEH1]  
“eis as razões pelas quais eu já não estou cá”
  - c. *As crianças já estão adormecido* [CASTEH1]  
“As crianças já estavam a dormir”

Como se constata, o que se passa internamente a nível do SN não é o mesmo que se passa no âmbito da frase, já que não é possível efectuar-se uma análise uniforme de todos os fenómenos de concordância na segunda. Nesta conformidade, e ao contrário do que sucede com o PE, não existe paralelo entre concordância verbal e concordância interna ao SDET. Ainda assim, o alto peso de pluralização evidenciado pelos determinantes pré-nucleares permite considerar que a posição DET é marcada com a morfologia de número, embora as ocorrências com possessivos do tipo das apontadas por Scherre (1988:163) (ausência de marcação na primeira posição e inserção de marca PL no possessivo em segunda posição) parecem retirar alguma robustez a este pressuposto no PB. Contudo, como veremos no ponto 4.1.2.6.1, esta particularidade não encontra paralelo no PA, já que, ao contrário do que sucede no Brasil, a expletividade do definido não é tão sistemática na fala almojarifana como no PB.

Relativamente à concordância entre SN e S<sub>ADJ</sub>, se tomarmos como exemplo uma oração pequena (OP) (exemplo [200]) e levarmos em conta que a CONC emerge entre SUJ e V até ao momento do *spell-out*, teremos a configuração sintáctica da Fig. 23 para representar os predicativos adjectivais e respectivo movimento sintáctico para CONC, proposta por Chomsky (1996 [1995]:175) para este tipo de orações:

[200] PE:  
*O João é alto.*

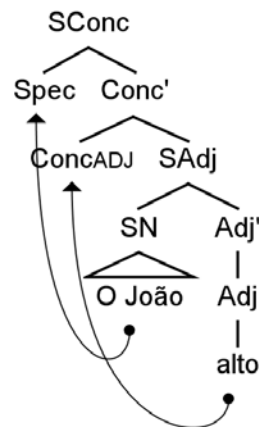


Fig. 23. Sintaxe na OP para concordância  
 (a partir de Chomsky, 1995:175).

Na configuração proposta, CONCADJ diz respeito à mnemónica de um conjunto de *traços-phi* relacionados com o adjectivo (Silva, 2010:26). O SN sobe para SPEC, enquanto o ADJ ascende a CONCADJ, gerando a estrutura interna entre SN-ADJ nas construções predicativas, ou seja, a configuração da OP. Na estrutura S-, o complemento de SV [SN o João] sobe para receber caso nominativo e CONC verbal, originando a estrutura *O João é alto*. De acordo com a estrutura minimalista da concordância (Fig. 18), o SN *o João* estabelece três tipos de relações: (i) Uma relação de Caso com [T CONC-S], daí resultando a relação verbal [[T CONC-S] V]; (ii) Uma relação de concordância com CONC-S, originando a complexidade verbal de *ser*; (iii) Uma relação de concordância com CONC da Fig. 23, estabelecendo a complexidade adjectival. Note-se, então, dois aspectos neste tipo de construções: (i) a OP impõe como requisito que o SN esteja fora do SAdj; (ii) a sintaxe na OP para a concordância estabelece paralelismo com o movimento de N para NUM (Fig. 22).

Outro aspecto a ter em consideração na CPL-var do PB e do PA é a assimetria entre posições pré e pós-nucleares. Para compreenderem esta, os teóricos têm recorrido também à MD, que advoga que o *output* da sintaxe é o *input* da morfologia (esta lê aquela). Deste modo, cada morfema contém em si a informação sobre os contextos em que pode ser inserido. Embick & Noyer (2001) desenvolveram uma análise sobre a marca de definitude no dinamarquês e no sueco, aplicando as assunções da MD, conjuntamente com as operações sintácticas que envolvem não só os movimentos do núcleo de N para DET mas também aspectos relacionados com o caso. Como se sabe, em

línguas como o latim, a classe de palavras que recebe o caso é determinada pelo nome, espalhando-se a morfologia por todos os contextos:

[201] LATIM:  
*Video illam cicadam pulchram.*

Nestas situações, estamos perante morfemas dissociados do mesmo tipo dos que ocorrem no PE, aos quais se opõem outros, que não se espalham pela cadeia sintagmática (os morfemas *singleton*), já que estes últimos aparecem apenas numa única posição, como acontece quer com os morfemas do pretérito perfeito quer com os que se fixam em determinada posição do SN, originando a variação (Costa & Silva, 2006b:44). Este aspecto permite predizer o seguinte: se há variação morfológica, então também é possível existir especificação na variação entre morfemas dissociados e morfemas *singleton*. De facto, reside nestes a primeira fonte de variação que distingue o PB e o PA do PE.

Outra fonte de variação radica, ao que tudo indica, na visibilidade da relação entre especificador e núcleo. Alguns idiomas, como o francês, apresentam concordância entre o pronome objecto e o particípio passado (exemplo [202]), o que não acontece no PE (exemplos [203a] e [203b]):

[202] FRANCÊS:  
*Je les avais ouvertes.*

[203] PE:  
a. *Eu tinha-as aberto.*  
b. *\*Eu tinha-as abertas.*

Ora, este tipo de contraste permite dizer que, quando duas categorias estão na posição especificador/núcleo, a relação paramétrica deixa perceber se a concordância é morfológica visível. No PE e PB/PA, a concordância é visível nas construções com sujeito pré-verbal, (exemplos [194a] e [196]), o que não acontece nas realizações de Almojarife (exemplo [198]) e do Brasil com sujeito pós-verbal, já que, nestas línguas, o sujeito não desencadeia a concordância.

No que concerne à posição da estrutura sintáctica na qual é inserido o morfema de número, é necessário ter em conta que este tipo de marcação é funcional, enquanto a do género é lexical (Müller, 1994a; White *et alii*, 2004; Corrêa, Name & Ferrari-Netto,

2004). Como o número constrói referência, deve aparecer na posição de DET, seja no PE seja no PA ou no PB. Contudo, a concordância do PE é caracterizada por morfemas dissociados, enquanto no PA e PB pode ser determinada por morfemas do tipo *singleton*. Já no que concerne à concordância verbal, quando o sujeito é pré-verbal faz-se apelo à relação SPEC/verbo para justificar a inserção de marcas neste. Logo, é possível predizer que só se espera concordância quando há relação SPEC/núcleo, mas apenas algumas línguas optam pela visibilidade morfológica, como acontece com o árabe (ponto 2.6.2.1 do presente trabalho). Por conseguinte, a ocorrência da concordância no PA e no PB é estabelecida de acordo com as configurações das Figs. 23 (concordância entre sujeito e verbo) e 24 (concordância interna ao SN, exemplo [199a]):

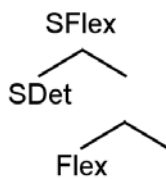


Fig. 24. Sintaxe da concordância sujeito/verbo (PA e PB).

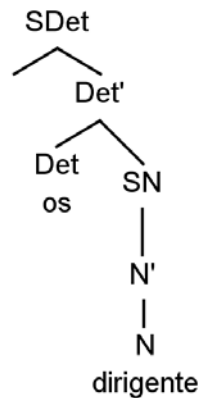


Fig. 25. Sintaxe da concordância interna ao SN (PA e PB).

Relativamente aos possessivos, que no PB evidenciam tendência para receberem a marcação PL quando estão em posição pré-nuclear, mas em segundo lugar da cadeia sintagmática (Scherre, 1988:163), a situação é justificada em termos da relação de definitude que estes itens gramaticais estabelecem com o nome. De facto, ocorre correlação entre posição e definitude, já que o estatuto dos possessivos é diferente quando se encontra em posição pré ou pós nominal. Assim, caso esteja colocado antes do nome (exemplo [204a]), o possessivo é forte portador de definitude e assume a função de núcleo, o que não sucede quando se encontra depois do substantivo (exemplo [204b]), visto deter o estatuto de sintagma nesta posição (Castro & Costa, 2003:108):

- [204] PE:
- a. *O meu livro.*  
↳ núcleo
  - b. *Um livro meu.*  
↳ sintagma

Paralelamente, existem fortes restrições de adjacência entre o definido e o possessivo, já que não admitem a interferência de nenhum elemento entre eles (exemplo [205a]), o que não acontece quando o segundo se encontra em posição pós-nominal (exemplo [205b]):

- [205] PE:
- a. *\*O inteiramente meu livro.*
  - b. *Um livro inteiramente meu.*

Não obstante, essa restrição desaparece no PA e em determinadas regiões do Brasil, visto o definido assumir uma função expletiva, isto é, ser dispensado em grande parte das realizações (exemplo [206]), sem que se perca a noção de definitude:

- [206] PA:  
*arramê mias bagagem, pegué vim* [CELESH3]  
“arrumei a minha bagagem, peguei nela e vim”

Como o número é inserido na categoria que permite a ancoragem com a posição de referência, verificamos que esta é garantida pelo possessivo, que recebe, assim, o morfema de PL. Estabelece-se, então, a configuração sintáctica seguinte:

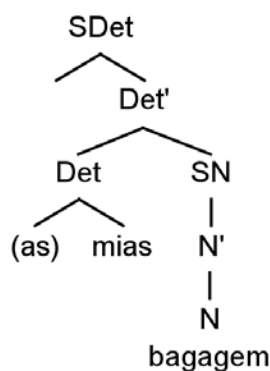


Fig. 26. Sintaxe do possessivo núcleo para a concordância.

Finalmente, a correlação entre posição e definitude é também advogável para os adjectivos que ocorrem em posição pré ou pós-nominal, justificando a assimetria



qualquer dos seus estágios de evolução, já que algumas das línguas analisadas neste estudo (p.e o português e o iorubá ou o santomense) exibem diferenças radicais nos traços de alguns dos seus itens, que vão determinar variação na interlíngua.

## **2.10. Fossilização**

Partindo do pressuposto que a ASL assenta na gramática da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), então o processo é caracterizado por uma tentativa de descoberta das regras (Chomsky, 1969) que, inicialmente, são confrontadas com as da L1. Este aspecto implica que o aprendiz interiorize mais rapidamente as particularidades linguísticas da L2 tidas como mais simples, em detrimento das consideradas complexas. Como o indivíduo em situação de ASL não entende plenamente (ou não tem tempo de assimilar) o sistema de flexões da LA, a sua interlíngua vai registar ausência ou redução destas, situação que se pode traduzir em cristalizações que acabam configurando variação sistemática. Deste modo, enquanto a criança em fase precoce de aquisição inata da L1 produz variação linguística, elaborando erros naturais no processo de desenvolvimento da linguagem, mas que, geralmente, ultrapassa de forma fácil, no caso da ASL, a variação linguística pode manter-se durante bastante tempo e, caso se torne permanente e estável, acaba por configurar uma fossilização. Assim sendo, a fossilização é um fenómeno exclusivamente inerente à ASL e aquisição de LE's, já que não ocorre na aprendizagem da L1 (Selinker, 1972:211).

### **2.10.1. Fossilização e variação**

Os estudos sobre fossilização que buscam revelar as causas que subjazem à mesma têm, quanto a nós, versado muito sobre descrições do fenómeno, mas adiantado pouco acerca dos diferentes tipos e causas que originam as variações sistemáticas ou permanentes. A este propósito, Selinker (1996:226) chama a atenção para o facto de, em diversos trabalhos, fenómenos distintos, como a recaída (*backsliding*),<sup>92</sup> a desaceleração no processo da aprendizagem da L2 ou a baixa proficiência, serem conotados à fossilização, já que não são objecto de uma observação em diacronia. Para Long (2003:490), a estabilização de uma determinada variação, enquanto fenómeno que



encerra todas as características da fossilização, excepto a imutabilidade, representa o primeiro passo no sentido de se prever que a fossilização está em vias de ocorrer. Neste sentido, a estabilização tem particular interesse no estudo da ASL, já que permite entender não só a natureza e causas da fossilização, ela mesma, mas também conhecer quais são as formas particulares da interlíngua que são potencialmente fossilizáveis. E no seguimento deste raciocínio, é ainda possível prever se determinada variação sistemática pode ou não vir a configurar mudança. Deste modo, tanto a variação prolongada como a fossilização que se manifestam na interlíngua impedem que o indivíduo alcance uma competência gramatical que se aproxime da dos falantes da LA, mesmo em contexto de aquisição formal que envolva um estudo intensivo e um processo de memorização bastante consciente. Além de afectar a produção fonológica, em primeira instância, a variação na L2, sobretudo em período precoce da sua aquisição, vai também manifestar-se acentuadamente a nível morfossintáctico, em segunda instância. Em situações de aquisição massiva, se o *input* da LA for prolongadamente esporádico, as motivações empobrecidas, isto é, a instanciação em cadeia de novos erros/parâmetros (Gonçalves, 2004:235), acabam por ter reflexo tanto na variação registada na organização frásica dos constituintes como na produção da flexão e concordância da L2, levando à cristalização diacrónica de formas variáveis.

Para Corder (1973:194), a ASL pressupõe três estágios de aprendizagem: o pré-sistemático, o sistemático e o pós-sistemático. No ponto 5.3.2 do nosso trabalho (variável extralinguística *idade*), é possível verificar que a geração mais idosa de Almojarife apresenta uma *performance* que permite inseri-la no estágio pré-sistemático, já que os falantes apenas revelam uma vaga consciência acerca da ordem sistemática dos itens da LA, não sendo capazes de corrigir erros. Por outro lado, poucos membros da geração menos idosa da comunidade terão alcançado o estágio pós-sistemático da aquisição, uma vez que são raros os indícios de se estar perante discursos consistentes, em que os falantes sejam capazes de corrigir os seus próprios erros autonomamente. Deste modo, o quadro geral da fala de Almojarife configurará o estágio intermédio da aquisição, o sistemático, correspondente igualmente à fase 3 de ASL proposta pela Teoria da “Processabilidade” (Plag, 2008a – Tabela 3.1), dado que a grande maioria dos seus utilizadores, apesar de revelar já alguma interiorização das regras da língua,

continua a produzir variação sistemática, nomeadamente a nível da concordância no SN. Este aspecto indica que os falantes não corrigirão a variação se não captarem ou obtiverem um *feedback* negativo (cf. Corder, 1973:196), o que leva a considerar a possibilidade de a variação registada actualmente no SN do PA se poder vir a fossilizar.

Schumann (1978:34), em estudos sobre pidginização, foi o primeiro a chamar a atenção para a possibilidade de os factores sociais conduzirem a fossilizações. Para o investigador, as situações de fossilização estão intrinsicamente ligadas ao modo como o aprendente se acultura ao grupo detentor da LA (*The Acculturation Model*). Defendendo que os primeiros estados de ASL são comparáveis aos processo de formação das línguas pidgin, o autor entende que, no caso de haver uma grande distância psicocultural entre o aprendente e o grupo falante da LA, a pidginização precoce da L2 torna-se persistente e fossiliza. No quadro de transmissão linguística irregular experienciado pela comunidade de Almojarife, o distanciamento em relação à LA foi bastante marcante em determinado momento da emersão e nivelamento da L2, fazendo com esta preenchesse apenas requisitos de comunicação interétnica. Assim sendo, o contexto criado mostrou-se propício para a fixação de formas transferidas das L1's, candidatando-as à fossilização. Ainda assim, é preciso ter em conta que a proposta de Schuman (1978) não leva em conta dois aspectos fundamentais: (i) os factores que determinam a aculturação do indivíduo não são estáticos, mas variáveis e dinâmicos, consoante as experiências que o aprendente vai atravessando; (ii) os aprendentes não estão apenas submetidos às condições sociais, já que se podem tornar sujeitos destas e construir o próprio contexto de aprendizagem.

Para Hale (1988:32), a fossilização resulta do facto de o falante não conseguir erradicar determinados parâmetros da L1, que acabam intervindo na ASL.<sup>93</sup> De certa maneira, o pressuposto seria posteriormente retomado por Schwartz & Sprouse (1996), quando estes determinaram que a GU actua na L2 através da L1. Como tal, a fossilização resulta da dificuldade que o aprendente tem em refixar os parâmetros na L2, conforme postulado pela FFFH. Por seu lado, Selinker & Lakshmanan (1993:215) entendem que a fossilização ocorre quando determinados factores da aquisição de uma L2 actuam de forma múltipla (*Multiple Effects Principle* – MEP), ou seja, funcionam em continuidade, levando à estabilização de certas formas. Entre os ditos factores podem

apontar-se a simetria de estruturas, a afectividade, o princípio da distância mínima, a interferência e, sobretudo a transferência, tida como o factor central da ASL para possibilitar a fossilização. A conjugação destes factores levaram a que acontecesse a transmissão em cadeia de novos/errados parâmetros (cf. Gonçalves, 2004:235) durante a aquisição do PAL1, pelo que terá de ser considerada a hipótese de ter ocorrido reinstanciação paramétrica neste, possibilitando a fossilização de determinadas formas do substrato na fala dos almoxarifanos.

Ainda segundo Selinker & Lakshmanan (1993:215), a fossilização apresenta duas formas: (i) a fraca, que determina que a transferência de uma língua é um co-factor privilegiado na ocorrência de efeitos múltiplos; (ii) a forte, que postula que a transferência de uma língua é um co-factor necessário na ocorrência de efeitos múltiplos. Apesar de estes dois tipos de factores serem fundamentais para ocorrer a fossilização, os autores admitem ainda que esta pode igualmente acontecer sem transferência, conforme o haviam demonstrado Bean & Gergen (1990:219) em estudos sobre variação registada na *performance* individual de interlínguas com fossilizações.

### **2.10.2. Potenciais causas da fossilização**

Várias têm sido as causas apontadas como potenciais fontes da fossilização, de entre elas podendo destacar-se as que resultam de fenómenos de esquiva (*avoidance*)<sup>94</sup> intencional por parte do falante (Nakuma, 1988), a ultrageneralização de determinadas regras Ellis (1994)<sup>95</sup> ou da falha na interacção entre o conhecimento interior e os estímulos externos para determinar o desenvolvimento das gramáticas mentais (Franceschina, 2005).<sup>96</sup> No caso da aquisição do PtgL2, são comuns, por exemplo, os erros de ultrageneralização, e consequentes fossilizações, que actuam sobre o sistema verbal, com os falantes regulando o paradigma de conjugação dos verbos da segunda (terminados em *-er* no infinitivo) e terceira classes (terminados em *-ir* no infinitivo) de acordo com o paradigma dos verbos de primeira classe (terminados em *-ar* no infinitivo) ou aplicando, continuamente, a forma da terceira pessoa singular a todas as outras:

[209] PA:  
*nós é pobre miserável aqui Praia Moxarife* [FRANCM2]

Neste caso, estamos perante transferências intralinguísticas, que fazem com que o falante ignore sistematicamente as regras das restrições, incluindo as que incidem sobre itens lexicais ou de qualquer outro tipo linguístico, fossilizando-as em diacronia (Ellis, 1994:284). Segundo Brown (1994:367), os falantes que incorrem na sobregeneralização acreditam que estão a usar a gramática correcta, em virtude de terem interiorizado regras sistemáticas. Deste modo, o fenómeno da sobregeneralização ocorre de forma inconsciente, necessitando quase sempre de *feedback* negativo para não ser fossilizado.

Tendo ainda em conta as distintas teorias sobre fossilização, poder-se-á dizer que, genericamente, o conceito aponta para as seguintes determinações: (i) a fossilização pode aparecer em determinados níveis da interlíngua; (ii) a fossilização pode ocorrer não só em diferentes estádios de aprendizagem da LA mas também em variados estratos etários; (iii) a fossilização pode acontecer a nível estrutural ou a nível da competência; (iv) a fossilização manifesta-se, geralmente, através de formas que se desviam das normas da LA; (v) há graus suaves e fortes de fossilização. Do exposto, conclui-se que a fossilização representa um estádio inevitável na ASL e com influências nesta.

O conceito de transferência tem sido observado como uma questão central quer em estudos de ASL (Chomsky, 1969) quer em estudos sobre fossilização (Ellis, 1994), tendo o primeiro autor levantado a possibilidade de ser um processo consciente, através do qual o aprendente constrói a interlíngua, testando hipóteses sobre a L2. Nesta perspectiva, e tendo Selinker (1972:229) advogado que algumas das regras registadas na interlíngua resultam de transferências da L1, então esta constitui a base para a aprendizagem da nova língua (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006). Desta forma, os erros registados na L2 têm origem nas diferenças entre gramáticas da LA e L1, sendo as transferências as causas da fossilização.<sup>97</sup> No entanto, e de acordo com a HI (Plag, 2008a, 2008b, 2008c, 2008d), é importante estabelecer a diferença entre traços conectados a questões universais do desenvolvimento da interlíngua e os que configuram transferências, já que algumas formas registadas na interlíngua poderão não passar de falsas transferências (Tabela 6.1). De facto, para estas acontecerem é necessário que o processador da interlíngua tenha disponível um dispositivo que accione o processamento da estrutura da L1 passível de transferência, isto é, que a estrutura da L1 a transferir encontre condições para ser

processada no sistema de desenvolvimento da L2<sup>98</sup> (Andersen, 1983c:182; Plag, 2008b:314).

Alguns estudos sobre ASL apontam a intervenção de duas outras causas para ocorrer fossilização de formas incorrectas na interlíngua: a falta de instrução formal na LA e, conseqüentemente, de *feedback* negativo sobre a variação produzida (Vigil & Oller, 1976:287; Corder, 1981b:72); a idade em que se dá a ASL (Brown, 1980; Larsen-Freeman & Long, 1991; Selinker & Lakshamanan, 1993; Lightbown & Spada, (2003 [1993])). Sobre o primeiro aspecto, pronunciar-nos-emos no ponto 5.3.1, dedicado à análise da variável social *escolaridade*. Quanto à questão da idade, Selinker & Lakshamanan (1993:207) estabelecem uma associação entre fossilização e escalão etário adulto, reservando a noção de desenvolvimento para a idade infantil. Por seu lado, Lightbown & Spada (2003 [1993]:70) entendem que a aprendizagem de uma L2 por adultos se revela mais eficaz nos estágios iniciais de aquisição e que estes podem ser capazes de comunicar com sucesso, apesar de apresentarem diferenças a nível de sotaque e produzirem formas que, por vezes, não se adequam ao discurso. Estes aspectos levaram ainda os autores a sugerir que, depois do período crítico, a aquisição não será inata, demarcando-se do modo como a mesma acontece na infância, seja em L1 seja em L2.

### **2.10.3. GU e fossilização**

Para White (1989, 1990, 1993, 2000, 2003), defensora de que a GU actua de forma igual na aquisição de L1 e L2, o falante em estado inicial de ASL já possui a gramática da L1, com princípios e parâmetros estabelecidos. Como o aprendente já conhece um sistema linguístico, vai ocorrer transferência do mesmo para aquele que se encontra em estágio inicial de aquisição, já que a gramática a que o falante recorre para testar as hipóteses é a da L1, que vai actuar como um filtro (Krashen, 1982, 1985, 1988), impedindo que certas formas do *input* da LA sejam captadas e apreendidas. Desta forma, os aprendentes são impedidos de atingir, na gramática da L2, a mesma proficiência que têm na gramática da L1, pelo que, a não percepção de estruturas relevantes da LA conduz à fossilização. Por seu lado, a fixação de determinados parâmetros, em

detrimento de outros, deve-se à variação inter e intrapessoal, o que caracteriza a interlíngua do aprendente (Cook, 1993:186).

Para Flynn (1989:104), a ASL é facilitada quando ocorre um ajustamento entre os parâmetros da LA e da L1. Para que tal aconteça, os aprendentes em ASL necessitam de ser expostos a *input* prolongado e correcto, caso contrário os parâmetros da L2 não se ajustam aos já estabelecidos na L1, instanciando-se novos/errados parâmetros na interlíngua (Gonçalves, 2004:235). Assim sendo, os falantes em processo de aquisição precisam de estabelecer novos valores para os parâmetros a serem adquiridos, o que compromete a apreensão e, no caso de esta se retardar, induz em fossilização. Contudo, para se determinar se uma forma linguística configura fossilização, Selinker (1972:223) propõe que ela terá que permanecer estabilizada pelo menos durante cinco anos. Por seu lado, Han (2004:14) questiona o período, uma vez que, primeiro, seria necessário precisar o tempo que um determinado indivíduo necessita para adquirir um factor particular na L2. De facto, segundo a autora, múltiplos são os factores que devem ser considerados como motivadores da fossilização, estando alguns deles ainda por determinar nos estudos efectuados até à data.

De tudo o que foi exposto acerca das possíveis causas motivadoras da fossilização, conclui-se que alguns factores se mostram determinantes para que certas formas linguísticas fossilizem, como é o caso do meio ambiente no qual o falante em aquisição se encontra inserido ou as questões relacionadas com aspectos cognitivos, neurobiológicos e sócio-afectivos. E olhando individualmente para estes, enquanto elementos de um todo, salta à evidência que apenas dois tipos de factores concorrem para que se dê a fossilização: os factores internos e os factores externos. Estes aspectos permitem então elaborar a seguinte tabela taxionómica, para, de forma sumariada, não só se simplificar toda a explanação acerca das fossilizações mas servir também como proposta metodológica para o nosso trabalho, quando tivermos de abordar questões relacionadas com a problemática da fossilização:

**Tabela 4.1.** Taxionomia dos factores motivadores da fossilização (Xueping, 2008:131).

Externo	Meio ambiente		p.e.: ausência de instrução; défice de <i>input</i>
Interno	Cognitivo	Representação do conhecimento	p.e.: influência da L1; défice de acesso à GU
		Processamento do conhecimento	p.e.: défice de atenção
		Psicológico	p.e.: esquiva; simplificação
	Neuro-biológico		p.e.: idade; défice de talento
	Sócio-afectivo		p.e.: défice de aculturação

Outro aspecto que consideraremos tem a ver com o facto de o conceito de estabilização estar muito próximo do de fossilização, mas ambos os fenómenos configurarem realidades distintas. A este propósito, Han (1998:13) e Selinker & Han (2001:278) esclarecem que a estabilização e a fossilização configuram um *continuum*, no qual a primeira é parte integrante do processo da segunda e, por tal, não podem ser equiparadas. Segundo a perspectiva de Long (2003:521), a estabilização é o primeiro sinal de que poderá ocorrer uma fossilização. Contudo, os fenómenos distinguem-se em virtude de a primeira ser permeável e reversível, enquanto a segunda é hermética e permanente. Portanto, a forma estabilizada pode ser corrigida, mas a fossilização é imutável.

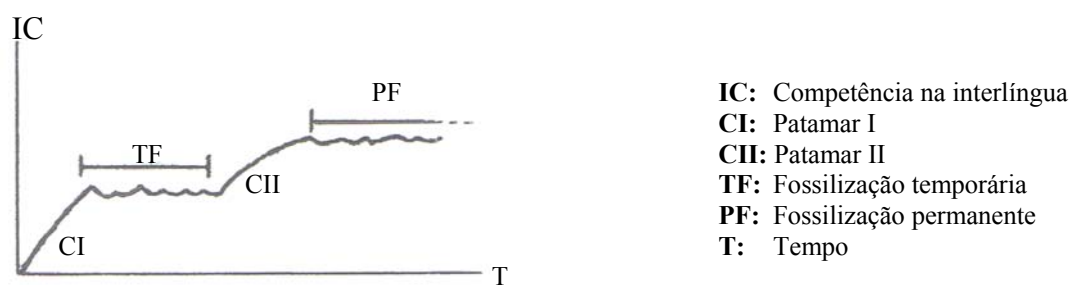
#### **2.10.4. Fossilização no PA**

##### **2.10.4.1. PA: transferência e fossilização fonológica**

A interlíngua dos falantes em ASL pode registar dois tipos de fossilização: (i) fossilização individual, determinada pela persistência na elaboração de certa forma durante o desenvolvimento da interlíngua do indivíduo; (ii) fossilização de grupo, motivada pelo nível de uso de uma forma durante o desenvolvimento diacrónico da língua de determinada comunidade. As segundas têm como ponto de partida alguns aspectos relacionados com as primeiras, uma vez que são originadas pela competência linguística fossilizada, primeiramente, ao nível de desenvolvimento das competências fonológica, gramatical, lexical e pragmática de indivíduos aprendentes de uma L2. Como tal, revelam-se na interlíngua de falantes de L2's que têm um contacto, de certa

forma prolongado, com a LA, atingindo um determinado nível de *performance* sem, contudo, deixarem de exhibir, repetidamente, certas variações, difundindo-as, depois, pelo grupo em que estão inseridos (Selinker, 1979:212).

Não obstante, note-se que as fossilizações poderão ser temporárias ou permanentes (Selinker, 1979:218) e ocorrerem a todos os níveis, desde o fonológico ao pragmático, sem deixar de passar pelos morfológico, sintáctico e semântico. As fossilizações temporárias, também designadas por estabilizações, revelam que a interlíngua fossilizada consiste de patamares de aprendizagem (Fig. 27), nos quais o desenvolvimento de determinados traços da LA é fixado ou inibido por períodos de tempo mais ou menos longos (Sims, 1989:70).<sup>99</sup> Como tal, as fossilizações deste tipo podem ser desfossilizadas, o que não acontece com as fossilizações permanentes, que resultam de uma conjugação de fenómenos conotados a aspectos sociais, psicológicos e interactivos e impedem que o estágio final de determinada forma da interlíngua alcance o mesmo estágio patente na LA:



**Fig. 27.** Fossilização temporária e fossilização permanente (Xueping, 2008:131).

No presente trabalho abster-nos-emos de abordar todo o género de fossilizações, já que as variáveis linguísticas a que recorremos para estudo do fenómeno da CPL-var do PA são do tipo marcadamente fonético-fonológico e morfossintáctico. Assim, incidiremos a nossa atenção sobre as fossilizações directamente relacionadas com estas variáveis.

O “fenómeno de Joseph Conrad”, ou seja, a fossilização fonológica, é tido como inevitável na ASL após a puberdade (Scovel, 1969:252).<sup>100</sup> As diferenças fonológicas entre interlíngua e LA resultam da aquisição incorrecta da pronúncia de algumas formas desta, que são afectadas pelo sistema fonético da L1, constituindo, possivelmente, a maior diferença entre ambas. Como determinadas línguas do grupo níger-congo



atlântico, caso do ramo bantu, não possuem a consoante /r/, é difícil aos falantes adultos destas articularem-na correctamente quando em processo de aquisição do PtgL2. Observando os exemplos [210] e [211], em especial o nome “doutor”, constatamos que a ausência da consoante /r/ nas línguas bantu provoca uma dificuldade de articulação nos falantes adultos em aquisição do PtgL2, levando à realização sistemática de fenómenos de apócope e lambdacismo na interlíngua (o santomense), que se conservaram diacronicamente, configurando fossilizações no PA, possivelmente temporárias (exemplo [212]):

[210] FORRO:  
*dotolo da e zesõ*  
(Ferraz, 1979:66)  
“o doutor deu-lhe injeções”

[211] PA:  
*dotôlo deu lemede* [PRINCEM3]  
“o doutor deu um remédio”

[212] PA:  
*tenho bastante filho, outra filha professora, outra veio **doutora***  
[MAURIH2]

Para Tarone (1978, 1995), à fossilização fonológica subjazem causas fisiológicas e psicológicas. Deste modo, as dificuldades articulatórias poderão assentar no facto de as línguas bantu e o português recorrerem a distintos pontos articulatórios do tracto bucal para produzirem os seus sons, levando os falantes do PtgL2 a enfrentarem dificuldades na articulação de sons que não existem nas suas L1’s (cf. Ellis & Beaton, 1995:161). Por seu lado, as justificações em termos psicológicos apontam para o fenómeno da formação de hábitos, que leva a percepção e a produção da fala a serem permanentemente influenciadas pelo sistema fonológico da L1, a ponto de o aprendente enfrentar sérias dificuldades em perceber ou elaborar uma nova fonologia (Tarone, 1978, 1995). Este aspecto pode gerar uma falta de empatia entre o aprendente e os falantes da LA, fazendo com que os primeiros só adoptem alguns padrões de pronúncia quando se sentem particularmente identificados com determinado grupo (Labov, 1972a:31). Debruçando-se ainda sobre a estrutura silábica de certas línguas (incluindo o português), Tarone (1978, 1995) defende também que o fenómeno da fossilização pode assentar em três

factores: a transferência, os processos relativos à aquisição da L1 e os fenómenos de carácter universal. A hipótese da transferência fonológica postula que o aprendente usa a estrutura silábica da L1 para tentar comunicar na L2, um aspecto que é facilmente constatado no santomense e que revela resquícios no PA, conforme analisado em detalhe no ponto 5.2.3.10.1 do presente trabalho. Assim, se a L1 do falante contém sílabas de acordo com a estrutura CV, a tendência do aprendente é transformar a estrutura silábica da LA em estruturas do mesmo tipo das que ocorrem na L1, até porque os falantes dão primazia às estruturas do tipo CV em virtude da sua força lexical (Clements, 2009:22). Neste caso, e como referimos no ponto 2.4.3, a saliência perceptual impõe-se à frequência na LA, em virtude de a estrutura CV ser mais facilmente processável (Thomason & Kaufman, 1988:331).

A hipótese sobre os processos relativos à aquisição da L1 entende que o aprendente da L2 tende a fazer aquilo que o aprendente da L1 faz com a estrutura silábica. Deste modo, os defensores desta linha, assentando pressupostos nos princípios advogados pela perspectiva psicolinguística de interlíngua (Selinker, 1972) de que a GU deixou de estar disponível para actuar na aquisição da L2 (Clahsen & Muysken, 1986; Meisel, 1997; Muysken, 2001), postulam que os processos que actuam na interlíngua são independentes daqueles que incidem sobre a L1 e o que ambos os falantes fazem é apenas simplificarem as sílabas difíceis (Oller, 1974:114). Porém, Tarone (1978, 1995) discorda desta hipótese, defendendo que os falantes em ASL usam estratégias para pronunciarem os sons difíceis. Assim, no caso de ocorrerem encontros consonantais, em vez de proceder a uma simplificação, o falante da L2 insere vogais na estrutura consonantal (epênteses ou paragoges), a fim de retomar a estrutura universal CV (Tarone, 1978:328). Portanto, estes fenómenos de inserção configuram estratégias próprias da fonologia dos aprendentes em ASL, levando a que a estrutura silábica da interlíngua seja diferente daquela que caracteriza a LA. Casos de inserção epentética ou paragógica são comuns no forro e no PA, como se pode também constatar no ponto 5.2.3.10.1 do nosso trabalho.

As fossilizações, em particular as fonológicas, têm, então, sido explicadas em termos de factores diversos: idade, resistência à aculturação, pressão comunicativa, ausência de aprendizagem formal, natureza do *input* (negativo ou positivo), perda de capacidade

discriminatória a nível dos padrões trocaicos e jâmbicos em idade pós-puberdade, falta de empatia em relação aos falantes da LA, formação psicológica de hábito, qualidade de voz para expressar emoções,<sup>101</sup> etc.. Contudo, e ao que tudo parece indicar, residirá na transferência a responsabilidade maior pelas fossilizações.

A fossilização fonológica é definitiva na maioria dos casos, permanecendo em indivíduos que adquirem um grande domínio de vocabulário e sintaxe da LA (Scovel, 1969:252). Se a distância entre os sistemas fonológicos da L1 e da LA for acentuada, maior será a tendência para os aprendentes da L2 construírem os sons que não existem na L1, por paronímia fónica com aqueles que fazem parte desta (exemplos [160], [162] e [163]). A inabilidade dos falantes da L2 em pronunciarem correctamente os sons da LA conduz, assim, à fossilização fonológica (Levelt, 1989:286). Por outro lado, como o conhecimento da L1 é o ponto de partida para a aquisição da L2 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), a produção fonológica desta vai seguir o mesmo processo de articulação fónica da L1. Deste modo, só dificilmente se poderá desassociar a fossilização fonológica do fenómeno da transferência, sendo a sua cristalização determinada pelo facto de os falantes não serem capazes de refixar, na L2, os parâmetros desactivados das matrizes fonéticas da LA (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003) nem conseguirem erradicar os novos/errados parâmetros transferidos das L1's (Hale, 1988:32; Gonçalves, 2004:235). Neste caso, os substratos do grupo níger-congo atlântico são a fonte de determinadas formas que transitaram para a interlíngua (santomense) e que acabaram fossilizadas no PA, devido à reinstanciação em cadeia de novos/errados parâmetros (cf. Gonçalves, 2004:235).

#### **2.10.4.2. PA: transferência e fossilização morfossintáctica**

As fossilizações morfológicas ocorrem em maior grau quando a LA possui uma gramática elaborada em morfemas, por oposição a uma L1 cuja gramática detém um reduzido sistema flexional. Isto é exactamente o que sucede quando comparamos as gramáticas do PE, por um lado, e do santomense e seus substratos do grupo níger-congo atlântico, por outro lado. O sistema de artigos portugueses e a flexão sufixal em género e número não encontram correspondência nestes dialectos. Como tal, podem não ser

fonologicamente interpretados pelos falantes em processo de aquisição do PtgL2, originando a sua ausência na interlíngua (Baxter, 2004:120; Kihm, 2003:347), como acontece a nível da estrutura do SN, que não exhibe determinantes no santomense e apresenta variação dos mesmos no PA, sendo esta originada pela conjugação de fenómenos de transferência e empréstimo.

Por outro lado, todas as línguas têm as suas próprias regras sintáticas. Uma das manifestações típicas de fossilização sintática registada em falantes do PtgL2 que têm línguas do grupo níger-congo atlântico como L1's prende-se com o uso do tempo e modo verbais, já para não falar no sistema de marcação do tempo pretérito e do participio passado. Como o sistema verbal das suas L1's não detém idênticas características, os falantes não conseguem fazer uma distinção clara das diferentes formas da LA, e acabam por recorrer às suas formas cognitivas para marcarem o tempo verbal, originando-se, assim, as fossilizações.

Alguns autores (Pienemann & Johnson; 1987; Pienemann, 2000; Plag, 2008a, 2008b, 2008c, 2008d) opinam que a ASL atravessa seis estágios incontornáveis de desenvolvimento da interlíngua (Tabela 3.1), iniciando-se com um desenvolvimento gramatical nulo, que leva à articulação de palavras simples, as quais, por sua vez, podem também representar fórmulas, tal como acontece na fase em que as crianças em aquisição de L1 produzem frases holofrásticas. No estágio seguinte a interlíngua começa a apresentar frases simples, mas completas, se bem que o falante ainda não faça a distinção entre os elementos que as compõem. Portanto, as palavras são ordenadas em função dos seus significados ou focos de informação nelas contidos. Tal como acontece com a produção das crianças em processo de aquisição de uma L1 na fase pós-holofrástica, as palavras, articuladas em função de relações semânticas, não apresentam marcas sintáticas e morfológicas nem flexão contextual. Plag (2008b:314) refere que a maioria das línguas crioulas não ultrapassa este estágio de desenvolvimento, podendo algumas revelar ligeira transição para o terceiro estágio (apresentam indícios primários de flexão contextual), mas sem conseguirem ultrapassá-lo. No caso do santomense, um crioulo do tipo basilectal, é evidente a estagnação de alguns dos seus falantes no segundo patamar do desenvolvimento da interlíngua, dado evidenciarem escassas produções de flexão inerente e quase total ausência de flexão contextual. Ainda assim,

os registos do nosso *corpus* revelam que a maioria dos falantes de Almojarife realiza flexão contextual em maior ou menor grau. Portanto, o terceiro estágio é aquele que melhor caracteriza a produção da generalidade da comunidade (Tabela 3.1), já que se nota nos seus membros a capacidade de conseguirem identificar o início e o fim da construção frásica, a ponto de trocarem a ordem dos elementos que podem ocorrer nestas duas posições (exemplo [213]) ou estabelecerem concordância, embora variável, a nível dos elementos do SN.

[213] PA:  
*lá pas 14 que nós conseguimos trazê isso pa casa* [OSVALH1]

Paralelamente, a flexão contextual registada em itens incorporados na estrutura do SN denota que a aquisição da pré-sintaxe foi alcançada. Não obstante, e apesar de alguns falantes revelarem algum domínio deste tipo de flexão e aplicarem as estruturas básicas da LA, nenhum deles evidencia uma aquisição completa das regras gramaticais que caracterizam o quarto estágio de desenvolvimento da interlíngua em direcção à LA. Como tal, são incapazes de, por exemplo, ultrapassar na totalidade a variação a nível da estrutura do SN ou caracterizar elementos particulares que compõem a cadeia frásica, movendo-os do interior desta para a posição inicial ou final.

Como a aquisição dos estágios é fixa, não é possível aos falantes abolir qualquer das etapas referidas. Este aspecto pressupõe que a capacidade de realizar as tarefas de um determinado estágio implica a competência de efectuar as do estágio anterior. Ora, é neste ponto que reside a diferença entre as aquisições de L1 e de L2, já que os falantes desta, em determinada fase da sua ASL, produzem construções típicas de diferentes estágios, comprovando-se que fossilizaram, suave ou definitivamente, formas de fases anteriores. Este aspecto evidencia ainda que a fossilização acontece por etapas (Fig. 27), já que, na aquisição de L1, se alguma destas for superada, o falante não apresenta variação característica dessa mesma etapa no patamar seguinte de aquisição, o que não sucede no PA, no qual o desenquadramento distribucional entre flexão inerente e flexão contextual transita de estágio para estágio.

Contudo, deve ser sempre levado em linha de conta que a aquisição dos diferentes estágios varia de indivíduo para indivíduo (Fernandes-Boëchat, 1989:212), comprovando-se que ela assenta em factores diversos, que não apenas os

morfofossintáticos. Como tal, as diferenças individuais também actuam a nível da fossilização, com alguns aprendentes a fossilizarem formas de um determinado estágio e outros a fossilizarem formas de um patamar distinto. Assim sendo, a observação das fossilizações pressupõe que a busca das suas motivações tenha que ser vista não só levando em conta as condições internas e externas que condicionam a ASL mas também considerando o tempo durante o qual determinada variação permanece estabilizada.

Seguidamente, passaremos a tecer considerações em torno da aquisição massiva de L2's, a fim de se observar até que ponto a variação, que começa por ser individual, se instala a nível da comunidade, determinando alterações no padrão da fala desta.

### **2.11. Aquisição massiva de L2's: línguas crioulas e teorias de aquisição de L2's**

Pesquisas recentes em ASL têm permitido uma nova visão acerca das propriedades das línguas crioulas e natureza da crioulização (Kouwenberg & Patrick, 2003; Lefebvre, White & Jourdan, 2006a; Siegel, 2008). Com base nestes estudos e nos pressupostos da Teoria da “Processabilidade” (Pienemann, 1998, 2005b), Plag (2008a, 2008b) formulou a Hipótese da Interlíngua (*Interlanguage Hypothesis* – HI), que advoga que os crioulos representam interlínguas convencionais em estágios primários de aquisição. A HI tem respaldo nos estudos de Andersen (1980, 1983a) e observa os processos de ASL que ocorrem nos estágios iniciais do contacto (cf. Kouwenberg & Patrick, 2003), a fim de avançar pressupostos acerca da forma como se processa a flexão morfofossintáctica dos crioulos.

#### **2.11.1. Universais linguísticos vs. transferência: a Hipótese da Interlíngua (*Interlanguage Hypothesis*)**

Dado que podem ocorrer transferências sem ASL (p.e. situações de diglossia ou aquisição de L1 em contexto bilingue) e processarem-se, durante a crioulização, estruturas que não resultam de transferência, a HI resgata o pressuposto de que os primeiros estados de ASL são comparáveis aos processos de formação das línguas pidgins (Schumann, 1978). Neste aspecto, os crioulos configuram interlínguas convencionais em estágios primários de aquisição, sendo os patamares avançados da aquisição explicados em termos de factores da língua-E, como, por exemplo, o tipo de acesso ao superstrato. Assim, estes factores terão pouco a ver com a simplicidade

gramatical dos crioulos, que será uma consequência da fase inicial de emersão das línguas nascidas por contacto.

Embora os crioulos apresentem morfologia reduzida, alguns deles possuem flexões a nível nominal (ponto 1.2.2.2), verbal e adjectival. Estas verificam-se unicamente a nível dos morfemas presos que expressam categorias gramaticais, uma vez que os morfemas gramaticais livres, como os marcadores de TMA, não configuram “morfologia flexional” no sentido clássico. Exemplos de ambos os tipos de morfologias podem ser consultados em alguns trabalhos sobre crioulos (p.e. Hudson, 1983; Holm, 1988/1989; Stolz, 1989; Kouwenberg, 1994, Bakker, 2003; Baptista, 2003; Veenstra, 2003). Luís (2008) providencia também algumas evidências de sufixos que englobam noções de tempo e aspecto em conjugações verbais de alguns crioulos indo-portugueses.

#### **2.11.1.1. Teoria da “Processabilidade” e flexão morfológica das línguas emergentes por contacto**

Os processos de flexão, independentemente de serem inerentes, contextuais ou morfofonémicos, fazem uso dos seguintes elementos formais: afixação, clitização, reduplicação, metafonía e supletismo. Os pidgins e crioulos exibem ausência de flexão quase total, em virtude de apresentarem um desenquadramento distribucional entre flexão inerente e flexão contextual. A distinção entre flexão contextual e flexão inerente dá-se a nível da informação que é necessária ou não para tornar a comunicação eficiente. A aplicação da concordância ou de casos estruturais envolve troca de informação intra e interfrásica, ao passo que a flexão inerente não requer troca de informação entre constituintes distintos. A concordância sujeito-verbo e o uso do caso sujeito, por exemplo, requerem um processamento-S, enquanto a atribuição dos casos objecto ou genitivo exigem troca de informação frásica entre núcleo e complemento (Fig. 15). Por seu lado, a concordância de número e/ou de género obrigam à troca de informação intrafrásica.

Aplicando estes processamentos aos crioulos, constata-se que eles possuem, quase exclusivamente, casos típicos de flexão inerente (p.e. pluralização de nomes e/ou flexão verbal de tempo e aspecto), ou seja, estruturas que praticamente não requerem troca de informação entre os seus constituintes, revelando tendência para a *interface* léxico-sintaxe mínima e transparente (Kihm, 2003).<sup>102</sup> Assim, os casos de flexão contextual são

raríssimos, enquanto situações de flexão morfofonémica não acontecem de todo (Crowley, 2008; Plag, 2008a). Partindo ainda do pressuposto que os crioulos configuram casos típicos de ASL, então é lógico predizer que estamos perante interlínguas, conforme o confirma a questão da morfologia flexional. Em adição, também se verifica que o valor semântico dos afixos flexionais não é, por si só, um garante da sua sobrevivência geracional (Plag, 2008a:125-126), como o demonstra, por exemplo, o facto de os afixos marcadamente inerentes de tempo/aspecto dos crioulos indo-portugueses precederem a aplicação de flexão contextual, como acontece com a concordância ou a atribuição de caso. Assim sendo, a preservação de marcadores flexionais aponta mais para uma questão de processamento do que para uma questão de transparência semântica ou relevância comunicativa. E nesta conformidade, poder-se-á afirmar que os falantes responsáveis pela emergência dos crioulos recorrem exactamente aos mesmos processos mentais dos outros falantes em situação de ASL. Estes aspectos levam então a concluir que a tipologia da flexão crioula emerge naturalmente, ou seja, em consequência de constrangimentos universais que actuam quer sobre o processamento linguístico quer sobre a aquisição de línguas, exibindo, em sequência de tal, os estágios típicos de desenvolvimento das interlínguas.

O facto de os falantes dos crioulos estabilizarem as suas interlínguas em estágios primários do desenvolvimento (Plag, 2008a:125-126) tem a ver com dois aspectos primordiais: a motivação e o *input*. Na grande maioria dos casos de crioulição, o acesso ao superstrato era bastante limitado, pelo que não havia exposição à LA que permitisse um desenvolvimento da interlíngua para estágios mais avançados de ASL. De facto, é tido como consensual que todo e qualquer conhecimento linguístico é indissociável das capacidades cognitivas do aprendente, pelo que a ASL representa uma construção de sistemas do conhecimento, direccionada para a compreensão e uso. Dado que a capacidade de retenção do conhecimento é limitada, o falante em situação de aprendizagem necessita da experiência e da prática rotineira para ir, progressiva e automaticamente, utilizando os conhecimentos que vai interiorizando (McLaughlin, 1987; McLaughlin & Heredia, 1996; Clements, 2009). Simultaneamente, as condições sociohistóricas em que ocorriam as nativizações das línguas fazem pressupor que os responsáveis pela crioulição não se sentiriam profundamente motivados para



aprenderem na perfeição as línguas dos seus opressores, antes olhariam para elas como meros veículos que serviam a comunicação interétnica (Hymes, 1971a; Silverstein, 1972; Baker, 1994; Baker, 2001; Smith, 2006). Neste aspecto, a hipótese da filtragem afectiva (*Affective Filter Hypothesis*), proposta por Krashen (1982, 1985, 1988) não deixa de ser aqui relevante, uma vez que advoga que, caso as condições psicológicas do aprendente estejam afectadas por aspectos subjacentes ao afecto (motivação, atitude, autoconfiança e ansiedade), ele revela tendência para filtrar o *input* e não processar a aquisição. Assim, se por um lado este tipo de motivação determinou a aquisição de material lexical e algumas propriedades estruturais do superstrato, por outro lado dispensou a aquisição de formas complexas e mais avançadas, como acontece com a atribuição de casos ou a aplicação da concordância morfológica. Portanto, é de esperar que, mesmo em estágios mais avançados da criouliização, se continuem a evidenciar manifestações características de estágios primários de ASL. Este aspecto permite ainda prever que o melhor acesso ao superstrato, a elevada motivação para aprender a LA ou o prolongado contacto com o superstrato, em situações que diluem o afastamento linguístico relativamente, conduzem à aplicação de estruturas crioulas mais próximas dos estágios avançados das interlínguas (Plag, 2008a:129).

Em virtude da variabilidade do desenvolvimento das L2's, é possível observar diferentes graus de traços flexionais em diferentes crioulos, dependendo do tipo de línguas envolvidas no contacto e das situações sociohistóricas em que este acontece. Caso as línguas em contacto sejam morfológicamente elaboradas e apresentem traços próximos, então as hipóteses de sobrevivência da flexão aumentam consideravelmente (Bakker, 2003; Clements, 2009). Esta particularidade pode ser interpretada como transferência em termos de Teoria da "Processabilidade", uma vez que algumas estruturas da L1 poderiam processar-se na LA e, como tal, transitam para a interlíngua, originando padrões não-uniformes na língua utilizada por grupos de aprendentes detentores de L1's tipologicamente distintas (Pienemann & Håkansson, 2007:492). O pressuposto de que os crioulos, enquanto produtos emergentes a partir de pidgins radicalmente reduzidos, evitam os traços desnecessários para a comunicação (McWhorter, 2006:180) e emergem como línguas mais simples do que as línguas não-crioulas (McWhorter, 2000:93), tem sido criticado em vários domínios pela forma

obscura e vaga como foi avançado (Siegel, 2004a; Klein, 2006a, 2006b; Kihm, 2008; Plag, 2008a). Contrariamente, a HI defende que a não-aquisição de categorias flexionais é uma consequência previsível dos constrangimentos que actuam sobre o processamento da língua nas fases primárias da ASL. Assim, a ausência de flexão nestas fases acontece porque, simplesmente, os falantes das interlínguas não detêm os recursos para o processamento da flexão. Quando o desenvolvimento desta se inicia, o ponto de partida é a flexão inerente, que não requer processamentos complicados. Posteriormente, nos estágios mais avançados de desenvolvimento da interlíngua, flexões de outro tipo podem ocorrer por influência de factores externos, especialmente por indução da LA em situações de interacção linguístico-cultural estreita (Winford, 2003a:313).

#### **2.11.1.2. Teoria da “Processabilidade” e processamento sintáctico das línguas emergentes por contacto**

No que concerne aos estudos sobre crioulos, ainda permanecem relativamente obscuros quais os mecanismos responsáveis pela natureza não-marcada de algumas das suas estruturas linguísticas. Outro dos desafios em análises crioulisticas prende-se com a tentativa de se clarificar, com exactidão, se determinada estrutura resulta de uma transferência do substrato ou se, pelo contrário, é regida por tendências universais. Aliás, em algumas situações parece ocorrer mesmo uma convergências de ambos os fenómenos, pelo que a HI pode ajudar a clarificar se as similaridades entre interlínguas e crioulos, por um lado, e entre desenvolvimento da interlíngua e pidginização/crioulização, por outro lado, não são meramente acidentais. De facto, os paralelismos entre a formação de crioulos e a ASL dão-se a nível do desenvolvimento de ambos, da natureza da variação registada nas interlínguas e nos crioulos, do papel da interacção e do factor idade.

Um fenómeno que estabelece paralelismo evidente entre ASL e pidginização/crioulização é o do grau da variação (*variability in “attainment”*), um termo usado em pesquisas de ASL para descrever o patamar de aproximação da variedade linguística do aprendente em relação à LA. Embora o termo *attainment* não seja adequado para os contextos de pidginização/crioulização (Baker, 1994:72), não deixa de estabelecer correspondência com a designação ‘aproximação ao superstrato’ (*proximity to the superstrate*), utilizada no campo das pesquisas crioulisticas.

Efectivamente, ambas as variações dependem da natureza da interacção entre os falantes envolvidos no contacto, já que esta é caracterizada pelos diferentes graus do limitado acesso ao superstrato, da limitada participação nas instituições e práticas reguladas pela língua lexificadora e baixa (ou nula) motivação relativamente à LA. Como tal, o *input* e a interacção, bem como a atitude e a motivação, têm influência notória no grau de variação registado quer na ASL quer na pidginização/crioulização.

A intervenção do factor idade estabelece também um paralelismo evidente entre interlínguas e crioulos. Bastantes estudos têm demonstrado que, na ASL, os aprendentes adultos não só são menos bem sucedidos do que as crianças (Lenneberg 1967; Chomsky, 1972; Brown, 1980; Larsen-Freeman & Long, 1991; Ellis, 1994; Beck, 1998) como se revelam também mais sensíveis a outras questões que interferem no grau de variação, como as sociológicas, as socioculturais e as psicológicas (Hyltenstamm & Abrahamsson, 2003:586). Nos contextos iniciais de línguas em contacto, adultos em condições socioculturais e psicológicas desfavorecidas configuram o público aprendente, deixando antever, à partida, um baixo grau de aproximação à LA.

Os paralelismos referidos permitem entender que algumas características universais da interlíngua podem ser utilizadas como ferramentas heurísticas para o estudo da formação dos crioulos, numa perspectiva comparada. Relativamente à questão das transferências em ASL, é geralmente aceite que podem ocorrer em todos os subsistemas, com especial incidência nos aspectos fonológicos. Alguns estudos crioulisticos têm entendido que similaridades estruturais entre L1 e crioulos representam evidências de transferências. Com base neste pressuposto, é possível estabelecer a assunção patente na tabela seguinte:

**Tabela 5.1.** A assunção: similaridades estruturais configuram transferência (Plag, 2008b:311).

	<b>Traço</b>	<b>L1 / substrato</b>	<b>LA / lexificadora</b>	<b>Interlíngua / Crioulo</b>
<i>a.</i>	traço x	+	-	+
<i>b.</i>	traço x	-	-	-

Em *a.* verifica-se que uma determinada estrutura está presente tanto na L1/substrato como na interlíngua, mas ausente na LA/superstrato, pelo que é forte a probabilidade de se estar perante uma transferência. Opostamente, como a estrutura *b.* não marca presença nem na L1/substrato nem na LA/superstrato, então também não pode ocorrer na

interlíngua. Contudo, para se confirmar se uma determinada estrutura representa transferência, é fundamental concluir que a mesma não emerge universalmente no desenvolvimento da L2, isto é, de forma independente, apesar de o mesmo tipo de estrutura existir na L1 do aprendente (Plag, 2008b:311). Assim, se tomarmos em consideração as questões inerentes aos universais de ASL, o cenário representado na tabela anterior torna-se mais complexo, conforme demonstrado na Tabela 6.1:

**Tabela 6.1.** Os factos: similaridades estruturais podem não representar transferência (cf. Plag, 2008b:311).

	<b>Traço</b>	<b>L1 / substrato</b>	<b>LA / lexificadora</b>	<b>Interlíngua / Crioulo</b>
<i>a.</i>	traço x	+	–	+
<i>b.</i>	traço x	–	–	–
<i>c.</i>	traço x	–	–	+ (!)
<i>d.</i>	traço x	+	+	– (!)

Como se constata agora, emergem novas estruturas (*c.*), enquanto outras, que são compartilhadas por ambas as línguas, desaparecem (*d.*). A exemplificação deste tipo de estruturas pode ser encontrada nos estudos sobre aquisição do alemão L2 por falantes de sueco L1, dois idiomas que compartilham o traço da inversão da ordem frásica em orações principais (V2nd) (Håkansson, Pienemann & Sayehli, 2002), de acordo com o representado na Tabela 7.1:

**Tabela 7.1.** Inversão da ordem frásica na interlíngua alemã (Plag, 2008b:311).

<b>L1: sueco</b>	<b>LA: alemão</b>	<b>Estágios da interlíngua</b>
V2nd	V2nd	1. SVO
*ADV SVO	*ADV SVO	2. ADV SVO
		3. ADV VSO

Nas duas línguas, a colocação do advérbio em primeira posição sem inversão é agramatical, uma similaridade que não é compartilhada pela interlíngua no seu estágio 2 de desenvolvimento, já que se mantém a ordem frásica SVO do estágio 1 da aquisição, configurando-se o esquema *c.* da Tabela 6.1 e não se confirmando a representação *b.* da Tabela 5.1. Note-se também que, apesar de o alemão e o sueco possuírem a estrutura V2nd, esta só emerge em sequência do desenvolvimento da interlíngua, ou seja, no estágio 3 (Håkansson, Pienemann & Sayehli, 2002:259). A questão da ordem frásica

levanta ainda outro problema: dado que a interlíngua emerge com a ordem SVO, inexistente na L1 e na LA, então a ocorrência deste tipo de traço não pode ser vista como uma transferência, caso a L1 possua a ordem SVO e entre em contacto, por exemplo, com uma língua do tipo VSO, ou seja, se configurem os traços patenteados em *a*. (Tabelas 5.1 e 6.1) e que apontariam para a evidência de transferência.

### **2.11.2. Hipótese do Desenvolvimento Moderado da Transferência (*Developmentally Moderated Transfer Hypothesis*)**

Face às questões apresentadas, é importante determinar quais os traços conectados a questões universais do desenvolvimento da interlíngua e quais os que resultam de transferência. Neste aspecto, a Teoria da “Processabilidade” (Pienemann, 1998, 2005b) e a HI (Plag, 2008a, 2008b) permitem não só elaborar assunções e predições mas também testá-las empiricamente (Figs. 13 e 14; Tabela 2.1, nota de fim de texto 63 e Tabela 3.1). De acordo com a hipótese de que a GU actua na L2 através da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), a FL0 da interlíngua tem que ser observado como a FLf da L1, o que leva a predizer que ocorrem transferências, principalmente na fase inicial da aquisição da L2. Contrastando com esta posição, Pienemann *et alii* (2005a, 2005b) e Pienemann & Håkansson (2007) avançaram com a Hipótese do Desenvolvimento Moderado da Transferência (*Developmentally Moderated Transfer Hypothesis*), que advoga que os aprendentes de uma L2 não iniciam a aquisição a partir da sua L1, mas sim activando o sistema de processamento e articulando os processos disponibilizados pela L2. De acordo com este pressuposto, qualquer estrutura da L1 só é transferida no caso de encontrar condições para ser processada no sistema de desenvolvimento da L2, isto é, caso o processador da interlíngua tenha disponível um processo que accione o processamento da estrutura da L1 passível de transferência. Como se depreende, a Hipótese do Desenvolvimento Moderado da Transferência não exclui a possibilidade de ocorrerem transferências, mas defende que elas são constringidas pela “processabilidade” da L2, ou seja, que elas se dão em tempo distinto daquele que é avançado pela sugestão de que a GU actua na L2 através da L1. Este aspecto pode ser de novo empiricamente observado nas construções que envolvem V2nd produzidas por falantes de sueco L1 em aquisição do alemão L2 (Tabela 7.1), já que os mesmos apenas produzem este tipo de estruturas no estágio 3 do desenvolvimento da

interlíngua (Håkansson, Pienemann & Sayehli, 2002:259). Este fenómeno apenas ocorre depois de os falantes terem ultrapassado dois estágios de desenvolvimento da interlíngua que apresentam estruturas sem qualquer correspondência na L1 e na LA, apontando para a possibilidade de o efeito apenas emergir quando o sistema de processamento é capaz de reorganizar a gramática da interlíngua com recurso aos apropriados conhecimentos da L1. Desta forma, não deixam de ser aqui pertinentes quer o princípio advogado por Montrul (2004:365) de que o desenvolvimento sintáctico é conduzido pela aprendizagem de novos itens lexicais e morfológicos quer determinados pressupostos da Hipótese da Maturação, que advogam que certos aspectos operativos da aquisição se dão de modo descontínuo, isto é, em período mais tardio do que aquele que seria previsto acontecerem (Borer & Wexler, 1987; Genesee, 1989; Radford, 1990, 1997a, 2000; Vainikka, 1993/1994; Meisel, 1994a, 1997).

Note-se que, em virtude de a transferência ocorrer apenas se o processamento o permitir, isto é, em estágios hierárquicos apropriados do desenvolvimento da interlíngua, pode também acontecer no estágio inicial, conquanto as estruturas transferidas sejam processáveis neste. É o que sucede, por exemplo, com a ordem básica SVO, quando ela se regista também na L1. De facto, após um breve estágio inicial em que formulam frases holofrásticas, os aprendentes de L2's passam a produzir, predominantemente, construções com a ordem canónica, isto é, SVO ou SOV, independentemente de a L1 e/ou a LA possuírem ou não estas mesmas ordens (Håkansson, Pienemann & Sayehli, 2002:259). A explicação para este comportamento inicial dos aprendentes é explicada por Pienemann, Biase e Kawaguchi (2005:229) em termos de Gramática Lexical Funcional, já que ocorre uma associação fixada entre a estrutura argumental, a estrutura funcional e a estrutura constituinte, designada pelos autores de “sequência não-marcada”.<sup>103</sup>

De acordo com esta teoria, o mapeamento directo não requer nenhum processador específico da língua nem qualquer armazenamento, mas permite que o aprendente produza sequências SOV ou SVO. Deste modo, e ao contrário do que seria de esperar de acordo com o predito pela hipótese de que a GU actua na L2 através da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), os aprendentes ingleses de japonês L2, por exemplo, começam por produzir a sequência SOV (Biase &

Kawaguchi, 2002:293), em vez de transferirem a sequência nativa SVO. Este aspecto, se observado na perspectiva da transferência vs. desenvolvimento universal, evidencia que nem a sequência SOV nem a ordem SVO produzidas por aprendentes de L2's podem ser vistas como transferências da L1, já que ambas são realizadas nos estágios iniciais da aquisição.

Observando-se agora a ordem básica de SN's com sujeitos e objectos não pronominalizados de diferentes crioulos, verifica-se que todos eles compartilham a estrutura canónica da ASL, apesar de os substratos e as línguas lexificadoras apresentarem, entre si, similaridades e/ou diferenças:

- [214] a. CRIOULO DO HAITI:  
 Crioulo do Haiti: S AUX VO  
 Gbe: S AUX VO / OV  
 Francês: S AUX VO  
 (Lefebvre & Brousseau, 2002)
- b. SURINAMÊS ou SRANAM:  
 Surinamês: S AUX VO  
 Gbe: S AUX VO / OV  
 Inglês: S AUX VO; X S AUX V  
 (Bruyn, 2002)
- c. PALANQUERO:  
 Palanquero: S AUX VO  
 Kikongo: S AUX VO  
 Espanhol: S AUX VO  
 (Bentley, 1967 [1887]; Laman, 1936; Schwegler, 1991)
- d. NEGERHOLLANDS:  
 Negerhollands: S AUX VO  
 Kwa: S AUX VO  
 Holandês: V2nd, X AUX SOV  
 (Muysken, 2001)
- e. CRIOULO DE BASE HOLANDESA DE BERBICE:  
 Crioulo de Berbice: S AUX VO  
 Ijo Oriental: SOV  
 Holandês: V2nd, X AUX SOV  
 (Kouwenberg, 1992, 1994)

Os exemplos evidenciam paralelismos estreitos entre estruturas dos estágios iniciais da interlíngua e estruturas crioulas, cujas emergências não podem ser totalmente

explicadas em termos de transferência. Assim, e de acordo com a HI, é possível prever que algumas línguas emergentes por contacto apresentem também a ordem SOV, como de facto acontece:

[215] PIDGIN DE NAGALAND:  
*Kikatemla modu kha-yas-ile*  
Kikatemla vinho comer-PROG-PRET  
“O Kikatemla estava a beber vinho” (Bhattacharjya, 2007:240)

De um modo geral, a HI, em congruência com a Teoria da “Processabilidade”, sustenta que, nos crioulos, a ordem das palavras reflecte uma sequência não-marcada, não havendo evidência em favor da transferência (Plag, 2008b:320).<sup>104</sup> Relembre-se também que, de acordo com os princípios da mudança evolucionista da língua, a saliência perceptual se pode impor à frequência na LA, fazendo com que uma estrutura não-marcada universalmente, por ser mais facilmente processável (Thomason & Kaufman, 1988:331), seja adoptada pelos falantes da L2 (Clements, 2009:22).

Tendo ainda em conta que a emergência da flexão contextual se instancia a partir do nível 4 (Tabela 3.1) da hierarquia da “processabilidade” (Plag, 2008a, 2008b), pode inferir-se que os crioulos não alcançam este patamar de desenvolvimento sintáctico. A causa tem então origem em factores de natureza psicolinguística, nomeadamente nos traços universais que motivam a ASL, responsáveis pela emergência de muitas das estruturas crioulas, geralmente consideradas não-marcadas.

Não obstante o contributo da HI para ajudar a clarificar o porquê de a flexão contextual se encontrar ausente em pidgins e crioulos, ficam em aberto questões para as quais a hipótese não fornece respostas satisfatórias: (i) Porque é que algumas flexões inerentes são preferidas em detrimento de outras? (ii) Porque é que a flexão morfofonémica é também aplicada em pidgins e crioulos, contrariando o pressuposto de que não deveria ocorrer nestas línguas; (iii) Porque é que, nestas mesmas línguas, a flexão inerente apenas recorre a um subconjunto de significados formais disponíveis, desfavorecendo a afixação e excluindo os casos de metafonia e supletismo?<sup>105</sup>

### **2.11.3. Transmissão linguística irregular**

Os conceitos de transmissão linguística irregular e de pidginização/crioulização reportam-se a processos históricos de contacto massivo e prolongado entre línguas, nos



quais intervêm um idioma que é tomado como LA pelos falantes dos demais dialectos. Todavia, o primeiro conceito é mais abrangente do que o segundo, visto envolver processos de natureza linguística e sociohistórica que conduzem ao emergir quer de uma nova língua (pidgin ou crioulo) quer de uma nova variedade linguística, que se tornam predominantes na situação de contacto. Portanto, a criouliização pressupõe uma situação de reduzido acesso à LA, que vai determinar o desencadeamento da reestruturação linguística e, conseqüentemente, o aparecimento de uma nova língua. Esta é caracterizada por acentuada não-aquisição de morfologia flexional, o que a torna bastante diferente das que lhe concederam os modelos iniciais para aquisição e conseqüente nativização. Por seu lado, a transmissão linguística irregular configura um *continuum* de níveis diferenciados de socialização/nativização das L2's adquiridas em situações sociohistóricas específicas. Desta forma, a ecologia destas línguas tem como conseqüência resultados distintos entre si, caracterizados por não-aquisição de morfologia flexional, a qual pode ir de grau mais leve a grau mais acentuado.

#### **2.11.3.1. Vectores para a emergência da transmissão linguística irregular**

As massas populacionais detentoras de diferentes L1's, quando constituindo uma comunidade, têm que adquirir unicamente por contacto, para efeitos de comunicação funcional, uma L2 de emergência comunicativa. Esta interlíngua transmite, lexicalmente, aquilo que outras línguas expressam de modo gramatical, sendo caracterizada por vocabulário reduzido mas utilitário, não-realização de segmentos com pouco peso fonético e uma acentuada redução gramatical, uma vez que os falantes apenas incorporam no novo dialecto os itens que preenchem funções gramaticais básicas. Por norma, a redução fica a dever-se a três factores principais: (i) pouco contacto com os detentores da LA, principalmente porque estes controlam o poder político e se configuram como um grupo minoritário de difícil acesso; (ii) o facto de a esmagadora maioria dos falantes em aquisição ser adulta, ou seja, iniciar a aprendizagem da L2 após o período crítico da aprendizagem; (iii) a ausência de aplicação de uma norma que oriente formalmente o processo de construção da interlíngua, já que esta responde apenas a questões de urgência comunicativa.

De modo geral, o que antes se definiu foi a emergência de um pidgin, isto é, de uma L2 a partir de *input* reduzido (modelos/estímulos fragmentados e deficientes) e que não é L1 de nenhum dos diferentes grupos em contacto. O distanciamento em relação à LA determina, desde logo, que o pidgin resulte essencialmente dos processos universais de aquisição da L2 (simplificação morfológica ou nivelção) e transferências da L1 (Schumann, 1978; Gilbert, 1983; Bickerton, 1984a, 1988), pelo que não pode ser explicado apenas em termos de simplificação da LA (Ferguson, 1971). Nesta fase do seu desenvolvimento, o pidgin não desempenha funções interactivas (coesão social), expressivas (ideias abstractas), poéticas e metalinguísticas isto é, algumas das funções que caracterizam as línguas que não emergem por contacto massivo. Portanto, acaba por funcionar de forma pragmática/pré-sintáctica, com base em discursos universais independentes da sintaxe (Givón, 1979, 1984), podendo ser definido, convencionalmente, como *um simplificado sistema linguístico não-nativo, lexical e estruturalmente reduzido, instável, com padrões variáveis e inconsistentes (devido à influência das diferentes línguas nativas em contacto), criado e desenvolvido rapidamente para ser usado como meio de comunicação emergencial (L2) em situação restrita de contacto funcional entre adultos que são falantes nativos de línguas mutuamente ininteligíveis*. Esta caracterização não é meramente linguística, reflectindo uma resposta social necessária e adequada à realidade imposta pelo contacto. Por outro lado, o pidgin impossibilita o bilinguismo, por razões sociohistóricas e políticas.

Contudo, se a situação de contacto se mantiver, a interlíngua de emergência evolui naturalmente, tornando-se o dialecto que preenche as funções sociais na rede de interacção linguística, ao mesmo tempo que vai ganhando estatuto de L1 para os descendentes dos falantes do pidgin e das outras línguas de substrato que vão engrossando a comunidade (Cornips & Hulk, 2006). A limitação estrutural da língua começa a revelar-se insuficiente para dar resposta às novas exigências funcionais da comunidade, pelo que esta leva a efeito a expansão do pidgin (Mühlhäusler, 1986; Siegel, 2008), gramaticalizando os itens lexicais importados da LA e incorporando novos dispositivos gramaticais das línguas do substrato (Thomason & Kaufman, 1988:73-74). Da troca entre expansão funcional e expansão gramatical emerge o crioulo, enquanto código linguístico que ganha identificação e referências próprias para passar a

preencher as necessidades sociolinguísticas da comunidade. Assim, e de um modo geral, a noção de crioulização surge relacionada com a emergência/desenvolvimento de uma L1, com *input* restrito, isto é, a partir de estímulos/modelos defeituosos de variedades de uma L2. Por seu lado, os efeitos do processo de aquisição variam de acordo com a natureza da L2, enquanto elemento que fornece os DLP's que actuam como estímulos para a aquisição da L1. Nesta fase, a língua nativizada detém ainda duas funções/características: a de L1 para os falantes nascidos no seio da comunidade; e a de L2 para os membros transplantados para a comunidade (DeCamp, 1971:16). Temos então crianças e adultos intervindo no processo de crioulização, com as primeiras emprestando a capacidade da competência linguística, enquanto os segundos a desdobram no processo de uso (Kroch, 1989a; Sprouse & Vance, 1999; Siegel, 2004b). Na comunidade processam-se então dois tipos de interacção linguística: a interacção na horizontal, entre todos os membros da comunidade; e a interacção na vertical, entre os falantes dos distintos substratos e seus descendentes directos, detentores da L1 nativizada. Quanto aos contactos com a LA, embora esporádicos, não cessam, ocorrendo, essencialmente, entre trabalhadores braçais e seus capatazes e/ou entre serviçais e seus amos. Assim, no momento em que a interlíngua da comunidade é objecto de expansão estrutural com a introdução de novas formas oriundas do substrato, estas entram em conflito com formas da LA.

Segundo Kihm (1980:12), para que uma língua seja considerada um crioulo, deve obedecer ao triplo “paradoxo crioulo”:

1. crioulos de bases diferentes assemelham-se mais entre si do que com a LA;
2. abstraído das diferenças lexicais, crioulos de bases diferentes estão mais próximos uns dos outros do que as LA's estão entre si;
3. crioulos da mesma base lexical não são, devido a isso, mais parecidos entre si do que os crioulos de bases diferentes.

Para alguns linguistas, estas similaridades são simplesmente conjunturais e decorrem da mera presença ou influência das línguas de substrato no processo de formação dos crioulos. Algumas propriedades são discutidas por Muysken (1981a; 1988) e vistas como resultando de valores não-marcados de parâmetros (Bickerton, 1988, 1999;

Roberts, 1997, 1999). Nesta etapa, já não faz então sentido continuar a observar a interlíngua em termos de pidginização ou criouliização, uma vez que o processo de transmissão da mesma varia em função de factores vários, provocando alterações significativas na sua estrutura. Por outro lado, é ainda de ter em consideração que a interlíngua poderá também entrar em contacto com outras línguas, como aconteceu no Brasil, quando o PtgL2 falado pelas comunidades escravas transplantadas passou a contactar com os dialectos dos grupos ameríndios ou oriundos de outras partes de África. Assim, o panorama que nos é oferecido é o de uma língua falada detentora de uma gramática reestruturada, mas que continua a incorporar, por empréstimo, traços das línguas com as quais passou a ter contacto (cf. Thomason & Kaufman, 1988; Lefebvre, 2003; Montrul, 2006; Lefebvre, White & Jourdan, 2006b; Siegel, 2008).

Nesta fase do novo contacto linguístico, os modelos linguísticos emergentes podem ser tidos como vigentes pelos descendentes dos membros da comunidade, que estabelecem uma norma linguística a partir dos mesmos, isto é, ao invés de facultarem a emergência de um novo sistema linguístico distinto da LA (um crioulo), apenas dão origem a uma nova variedade desta. Contudo, este novo código linguístico não deixa de exibir as características defectivas de que foi alvo durante o seu processo de socialização/nativização. Nesta conformidade os dialectos emergentes por transmissão linguística irregular, tal como os pidgins e crioulos, devem ser vistos como fenómenos que conjugam factores linguísticos e sociohistóricos na sua emergência.

#### **2.11.3.1.1. Sociohistória e transmissão linguística irregular**

Sociohistoricamente, o transplante de massas populacionais escravizadas entre os sécs. XVI e XIX por colonizadores europeus para novos contextos geográficos ou geográfico-culturais configurou-se como um dos factores para a emergência de línguas pidgins ou crioulas. A constante necessidade de mão-de-obra para implementação e expansão de empreendimentos agrícolas ou exploração mineira, sobretudo no continente americano, foi preponderante no processo de escravização de africanos, em virtude de as populações ameríndias, pelos seus hábitos culturais e de recolectagem, resistirem ao processo de aculturação e não possuírem características físicas para suportar esforço físico intenso.

No processo de transplantação para o *Novo Mundo*, os escravos resgatados no continente africano eram inicialmente conduzidos para o entreposto de São Tomé, onde permaneciam durante algum tempo a fim de prestarem serviço nos engenhos da cana-de-açúcar dos colonos, entretanto aí fixados. Detentores de diferenciadas L1's, os escravos viam-se na necessidade de aprender o PtgL2, enquanto código linguístico que respondia às necessidades emergentes de comunicação não só entre os membros da comunidade escrava (cf. Hymes, 1971a; Silverstein, 1972; Baker, 1994; Baker, 2001; Smith, 2006) mas também entre estes e os seus capatazes e senhores. Posteriormente, foi-se construindo um cenário complexo de interacções sociolinguísticas na ilha, que acabaram por deter papel preponderante, em maior ou menor grau, no processo de socialização/nativização da L2: constante chegada de novos escravos resgatados de vários pontos do continente africano, alguns já falantes de um PtgL2 defectivo; aumento da população nativa descendente dos escravos, que, conjuntamente com os membros transplantados, conduziu ao desequilíbrio populacional entre colonos e escravos, com os segundos constituindo a grande maioria da massa demográfica fixada na ilha; envio de cerca de duas mil crianças judias com cerca de oito anos de idade, de ambos os sexos, e oriundas de Castela, retiradas pelo rei João II de Portugal aos seus pais, a fim de pressionar os judeus residentes em Portugal a optarem pela religião cristã; incremento da população mestiça, já que o mesmo soberano havia permitido que os colonos portugueses tomassem uma escrava negra por companheira, população essa que, pelo seu bilinguismo e condição de homens livres, terá contribuído para um estreitamento das distâncias na rede de interacção sociolinguística entre colonos e escravos; êxodo dos colonos portugueses, em finais do século XVI e princípios do século XVII, que abandonaram o arquipélago por falta de segurança e regressaram a Portugal ou se fixaram no Brasil, levando o santomense a deixar de contactar com a LA e a estabilizar-se como basilecto (Ferraz, 1979:19); ocupação do arquipélago pelos holandeses em meados do século XVII e contacto da população residente com uma nova realidade linguística; chegada de novos fluxos de africanos a partir do século XIX, contratados em Angola e Moçambique, alguns falando PtgL2 defectivo marcado por características bantu, e que emprestaram mão-de-obra “contratada” às roças do cacau e do café; fixação de trabalhadores provindos de Cabo Verde, nos anos 50 do século XX, falantes do CCV;

criação de um sistema de ensino e aplicação de uma norma que orientava o processo de aquisição linguística; incorporação de elementos nativos nos quadros profissionais da administração e do corpo militar coloniais; retirada definitiva dos portugueses após a Revolução dos Cravos, em 1974; integração de professores nativos nos novos quadros do ensino, detentores de um código linguístico cujo registo se aproxima do modelo do português popular falado na ilha. Como se verifica, a conjuntura sociohistórica e linguística do arquipélago proporcionou que, diacronicamente, o crioulo se estabilizasse como L1 sob a forma de basilecto, ou seja, se elaborasse um código linguístico que favoreceu e fixou a transferência das estruturas dos substratos. Posteriormente, o contacto do santomense com o português determinou a emergência, em situação de transmissão linguística irregular, do PA reestruturado, marcado por variação que caminhava no sentido da aquisição das regras da LA. Como tal, as diferenças nos registos da comunidade foram-se acentuando geracionalmente em direcção à língua de maior prestígio, registando-se um quase apagamento de determinadas características estruturais dos substratos, directo ou ancestrais.

Os estudos de Baker (1982) sobre os crioulos das ilhas Maurício e Reunião revelaram que o maior ou menor acesso aos modelos da LA é determinante para o tipo de língua que emerge do contacto linguístico, o que viria a fundamentar o princípio de que o estudo sobre línguas em contacto não pode ser demarcado da noção de crioulição variável (Bickerton, 1984a). Nestes estudos, a crioulição do tipo clássico configurará um dos extremos do processo de aquisição de uma LA, em que o processo de reestruturação, face ao escasso acesso à LA, gera uma língua distinta daquelas que lhe forneceram os modelos para a aquisição/nativização. Por seu lado, o processo de transmissão linguística irregular configura um *continuum* de diversificados níveis de socialização/nativização de uma L2 adquirida de modo imperfeito ou de aquisição defectiva de uma L1 por grupos populacionais em contextos sociohistóricos específicos, como sucedeu em São Tomé.

#### **2.11.3.1.2. Transmissão linguística irregular: parâmetros linguísticos**

A transmissão linguística irregular tem como alicerce as situações de comunicação social defectivas em que emerge ou se desenvolve um código linguístico de urgência,

que assenta, por sua vez, em termos retirados do léxico da LA. O manancial dos itens lexicais presente na variedade simplificada da LA é mais ou menos extenso, apresentando formas que vão desde soluções sintáctico-gramaticais transferidas da L1 até aspectos mais efectivos da língua lexificadora. Deste modo, os processos de pidginização/crioulização levados a efeitos por comunidades compostas por falantes adultos de línguas tipologicamente distintas geram elementos da LA que contêm reduções e simplificações, mas também itens relexificados com base nas L1's. Portanto, os pidgins e crioulos vão exibir formas completamente inovadoras, isto é, inexistentes na LA (Ellis, 1985; Gass, 1988; Winford, 2003a; Siegel, 2006, 2008).

O fenómeno da redução visa fornecer soluções que tornam mais simples algumas formas gramaticais das outras línguas envolvidas no contacto e pode ser justificado por questões diversas. Em termos de uma abordagem construtivista (*constructivist approach*), a emergência de determinados itens do novo código linguístico ficar-se-á a dever não ao facto de os falantes terem a L2 como alvo, mas sim à necessidade de estabelecerem uma comunicação meramente interétnica (Hymes, 1971a; Silverstein, 1972; Baker, 1994; Baker, 2001; Smith, 2006), recorrendo, para tal, aos meios cognitivos que já se encontram disponibilizados (Plag, 2008a, 2008b). A constante integração de novos membros na comunidade pode conduzir a mudanças semelhantes às registadas pelos fluxos de imigrantes, uma vez que novas variedades são introduzidas na referida comunidade (Horvarth, 1985:84). Em trabalho levado a cabo no âmbito do European Science Foundation Project,<sup>106</sup> Klein & Perdue (1992) observaram, diacronicamente, a ASL por contacto entre grupos de imigrantes, tendo constatado que tanto os morfemas livres como os morfemas presos de função meramente gramatical se encontravam ausentes das variedades básicas faladas pelos ditos imigrantes (*foreigner talk*). Outro dos factores que intervém no processo de redução das interlínguas em emergência tem a ver com questões do tipo psicolinguístico, já que as limitações perceptuais por parte de falantes adultos de línguas tipologicamente distintas da LA podem levar estes a não estabelecerem a segmentação entre diferentes sequências morfológicas, que acabam sendo tratadas como um todo indivisível (DeGraff, 1999b:482), deixando-se, conseqüentemente, de se estabelecer a correspondência entre as características formais e as formas morfofonológicas de determinados itens da LA

(Prévost & White, 1999 – MSIH). Seja por um motivo seja pelo outro, certo é que a redução do acervo das funções desempenhadas pela comunicação verbal e a não-aquisição de matéria gramatical incidem, com maior acuidade, sobre as estruturas mais abstractas da língua, já que estas apresentam uma carga semântica menos transparente, logo menos funcional em situações de comunicação de emergência.

Lefebvre (1986; 1998) e Jake & Myers-Scotton (1998) avançaram com princípios advogando que a criouliização é um processo mental e sociolinguístico. Os resultados de Lefebvre evidenciam que acontece relexificação na ASL, confirmando-se que a mesma necessita de monolinguismo para poder ocorrer (Thomason & Kaufman, 1988:73-74). Outro aspecto a ter em conta é que, nas hipóteses de relexificação, o inventário de núcleos funcionais e suas propriedades permanece invariável na sua transmissão da língua de substrato para o crioulo. Contudo, a elaboração de novas categorias implica desenvolvimento da sintaxe complexa e da morfologia gramatical de uma língua em contacto na qual elas não existiam anteriormente, isto é, a emergência, por exemplo, de um sistema de marcação de TMA onde antes eram apenas usados advérbios. Klein & Perdue (1992:87) salientam o facto de, na linguagem dos imigrantes, os elementos funcionais não serem adquiridos da mesma forma que os itens lexicais. Deste modo, categorias funcionais como determinantes ou marcadores de TMA, apesar de terem origem nas categorias lexicais do superstrato e entrarem como categorias lexicais no crioulo incipiente, são reanalisadas como categorias funcionais com a não-aquisição da flexão, isto é, como marcadores pré-verbais. Na base deste desenvolvimento gramatical está a mudança linguística interna, envolvendo processos de reanálise e extensão, mas também a influência de outras línguas, especialmente dos idiomas do substrato (Siegel, 2004b:362).

O prolongamento da situação de contacto leva o código linguístico de emergência a ter de responder ao incremento da rede de relações sociais das comunidades. Deste modo, ele é adoptado como veículo de comunicação entre os membros da franja social dominada, que vai abandonando, a pouco e pouco, o uso das L1's, principalmente nas situações de falantes geográfica e culturalmente transplantados, como sucedeu em São Tomé. A cimentação do uso do código de emergência e a necessidade de este dar resposta às novas solicitações é determinante para a recomposição da sua estrutura



gramatical, que leva ao surgimento da nova língua, qualitativamente distinta dos dialectos inicialmente envolvidos no contexto do contacto. É o que sucede, por exemplo, com o *tok pisin*, da Nova Guiné, ou o *kriol*, de Belize, antigas Honduras Britânicas, vistos como pidgins expandidos, em virtude de se terem desenvolvido gramatical e funcionalmente, mas sem se tornarem na língua nativa da maioria dos seus utilizadores.

Este aspecto vem demonstrar que a reestruturação gramatical não está unicamente dependente da nativização da língua, conforme defendido pela linha bickertoniana. De facto, os contextos de crioulização determinam a aquisição de L1's em moldes distintos dos da aquisição de L1's que não conhecem uma situação de transmissão linguística irregular, uma vez que a variação orientada pela fixação dos parâmetros da GU está dependente dos DLP's que actuam como gatilhos (*triggers*) para fixação das regras gramaticais. Quando a sociohistória propicia a crioulização, os DLP's transmitidos pelos falantes adultos às gerações seguintes são bastante variáveis e defectivos, o que vai fazer com que a gramática dos mais novos se aproxime da dos mais idosos, apesar de não ser totalmente idêntica. Como se viu no ponto 2.11.2 do presente trabalho (Hipótese do Desenvolvimento Moderado da Transferência), os crioulos não ultrapassam, por norma, o estágio 3 do desenvolvimento da interlíngua (Plag, 2008b:325), o que compromete a morfologia flexional da L2, ficando esta marcada pelas reduções provenientes das fases iniciais da aquisição. Deste modo, também a aquisição da L1, sustentada por este tipo de DLP's em cadeia, vai apresentar defectividade, assente, por exemplo, em morfosintaxe do género analítico, que visa compensar a falta de informação advinda da ausência de morfemas de género e número. Nesta conformidade, factores externos (*E-factors*) de natureza sociohistórica constituem-se também como uma das causas das mudanças que ocorrem nas línguas (Meillet, 1906; Schuchardt, 1980 [1909]). E na crioulização, estes factores associam-se aos recursos linguísticos da faculdade humana da linguagem (factores internos > *I-factors*) para construir um novo sistema linguístico, definido por Rizzi (1999:465) como a “invenção gramatical”.

É tido como ponto assente que existem muitos traços em comum entre pidgins e crioulos, uma vez que ambos perdem, em maior ou menor grau, estruturas gramaticais, consoante o seu processo sociohistórico de formação. Este aspecto é determinante na fase inicial da transmissão linguística irregular, visto que vai, igualmente, determinar o

grau de reestruturação gramatical, se bem que os factores sociolinguísticos e psicocognitivos dêem também o seu contributo para esta. Posteriormente, a LA determina os vectores de expansão estrutural da nova variedade linguística, já que a progressão é feita em direcção a ela, com todas as envolventes que isso acarreta em termos de pressão sociolinguística. Nesta fase, compreende-se que os processos de transferência de formas dos antigos substratos sejam menores, até porque os falantes já são bilingues. Contudo, estruturas fossilizadas não deixarão de acompanhar, em diacronia, as consecutivas fases de expansão estrutural, ainda que o peso da reestruturação se faça sentir com maior intensidade não só em virtude da pressão sociolinguística mas também por causas bionaturais conotadas à aquisição de L1's (faculdade da linguagem).

#### **2.11.3.1.3. Transmissão linguística irregular e gramaticalização: heterogénese GU/transferência**

Genericamente, a gramaticalização pode ser definida como a mudança operada a nível dos itens e construções lexicais, levando a que percam, total ou parcialmente, o seu significado lexical, a fim de, em determinados contextos linguísticos, passarem a deter funções gramaticais que se encontravam ausentes do código linguístico vigente. Uma vez gramaticalizados, esses elementos continuam a desenvolver novas funções gramaticais (Hopper & Traugott, 2003:16). Assim, nomes e verbos portadores de significado lexical podem transformar-se, ao longo do tempo, em formas auxiliares, marcadores de caso e flexão ou conectores frásicos. Contudo, o processo de transição deste tipo de palavras principais em palavras gramaticais não é imediato, já que têm de passar por um gradual processo de esvaziamento lexical, isto é, transitar de palavras principais a palavras acessórias e, só depois, fixarem-se como palavras gramaticais.

Em termos generativistas, Roberts (1993:384-389) define a gramaticalização como a transposição de um item de categoria lexical para categoria funcional, acompanhada de não-aquisição do seu conteúdo funcional. Assim, ocorre também uma recategorização sintáctica do item ou forma gramaticalizada, levando a que o mesmo perca o seu significado referencial conotado às situações interlinguísticas (universo exterior da língua) e passe a representar um significado gramatical ligado às questões intralinguísticas (universo interior da língua). Com a recategorização sintáctica, o item

ou a forma gramaticalizada adquire “*propriedades funcionais na sentença, sofre alterações morfológicas, fonológicas e semânticas, deixa de ser uma forma livre, estágio em que pode até mesmo desaparecer, como consequência de uma cristalização extrema. Esse trajecto se dá no tempo real quanto no tempo aparente. Num sentido mais amplo, a gramaticalização é a codificação de categorias cognitivas em formas lingüísticas, aí incluídas a percepção do mundo pelas diferentes culturas, o processamento da informação, etc.*” (Castilho, 1997a:31-32). Um exemplo bastante conhecido de gramaticalização é o processo pelo qual a forma lexical inglesa “*let us go*” foi reduzida para a construção simples “*let’s*”, a qual, posteriormente, se cristalizou em “*lets*”, como acontece, por exemplo, na frase “*lets you and me study together*”. Como se pode verificar, a forma perdeu o seu significado lexical de “permitir; deixar”, adquirindo a categoria de auxiliar, ao mesmo tempo que o pronome “*us*” se reduziu a um sufixo, primeiro, e a um mecanismo puramente fonémico, depois. Com base na hipótese da unidirecção (*Unidirectionality Hypothesis*), Hopper & Traugot (2003:7) propõem o seguinte modelo para definir os diferentes estágios da trajectória da gramaticalização:

item de conteúdo > item gramatical > clítico > afixo flexional

Os contextos sociohistóricos que são favoráveis à ocorrência de processos de pidginização/crioulização são-no, conseqüentemente, também propícios para que aconteçam reduções que potenciam posteriores gramaticalizações, uma vez que em situação de ASL não só a GU, que mantém os seus princípios acessíveis, actua na L2 através da L1 (Schwartz & Sprouse, 1996; Sprouse & Vance, 1999; Siegel, 2006; Sprouse, 2006), permitindo transferências a partir desta (Thomason & Kaufman, 1988:73-74), como também os falantes adultos revelam falhas no restabelecimento dos parâmetros (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashes, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), visto as categorias funcionais deixarem de estar disponibilizadas depois do período crítico do desenvolvimento.

Em fases posteriores do desenvolvimento, e conseqüente nativização do código linguístico vigente na comunidade, as reduções deixam de preencher os requisitos comunicativos impostos pelo processo de socialização, pelo que as novas gerações, normalmente bilingues, buscam nos itens do acervo lexical disponível (alguns dos quais

comprometidos por transferências resultantes, por exemplo, das limitações perceptuais dos adultos da primeira geração da ASL) aqueles que possibilitam papéis funcionais intralinguísticos, por forma a suprirem as necessidades interlinguísticas que emergem. Assim sendo, verifica-se que, na recomposição gramatical durante o processo de socialização/nativização de uma língua, intervêm dois fenómenos:

- (i) o da ampliação das funções dos itens gramaticais já existentes na fala da comunidade;
- (ii) o recurso aos itens lexicais para se expressarem relações gramaticais inexistentes, determinante no processo de gramaticalização da língua em emergência.

O sistema de marcação PL do santomense constitui um exemplo típico de gramaticalização conotado à influência do substrato, sendo efectuado com recurso ao item *inen*, do substrato edo, que é colocado em posição pré-nominal para funcionar como marcador de flexão em número (exemplo [38b]). Este tipo de estrutura, para além de evidenciar paralelismo com o sistema linguístico kwa (exemplo [39]), divide igualmente características com construções similares de outros CP's atlânticos:

[216] ANGOLAR:  
*ené*                      *omé*    (Ferraz, 1979:61)  
3PL    homem  
“os homens”

Para Holm (1988/1989:29), a evidência deste tipo de transferência é tão óbvia que nenhum argumento foi ainda proposto no sentido de atribuir tal gramaticalização à influência dos universais linguísticos. No caso do PA, detectamos reminiscências deste tipo de gramaticalização, agora com recurso ao pronome pessoal do português, 3ª pessoa do PL “eles” (exemplo [48]), revelando, mais uma vez que, para a produção linguística dos falantes bilingues de Almojarife, apesar de a língua activa (o PA) colaborar no processamento da fala, as estruturas da língua em repouso (o santomense) se mantêm latentes (Green, 1986:219), provocando interferências na primeira.

Contudo, a propósito da relação entre animacidade e marcação de número, Alexandre & Hagemijer (2007:42) chamam a atenção para, no santomense, a sequência *inen+N*

apenas se realizar com nomes de traço [+humano], expressando um significado definido que diz respeito a um grupo não-individualizado familiar ao ouvinte e ao falante (ponto 1.3.2.2.1 do presente trabalho). Nesta conformidade, o pronome pessoal *eles* poderá ser entendido, em alternativa, como uma forma desviada do objecto indirecto da frase (“*outra pessoa foi, tirou-lhes a roupa e viu a arma lá*”), já que, como foi visto a propósito das relações sintagmáticas e oracionais patentes na fala de Almojarife, é frequente os almojarifanos não fazerem a distinção entre formas rectas e átonas do pronome pessoal, com tendência para uso das primeiras em detrimento das segundas (ponto 1.3.2.3.1 – exemplos [147] e [148]). Contudo, e a confirmar-se esta possibilidade, a realização em observação apresentar-se-ia descontextualizada na situação em que foi produzida, tornando-se completamente desprovida de conteúdo lógico-semântico.

Outra possibilidade seria a de o falante estar a usar um possessivo e não um pronome pessoal (“*outra pessoa foi, tirou a roupa deles e viu a arma lá*”). Porém, importa referir que o mesmo informante faz um uso correcto do possessivo em outras construções (exemplo [217]), pelo que não é provável que se esteja na presença deste item gramatical:

[217] PA:  
*puxemo laço assim... rabo dele fica assim... [ZECAH1]*  
“puxámos o laço assim... o rabo dele ficou assim...”

Paralelamente, e como se verá no capítulo 4, dedicado à análise das variáveis independentes do tipo estrutural, as formas das L1's africanas exercem, conjuntamente com as do crioulo, importante influência no ponto de entrada do item com morfema de PL que origina a configuração do SN PL do PA. Deste modo, o que se verifica é uma tendência para a entrada da marcação ocorrer categoricamente na posição imediatamente anterior ao núcleo do SN, o DET, levando ao consequente apagamento das marcas de PL nos outros itens da cadeia sintagmática (*tirô eles rôpa*), em virtude de elas se revelarem redundantes (Kiparsky, 1972:195; Castro & Ferrari-Neto, 2007:75). Veja-se ainda que a forma “*tirou a roupa deles*” implicaria o uso de um nome massivo, categoria morfológica afectada pela redução gramatical que se verifica nos estágios iniciais de ASL por adultos e que, no caso da totalidade dos dados do nosso *corpus*, iria constituir a única ocorrência registada deste tipo de elemento gramatical.

Os exemplos [38b] e [48] evidenciam também como itens e formas do substrato se conjugam com os do superstrato no processo de socialização/nativização das L1's que emergem a partir de DLP's fornecidos por L2's defectivas, estando tal conjugação na base de uma heterogénese orientada por princípios universais e transferências dos substratos, tudo com vista a possibilitar uma comunicação efectiva. Assim, e embora o item gramaticalizado sofra uma alteração a nível meramente tradutológico, respondendo à pressão social que leva a que algumas formas sejam progressivamente abandonadas e substituídas por outras da LA (Thomason & Kaufman, 1988; Lefebvre, 2003; Montrul, 2006; Lefebvre, White & Jourdan, 2006b; Siegel, 2008), a verdade é que o sistema estrutural de marcação PL, por partilhar um traço de marcação estrutural com o do santomense, se mantém conforme o padrão das ancestrais L1's, quiçá por fossilização geracional. Isto mesmo parece comprová-lo o facto de a construção do exemplo [48] ter sido produzida por um falante da geração menos idosa de Almocharife e com um nível de escolarização que classificámos de médio na variável independente *escolaridade*.

Estas questões levam então a considerar que a reestruturação da variedade que emerge por transmissão linguística irregular ocorre em dois planos distintos, mas que confluem para a sua elaboração: por um lado, e a um nível mais profundo, actuando os dispositivos da faculdade da linguagem regulando estruturas das línguas do substrato e da LA que, conforme as diferentes situações sociohistóricas, vão conceder os distintos DLP's para accionarem a nativização da língua; e por outro lado, a um patamar mais superficial, acontecendo a escolha dos elementos lexicais que vão desempenhar as funções gramaticais requeridas pelas referidas estruturas.

Centrando a atenção no primeiro aspecto, o *bioprograma* de Bickerton (1981, 1983, 1984a, 1988, 1999), partindo da perspectiva avançada por Borer (1984) de que os parâmetros sintácticos são fixados a nível do léxico, apresenta uma distinção entre o que acontece na aquisição das L1's em situação que não configura transmissão linguística irregular e o que sucede no processo de nativização dos crioulos. Assim, enquanto às crianças da primeira é facultado todo um manancial de itens gramaticais que lhes vai permitir responderem às solicitações definidas pela sintaxe inata, às crianças do segundo não são disponibilizados tais itens gramaticais, o que as leva a terem de os criar a partir

dos elementos lexicais de que dispõem, os quais são, assim, alvo de erosão na sua função inicial: a de conceder o significado.

A gramaticalização da forma lexical inglesa "*let us go*" em "*lets*" revela que este é um fenómeno que actua também ao nível da evolução histórica das línguas que não emergem por contacto. Com efeito, todas as línguas são afectadas por diversos fenómenos no seu processo de desenvolvimento natural, que não só mostram que elas são entidades dinâmicas que apresentam variações ao longo do tempo como evidenciam também que os falantes não reagem todos do mesmo modo face às mudanças que decorrem nos seus idiomas. Se uma determinada forma linguística parece estar a ser substituída por outra, seja ao longo do tempo seja numa determinada dimensão socioeconómica ou demográfica da comunidade, então poderemos estar face a um processo de mudança.<sup>107</sup> Um dos fenómenos de gramaticalização que afectou o português e continua activo nas línguas adquiridas por contacto tem a ver com a gramaticalização do demonstrativo, que fez com que este passasse a deter as funções de artigo. Na passagem do latim para o português, o demonstrativo latino *ille/illa/illud*, com função deíctica/fórica ou enfática (positivo) (Andrade, 1999:161), forma acusativa, originou o artigo definido. No *papia kristang* também não existe este artigo, podendo a sua função ser desempenhada pelos demonstrativos *isi* ou *isti* (este) e *akeli* (aquele) (Hancock, 1969; Baxter, 1988). Por seu lado, o fanakalo<sup>108</sup>, um pidgin usado como língua franca, principalmente nas minas de ouro, diamante e carvão das zonas industriais mineiras da África do Sul, recorre ao item zulu *lo*, que faz as funções de demonstrativo, de artigo definido e de artigo indefinido (Ferraz, 1979:74). O uso deste último demonstrativo estabelece um claro paralelismo com o do MSPEC *se*, do santomense, que chegou a ser apontado como detendo função de artigo definido (Ferraz, 1979; Lucchesi, 1993):

[218] FORRO:  
*kanwa se* (Ferraz, 1979:74)  
canoa MSPEC  
“esta/aquela/a canoa”

A posterior socialização da língua e o prolongado contacto do PAL1 com o PE iriam determinar a aquisição do artigo definido da LA, originando-se formas híbridas do demonstrativo/artigo, conforme se pode constatar no exemplo seguinte:

[219] PA:  
*táqui*                    *issos*                    *home* [ZECAH1]  
está+LOC            DEM                    homem  
“estão aqui esses/os homens”

A diferença que ocorre a nível da gramaticalização entre as línguas que emergem em situação de contacto massivo e as que emergem sem ser por contacto massivo tem a ver com o ritmo de evolução e extensão do processo de variação, muito mais rápido no caso dos idiomas submetidos a contacto horizontal, um aspecto que alguns defensores da mudança evolucionista atribuem à característica poliplóide destas línguas (Mufwene, 2001:151-152). De facto, enquanto a gramaticalização de construções como as que referimos no inglês ou no português têm de superar estruturas fixas e normatizadas, a fim de se fixar a mudança, no caso das línguas emergentes por contacto, a erosão estrutural vai possibilitar uma reestruturação gramatical extensa e profunda em período relativamente curto, isto é, no tempo de uma ou duas gerações. Deste modo, e tal como defende Roberts (1997:23), o carácter criouliizante de uma determinada língua não se define em consequência das suas características estruturais, mas tão só em função do processo sociohistórico que determina a sua emergência. Aliás, nem de outra forma poderia ser, uma vez que, segundo o autor, a aquisição das línguas crioulas é determinada pela faculdade humana da linguagem, sendo as suas características estruturais e funcionais geneticamente codificadas da mesma forma que acontece com todas as outras línguas transmitidas geracionalmente.

### **2.11.3.2. Variedades criouliizantes**

O pressuposto de que os parâmetros sintácticos são fixados a nível do léxico (Borer, 1984; Bickerton, 1999) levantou também a hipótese de as línguas crioulas virem geneticamente programadas com parâmetros não marcados, princípio que Roberts (1997:12-13) considerou válido, mesmo nos casos em que a LA lexificadora é determinada por parâmetros marcados. Segundo o autor, os parâmetros dependem dos valores dos traços abstractos dos núcleos funcionais, acontecendo a variação paramétrica



em virtude de cada núcleo poder activar movimentos de diversos tipos. Como se viu a propósito da fixação do valor de parâmetros no processo de aquisição (ponto 2.2.2.1 do presente trabalho), os movimentos são determinados pelos valores «+», originando, por sua vez, a sintaxe complexa. Como as circunstâncias que subjazem à génese das línguas crioulas originam não-aquisição de morfologia, isto é, sintaxe simplificada, os valores não marcados de parâmetros serão os que se apresentam por defeito nos DLP's que vão funcionar como estímulo para as primeiras gerações de falantes do crioulo. Nesta conformidade, os referidos DLP's vão apresentar-se defectivos em relação às propriedades responsáveis pelo desencadeamento dos parâmetros de valor «+», fazendo com que as línguas crioulas sejam marcadas, por exemplo, por ausência de movimento do verbo, pela realização do sujeito referencial ou pela ordem SVO.

Não obstante, estas características não são exclusivas das línguas crioulas, já que, por exemplo, o inglês também não tem movimento do verbo, não realiza sujeitos referenciais nulos e obedece à ordem SVO. Deste modo, não se pode determinar o carácter crioulo de uma língua apenas em função da sua estrutura linguística, havendo necessidade de recorrer, para tal, à sua sociohistória particular de aquisição em contacto massivo. Como esta é determinante no processo de não-aquisição de morfologia e de outros elementos da LA, a fase inicial da transmissão linguística irregular vai aliar estes fenómenos à reestruturação gramatical em processo. Neste ponto, é importante referir que, na transição do sistema de marcação fonológico proposto por Jakobson (1966 [1932]) para o de Chomsky (1981b), o paradigma dos traços marcados perdeu o seu carácter restritivamente binário, sendo possível a sua multivalorização (Battistella, 1996:132). Neste aspecto, o pressuposto bickertoniano vai determinar que não existirá limite *a priori* para o número e tipo de instanciações não-defeito. Como tal, se o traço [+vozeado] contrasta unicamente com o [-vozeado], o mesmo não sucede, por exemplo, com o caso nominativo (um parâmetro não-marcado), que contrasta não-unicamente com todos os casos marcados, mas que contam como uma unidade no que concerne à marcação. Por outro lado, se levarmos em conta, por exemplo, que as construções anafóricas representam formas marcadas (Chomsky, 1981c) em virtude de caírem fora do escopo do SN (p.e. [*os livros*] uns dos outros), este aspecto, por si só, não nos revela qual será a construção não-marcada que expressa a mesma ideia. As possibilidades são

imensas, podendo recorrer-se a uma frase que seja semântica, mas não sintacticamente uma anáfora recíproca (exemplo [220]), aplicar-se uma forma que explicita a questão (exemplo [221]), ou utilizar-se uma construção ambígua (exemplo [222]):

[220] PTG:  
*Eles leram os livros um do outro.*

[221] PTG:  
*Cada um leu o(s) livro(s) do(s) outro(s).*

[222] CGB:  
*Kada un lei libru di utru.* [Kihm, 2008:414]  
“Cada um leu o livro do outro”

O último exemplo levanta uma questão: como em situação de socialização/nativização de uma L2 adquirida de modo imperfeito as crianças se encontram expostas a DLP's que podem não reflectir os parâmetros instanciados por defeito pelo bioprograma não-adulterado, as opções marcadas são limitadas em número, o que as leva a estabelecerem intrinsecamente uma relação finita com a opção não-marcada. Neste contraste de  $n_{marc} = 1$  vs.  $\{marc\}$ ,  $\{marc\}$ , apenas um conjunto limitado de formas permite originar o parâmetro  $n_{marc}$ , e vice-versa, isto é, pode determinar um “*sistema binário ampliado*” (Kihm, 2008:414). Este aspecto irá então ser fundamental na definição das características das línguas que emergem por transmissão linguística irregular, enquanto fenómenos altamente variáveis que conjugam factores psicolinguísticos e sociohistóricos na sua génese. Como tal, é possível avançar, desde já, com algumas particularidades que distinguem este tipo de línguas:

- (i) Não-aquisição ou variação a nível da morfologia flexional e de determinados itens gramaticais;
- (ii) Gramaticalização de itens lexicais para compensar lacunas do código linguístico;
- (iii) Alteração dos valores dos parâmetros sintácticos em função de valores não-marcados por defeito, já que estes não implicam a realização de movimentos aparentes (*covert movements*)<sup>109</sup> na cadeia frásica.

### 2.11.3.3. Transmissão linguística irregular e variação: gramáticas em competição

A emergência de pidgins e crioulos deve-se, em grande parte, à conjugação de duas situações: pouco acesso à LA no período inicial do contacto massivo entre línguas tipologicamente distintas e, posteriormente, durante a fase de socialização/nativização da interlíngua, uma fraca influência exercida na reestruturação gramatical desta pela LA. Em sequência de tal, os fenómenos que afectam a morfossintaxe dos crioulos são, regra geral, mais radicais e profundos do que aqueles que se registam nos dialectos resultantes de transmissão linguística irregular. Por norma, a afectação traduz-se em não-aquisição elevada – ou total – não só da morfologia flexional a nível do nome e do verbo mas também das regras de concordância a eles associadas, que praticamente não transitaram para determinados crioulos, como acontece com o santomense. Este dialecto, a exemplo de outros CP's atlânticos, apresenta, então, uma acentuada alteração nos seus parâmetros sintácticos, reflectida em gramaticalização e/ou transferências das línguas dos substratos do grupo níger-congo atlântico, como o demonstra o desenvolvimento de um sistema de partículas de tempo assente na reestruturação de advérbios e determinadas formas verbais, que se gramaticalizaram como marcadores das categorias gramaticais do verbo. Por seu lado, o valor marcado de parâmetro de sujeito nulo da LA foi alterado, adquirindo a função de valor não-marcado no santomense e nos outros CP's atlânticos. Como tal, uma das características que distingue estes CP's da sua LA é a obrigatoriedade de se realizar o sujeito referencial, excepto em determinados casos, como acontece com os sujeitos expletivos:

[223] CCV:  
*Sta faze calor oji.* (Baptista, 2002:254)  
PROG fazer calor hoje  
Hoje está calor.”

Mas se o santomense, enquanto basilecto, configura uma situação de afectação morfossintáctica profunda, já outras línguas – ou variedades de uma mesma língua – tidas como crioulas, ou sejam, os acrolectos e os mesolectos, apresentam diferenciados graus de reestruturação gramatical, como acontece com os crioulos do Barlavento e do Sotavento cabo-verdianos. Tal diferenciação é consequência do grau de intensidade do contacto entre línguas, seja na fase inicial deste seja no *continuum* pós-crioulo, que leva,

consequentemente, a um maior ou menor índice de reestruturação. Este, por sua vez, pode mesmo traduzir-se em descrioulização, isto é, na substituição de estruturas marcadamente crioulas por formas da LA.

Já as situações de transmissão linguística irregular configuram fenómenos de reestruturação morfossintáctica mais leves, centrando-se a variação mais a nível das formas de uso dos elementos gramaticais e não tanto no plano das alterações da estrutura linguística. Assim, na fase inicial do contacto, ao não atingir o grau de erosão morfológica registada no caso dos pidgns, a variação em situações de transmissão linguística irregular verifica-se na aplicação dos morfemas flexionais e uso dos elementos gramaticais. Este aspecto torna nubloso o tipo de estímulos responsáveis pela marcação dos valores dos parâmetros sintácticos, que não podem ser fixados na ausência de DLP's processáveis (Kihm, 2008:415), originando um acentuado grau de variação a nível do uso colectivo da língua, sem que se registre, nesta fase da emergência da nova língua, uma elevada quantidade de dispositivos originais relativamente à gramática da LA. Paralelamente, as lacunas são suprimidas com dispositivos da LA (Thomason & Kaufman, 1988; Lefebvre, 2003; Montrul, 2006; Lefebvre, White & Jourdan, 2006b; Siegel, 2008), originando-se um sistema de variação binária entre presença e ausência de mecanismos gramaticais da LA, uma vez que convivem na fala da comunidade duas ou mais gramáticas geracionais, isto é, diferentes valores de parâmetros sintácticos.

No caso da aquisição do PtgL1 por parte dos descendentes almxarifanos, deu-se com base em DLP's compostos por dados fragmentários e defectivos do PtgL2 adquirido pelos seu progenitores, configurando uma situação de transmissão linguística irregular que pode ser esquematizada da seguinte forma:

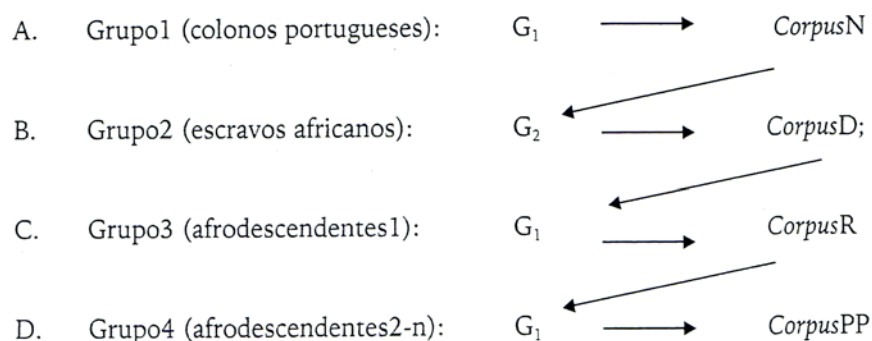


Fig. 28. Inglês: Modelo de aquisição para situação de contacto entre línguas (Lucchesi & Ribeiro, 2009:145).

Em S. Tomé, a relação entre os colonos portugueses (Grupo1) e os escravos adultos africanos (Grupo2) determinou a aquisição defectiva de uma L2 pelos segundos, com base nos dados fornecidos pelos primeiros (*CorpusN*). Esta variante é caracterizada quer por simplificações a nível dos mecanismos gramaticais mais abstractos e sem valor referencial quer por especificidades decorrentes de transferências e reanálises orientadas pelas gramáticas das L1 africanas (*CorpusD*). Posteriormente, a partir deste modelo de PtgL2 defectivo, os afrodescendentes do Grupo2 adquiriram o Ptg L1 característico de Almojarife. Este incorpora tanto as simplificações do *CorpusD* como transferências do substrato, que materializam eventuais fossilizações, sejam elas temporárias sejam permanentes. Por fim, ocorreu uma situação de transmissão linguística geracional entre os segmentos afrodescendentes (Grupo3 e Grupo4), com base no Ptg nativizado e reestruturado pelos afrodescendentes1 (*CorpusR*). Assim, a fala produzida pelos afrodescendentes2-n (*CorpusPP*) tendeu e tende a diminuir as interferências iniciais resultantes do contacto entre línguas, ao incorporar, continuamente, estruturas gramaticais do português falado tanto pelos outrora colonos portugueses como pelas actuais classes elitistas santomenses, que fizeram ou continuam fazendo a sua formação académica em Portugal.

Como se constata, aconteceu em São Tomé (e Almojarife) uma relação de implicação entre os diversos níveis. Daí resulta que a maior ou menor reestruturação na gramática do Grupo3 (afrodescendentes de primeira geração) será directamente proporcional à maior ou menor simplificação/contaminação produzida no *input* que configura o *CorpusD*. A partir de então, “o processo de transmissão da língua da segunda geração de afrodescendentes em diante se assemelhará mais a um processo de descrioulização” (Lucchesi & Ribeiro, 2009:146). Contudo, tal processo descrioulização não terá encontrado via livre para se poder consumir plenamente. Com efeito, a relação A-B-C repetiu-se ciclicamente na ilha, já que a chegada de africanos foi ininterrupta, primeiro traduzida nas levas de escravos em trânsito para outros continentes, com alguns emprestando temporariamente mão-de-obra aos engenhos do açúcar, e, posteriormente, reflectida nos contingentes de contratados para trabalharem nas roças do cacau e do café, muitos deles falando um PtgL2 de características idênticas ao do *CorpusD*.

Verifica-se então que, nos casos de transmissão linguística irregular, se observam duas etapas fundamentais que concorrem para a reestruturação da língua emergente: (i) o período inicial do contacto, marcado por variação mais, ou menos intensa, no uso dos mecanismos gramaticais, já que, nesta fase, a ausência de DLP's processáveis impede a fixação de parâmetros sintácticos; (ii) as fases seguintes de reestruturação linguística, com variação nas frequências de uso dos parâmetros sintácticos, fixados a partir do parâmetro da não-marcação fornecido por defeito às crianças e das pistas detectadas por estas nos DLP's defectivos, que as forçam a modificar o parâmetro por defeito em determinada direcção (Lightfoot, 1999a:439). Assim sendo, durante a primeira fase ocorre uma competição de formas gramaticais de todo o tipo, com os falantes a seleccionarem as que melhor respondem às necessidades de comunicação, cujo uso passa a ser generalizado. Seguidamente, as formas adoptadas acabam por se configurar como estímulos nublosos, isto é, como DLP's de parâmetros indeterminados (Kegl, Senghas & Coppola, 1999:2006), que facilitam diferenciados valores para a fixação dos parâmetros sintácticos que acontece nas fases posteriores de aquisição. Nestas, o nivelamento das formas nuelas pode mesmo levar à diluição diacrónica de algumas protoformas (Mufwene, 1999, 2001; Siegel, 2008), já que estas não conseguem resistir à pressão exercida quer pelos processos de gramaticalização determinados pela heterogeneidade da fala da comunidade quer pelas formas gramaticais que se cimentam por pressão da LA.

Nas variedades reestruturadas por transmissão linguística irregular, caso do PA – e variedades populares do PB –, são comuns, então, os seguintes quadros gramaticais:

- (i) Não-realização de dispositivos gramaticais marcadamente abstractos e de uso restrito na LA. É o que sucede, por exemplo, quer com o apagamento de marcas morfológicas de número, que afectam a concordância a nível do SN (sobretudo do SDET – exemplos [227a] e [227b]) e do SV (exemplo [224]), quer com a supressão do modo conjuntivo (exemplo [225]):

[224] PA:  
*vinte cinco de Abril nós ficou aqui muito abalado.* [SALVH3]  
“em vinte e cinco de Abril nós ficámos aqui muito abalados”

[225] PA:  
... *talvés esse semana... essa semana ele ão veio*. [MANOH3]  
“talvez nessa semana... nessa semana ele não tenha vindo”

(ii) Recomposição da estrutura gramatical da L1, ao encontro da LA, reduzindo-se o grau de variação de determinadas formas, como sucedeu com a marcação de género. Esta terá sido reintroduzida no PA (Figueiredo, 2009a) e na grande maioria das variedades populares do PB, mantendo-se a sua variação em maior escala apenas nas comunidades que conheceram um profundo processo de transmissão linguística irregular. A este propósito, veremos no ponto 4.1.2 do presente trabalho que a variável independente marcação de género não foi considerada relevante para observação do fenómeno da CPL-var que ocorre no SN do PA. Por outro lado, veja-se também como, em alguns casos, a pressão exercida pela LA conduz à eliminação da variação de determinadas formas, como sucede em alguns registos do falante [CASTEH1], detentor do 8º ano de escolaridade, apontados nos exemplos [156] e [157].

(iii) Conservação da variação no esquema presença/ausência do mecanismo gramatical da LA, como acontece a nível da concordância nominal de número (exemplos [190] e [191]) ou das marcas de pessoa-número do SV (exemplo [264]):

[226] PA:  
*nós fomos saber qual é lotes que nós vão receber*. [ABILH2]  
“nós fomos saber quais eram os lotes que nós íamos receber”

(iv) Manutenção do tipo ternário no uso da variação gramatical, isto é, de formas da LA (exemplo [227a]), da variante proveniente do processo de reestruturação da gramática (exemplo [227b]) e da variante zero (exemplo [227c]):

[227] PA:  
a. *todos os dias si ê nô i, ê perdo dia* [ANTOM1]  
“todos os dias que eu não for, eu perco o dia”  
b. *Correu com todos branco na roça*. [MINISM3]  
“Correu com todos os brancos da roça”  
c. *são coesa acontecida* [CASTEH1]  
“são coisas (acontecidas)/ são coisas que acontecem”

- (v) Alteração em várias estruturas típicas da LA, como é o caso das construções relativas (exemplo [160]), ou que envolvem partículas gramaticais como o artigo (exemplos [124] e [125]), as preposições (exemplos [141] a [144] e os clíticos (exemplos [147] a [151]);
- (vi) Alteração nas frequências de uso em relação à marcação de determinados parâmetros sintáticos, como sucede com a redução na aplicação do sujeito nulo (Duarte, 1993) (exemplo [224]), ou a não-realização de movimento do verbo (Morais, 1993; Rossi, 1993) (exemplo [228]), já que a ordem das palavras na estrutura frásica não tem consequências para a interpretação. As construções que configuram estes dois tipos de estruturas podem ser observados sob a óptica de uma expansão da gramática crioulezante, já que a marcação de valores consistentes para esses padrões sintáticos é fortemente generalizada no PE:

[228] PA:

*Depois eu sempre passávamo pá fila das pissoa* [OSVALH1]

A propósito desta realização, importa dizer que a concatenação da estrutura (*merge*) é o requisito mínimo para ocorrer sintaxe, pelo que não representa um parâmetro marcado. Em contrapartida, o movimento é a operação marcada, o que faz sentido do ponto de vista minimalista de uma economia da derivação e também se considerarmos a deslocação como uma “imperfeição” (*imperfection*) do *design* da língua (Chomsky, 1998).<sup>110</sup> Como a flexão implica comportamento sintático, surge envolvida no movimento do verbo. Desta forma, a não-aquisição de morfologia flexional determinada pela transmissão linguística irregular pode originar variação devido à não-movimentação do verbo para domínios funcionais mais altos.<sup>111</sup> A ausência de movimento aparente do verbo de “V-para-I” sucede também na língua inglesa (Pollock, 1989), que revela concatenação, mas cuja pobreza flexional origina a ordem ADV>V>O, registada nos exemplos [229a] e [229c], obtendo-se o diagrama da Fig. 29. No PE, contudo, é possível “movimento do verbo de V-para-I”, devido à elaborada flexão gramatical da língua (exemplo [230]), situação que acontece também no francês (exemplo [231]), podendo realizar-se a ordem V>ADV>O em ambas as línguas (Fig. 30).<sup>112</sup>



- [229] INGLÊS:  
 a. *John often kisses Mary.*  
 b. \**John kisses often Mary.*  
 c. *John has often kissed Mary.*

[230] PE:  
*O João beija frequentemente a Maria.*

[231] FRANCÊS:  
*Jean embrasse souvent Marie.*

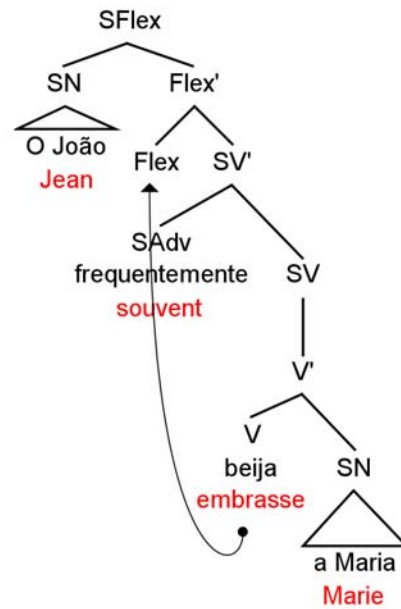
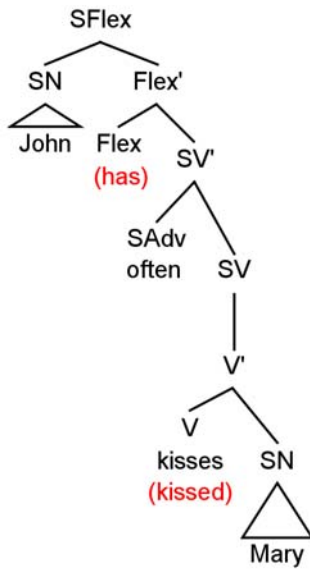


Fig. 29. Inglês: ausência de Mov-V.  
 (a partir de Silva, 2010:15)

Fig. 30. Português e francês: Mov-V-para-I.  
 (a partir de Silva, 2010:15)

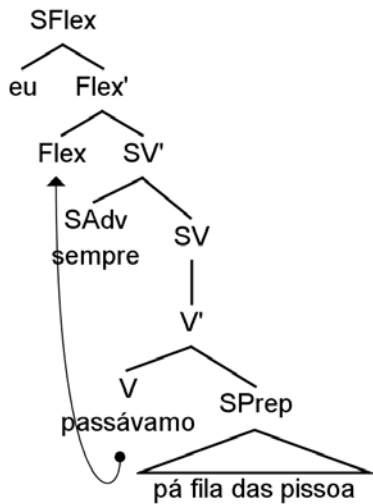


Fig. 31. PA: Mov-V-para-I.

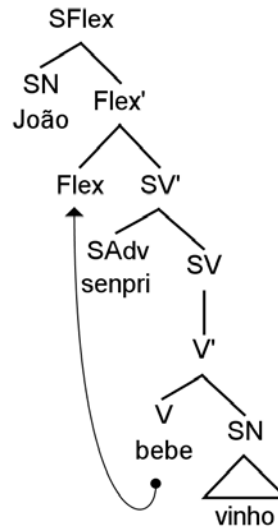


Fig. 32. CCV: Mov-V-para-I.

Todavia, note-se que a discussão em torno da questão do movimento do verbo não tem sido consensual na literatura actual. Analisando a posição de alguns advérbios do crioulo de base francesa do Haiti, entre os quais o advérbio “sempre”, DeGraff (1997:68-70) apresenta paralelismos entre esta língua e o inglês, e que a fazem divergir do idioma lexificador. A conclusão é que, no haitiano, tal como no inglês e ao contrário do francês, em virtude de ser uma língua de flexão “pobre”, o verbo apresenta ausência de “movimento de V-para-I”, isto é, instancia-se o parâmetro não-marcado que estabelece a força dos traços do verbo. Assim, neste tipo de línguas, o advérbio parece constituir uma barreira intransponível para a subida do verbo para a flexão, levando este a ganhar traços numa posição mais baixa da estrutura, isto é, na posição original.

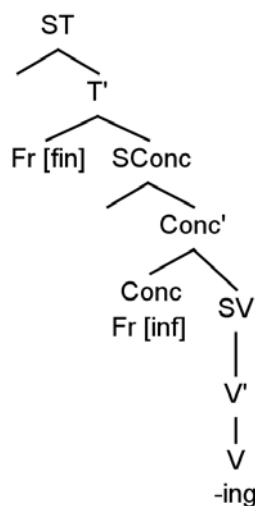


Fig. 33. Configurações de Mov-V para diferentes línguas (Silva, 2010:15).

Quanto à questão do movimento do verbo em francês, pode ser de dois tipos: (i) um movimento mais curto, para os casos em que vai aparecer no infinitivo (p.e. *ne pas manger*); (ii) movimento até ao núcleo do sintagma de tempo, para ganhar os traços de flexão relativos a T (p.e. *Jean embrasse souvent Marie*) (Fig. 30). Em português, o que distingue a colocação do ADV<sup>113</sup> é que sofre um movimento curto, semelhante ao movimento que ocorre nos verbos franceses no infinitivo (exemplo [233]). Assim, o movimento do verbo não é opcional em português, devido às restrições impostas por determinados advérbios, como os advérbios de negação, que surgem sempre antes do verbo. Por outro lado, este idioma regista também advérbios “baixos”, que surgem ligados ao sintagma. A não aquisição destas regras pelos falantes do PtgL2 vai

determinar conflito entre gramáticas, que origina a variação na colocação dos advérbios (exemplo [228]).

O estabelecimento de paralelismos com os estudos do movimento do verbo no inglês levados a cabo por Pollock (1989) e a tendência de DeGraff (1997) em generalizar as conclusões a todos os crioulos no que concerne à questão de estes representarem línguas não-marcadas tem que ser observada com alguma reserva, já que o primeiro não analisou o comportamento verbal no PE, que apresenta complexidade relativamente à posição do advérbio.<sup>114</sup> Efectivamente, apesar de os verbos principais se movimentarem nesta língua, não há evidência consistente de que se desloque sempre para FLEX (Costa, 1996:27), como o demonstram os exemplos seguintes com o advérbio “sempre”:

[232] PE:  
*O João vai sempre com a Maria ao cinema.*

[233] PE:  
– *O João sempre vai com a Maria ao cinema hoje?*  
– *Sim, sim. Ele sempre vai com ela.*

Como se pode verificar, embora o advérbio de frequência surja pós-verbalmente no exemplo [232], também pode ser colocado pré-verbalmente em interrogativas dubitativas, e em respostas às mesmas perguntas, se estas esclarecerem assertivamente a dúvida inicial (exemplos [233]). Verifica-se então que a posição do advérbio “sempre” em relação ao verbo não é regida por questões sintácticas, mas sim semânticas. De facto, no primeiro exemplo a sua semântica associa-se ao tempo verbal, enquanto no segundo caso se relaciona com o aspecto do verbo. Idênticas situações acontecem no CCV, no qual a posição do advérbio *senpri* (sempre) também varia, podendo ocorrer em vários lugares da cadeia frásica (Baptista, 2002:128):

[234] CCV:  
a. *Senpri João bebe vinhu.*  
*Sempre João bebe vinho*  
*“O João bebe sempre vinho.”*  
b. *João senpri bebe vinhu.*  
c. *João bebe senpri vinhu.*  
d. *João bebe vinhu senpri.*

Este aspecto vem evidenciar que o CCV exhibe marcação exactamente onde o postulado de Roberts (1997:12-13) aponta para que tal não aconteça.<sup>115</sup> Por outro lado, os exemplos [234a] a [234d] do CCV, bem como outros do PE (exemplos [235a] a [235i]) e do francês (exemplos [236a] a [236g]), demonstram que as posições do verbo e do advérbio não podem ser vistas nestas línguas como mero resultado de uma deslocação do primeiro para uma posição acima do segundo, a fim de preencher os requisitos de mapeamento:

[235] PE:

- a. *Frequentemente o João beija a Maria.*
- b. *O João frequentemente beija a Maria.*
- c. *O João beija frequentemente a Maria.*
- d. *O João beija a Maria frequentemente.*
- e. *Frequentemente o João tem beijado a Maria.*
- f. *O João frequentemente tem beijado a Maria.*
- g. *O João tem frequentemente beijado a Maria.*
- h. *O João tem beijado frequentemente a Maria.*
- i. *O João tem beijado a Maria frequentemente.*

[236] FRANCÊS:

- a. *Jean embrasse souvent Marie.*
- b. *Jean souvent embrasse Marie.*
- c. *Souvent Jean a embrassé Marie.*
- d. *Jean souvent a embrassé Marie.*
- e. *Jean a souvent embrassé Marie.*
- f. *Jean a embrassé souvent Marie.*
- g. *Jean a embrassé Marie souvent.*

(Kihm, 2008:423)

Tendo ainda em conta a questão dos traços do verbo, verificámos que este tanto pode aparecer marcado (*João bebe senpri vinhu*) como não-marcado (*João senpri bebe vinhu*) no CVV, em clara consonância com aquilo que sucede na sua LA. Ora, estes aspectos são também válidos para o PA (exemplos [228] e [237], estabelecendo-se, assim, um claro paralelismo entre alguns CP's, variedades reestruturadas de português e o próprio PE no que concerne à questão dos traços do verbo e seu "movimento de V-para-I":

[237] PA:

*Por acaso assistia sempre, todos os dia eu assistia televisão.*

[OSVALH1]

Considerando não só que a posição relativa dos marcadores de TMA dos crioulos, dos verbos e dos advérbios dos crioulos, línguas reestruturadas e línguas lexificadoras dependem da semântica dos advérbios e não de mecanismos sintácticos cegos (Kihm, 2008:422), a presença ou ausência do movimento do verbo para ajudar a sustentar que os crioulos são línguas sintacticamente não marcadas (Roberts, 1997:12-13) têm que ser vistas com alguma reserva, já que assentam em bases pouco consistentes. Por outro lado, a complexidade acerca do posicionamento do advérbio – e da possibilidade ou não de realização do sujeito referencial – mostra que a questão também não pode ser vista em termos de traços de marcação, uma vez que não é possível efectuar tal observação a uma escala meramente binária. Assim sendo, contempla-se o pressuposto de Muysken (1981b:415) de que línguas marcadas serão aquelas cuja interpretação conduz a uma falta de paralelismo na complexidade entre a sintaxe e a FL, que é exactamente o que sucede no caso do CCV, no PA, no PE, no crioulo do Haiti ou no francês. O que deixa em alternativa a questão de a complexidade entre a sintaxe e a FL representar o parâmetro por defeito, o qual é aplicado pela criança, a não ser que esta obtenha evidências para não o instanciar. Sob esta perspectiva, certas particularidades das línguas crioulas, como a colocação do advérbio na estrutura frásica, o uso de partículas de negação, o sistema de partículas pré-verbais de TMA, a ordem SVO, ou a ausência referencial de sujeitos nulos, têm que ser vistas, não como características de línguas não-marcadas, mas sim como instanciações inatas de parâmetros por defeito dessas mesmas propriedades, que podem originar uma série indefinida de instanciações não-defeito a adquirir pela criança, quando exposta a DLP's processáveis (Kihm, 2008:432). A este propósito, lembre-se que as formas não marcadas são as que se aplicam à língua como um todo, enquanto as formas marcadas são as que se aplicam a morfemas específicos.

No que respeita à afectação da morfologia de número no SDET, e consequente variação de concordância entre os seus elementos gramaticais, a criança mapeia dados a partir de várias possibilidades que são construídas na fala almoxarifana. Tomando como exemplo os SDET's que envolvem determinante+nome, a L1 nativizada a partir da L2 defectiva e fragmentária realiza, pelo menos nas primeiras fases da aquisição, as seguintes estruturas, ainda que o *input* não conceda evidências para a elaboração das mesmas:

1. DET<sub>[+PL]</sub> N<sub>[+PL]</sub> (exemplo [154]);
2. DET<sub>[+PL]</sub> N<sub>[-PL]</sub> (exemplo [227b]);
3. DET<sub>[-PL]</sub> N<sub>[+PL]</sub> (exemplo [199b]);
4. DET<sub>[-PL]</sub> N<sub>[-PL]</sub> (exemplo [227c]).

Posteriormente, os usos que não são consistentes com a LA poderão ir desaparecendo, uma vez que a *performance* se adequa às situações sociolinguísticas avaliadas como positivas, sendo estas determinadas, sobretudo, pela variante prestigiada. No caso concreto da comunidade de Almojarife, o prolongado isolamento restringiu o contacto com as formas prestigiadas do PE, mantendo geracionalmente a concorrência de gramáticas distintas e preservando o leque de formas flutuantes que constituem os DLP's para a aquisição/transmissão entre os níveis B-C-D (Fig. 28). Consequentemente, na fala almojarifana, as estruturas da gramática do português elitista continuam a conviver em concorrência com as estruturas típicas das situações de contacto e transmissão linguística irregular, isto é, com as formas resultantes dos processos de simplificação, transferência e reanálise. Por outras palavras: a variação registada no PA resulta da competição entre gramáticas e deve ser entendida como processos de alternância de código (*code switching processes*) (cf. Lucchesi & Ribeiro, 2009:149).

Este aspecto revela-se de primordial importância para entender a variação registada no SN da comunidade bilingue de Almojarife, uma vez que a alternância de códigos a nível do processamento inter e intrafrásico (Fig. 15) é regulada por mecanismos inerentes às situações de interacção verbal (ponto 2.3.1 do presente trabalho). De facto, alguns estudos têm mostrado que a alternância de códigos no processamento interfrásico é determinada por factores sociais e discursivos, enquanto a que atinge o processamento intrafrásico é condicionada por princípios sintácticos (Belazi, Rubin & Toribio, 1994:221). Nesta conformidade, a alternância de códigos, sobretudo a registada no interior da mesma sentença, terá implicação directa sobre o modo como o falante estrutura o seu conhecimento linguístico na altura de interagir verbalmente.

No caso concreto da alternância de códigos no interior da mesma frase, parece ocorrer “*uma tendência para predominância de uma gramática numa determinada porção da estrutura oracional e a predominância da outra gramática nos demais nódulos da*

*oração*” (Lucchesi & Ribeiro, 2009:149). A este propósito, convém lembrar igualmente que as transferências não se processam de modo caótico, antes necessitam de encontrar condições para serem processadas no sistema de desenvolvimento da L2 (Andersen, 1983c:182; Plag, 2008b:314) (ponto 2.10.2 do presente trabalho). Este aspecto alia-se aos anteriores no sentido de impedir que a fala almomarifana realize flexão contextual plena, quedando-se, como referimos no ponto 2.10.4.2, por produções que revelam apenas alguma flexão inerente e a posicionam entre os estágios T2 e T3 do desenvolvimento linguístico sugerido por Plag (2008b – Tabelas 2.1 e 2.3), os quais são caracterizados por realização de variação acentuada, mas estruturada. Assim, no plano variável do PA, será de considerar a tendência para a co-ocorrência de estruturas provenientes de uma mesma gramática numa mesma porção da estrutura da sentença nos processos de alternância de código (cf. Lucchesi & Ribeiro, 2009:149), configurando-se aqui as propostas avançadas pelo Princípio da Coesão Estrutural (Haiman, 1983:782).

## **2.12. Princípio da coesão estrutural e variação a nível do sintagma determinante**

Os casos de alternância de códigos são originados pela competição entre gramáticas alojadas na mente do falante, elaborando-se sentenças a partir das regras de uma das gramáticas e que se combinam com estruturas geradas pela outra gramática. Estas formas, apesar de reflectirem variação nos mais variados níveis da sentença, são orientadas e restringidas pelo Princípio da Coesão Estrutural (Haiman, 1983:782). Este determina, então, que as estruturas de uma mesma gramática apresentem propensão maior para co-ocorrerem nos nódulos mais coesos, isto é, naqueles que são internos a um único constituinte (p.e. número, género, pessoa), e não tanto nos nódulos mais periféricos, ou seja, nos que estabelecem uma relação de especificador/núcleo (p.e. entre um SDET sujeito e o núcleo da flexão sentencial). Assim, e no que concerne especificamente à estrutura sintagmática de um mesmo SDET, é possível prever, probabilisticamente, determinadas elaborações.

Tendo em conta, por exemplo, que a inserção de marca de género, por ocorrer primeiro (Müller, 1994a; Franseschina, 2002, 2005; Corrêa, Name & Ferrari-Netto, 2004; Montrul, 2004; White *et alii*, 2004), serve como gatilho para despoletar também a marcação de número, Lucchesi (2000a:264) constatou a aplicação do Princípio da

Coesão Estrutural (Haiman, 1983:782) no dialecto de HEL-Ba, uma vez que a marcação de género favorece categoricamente a marcação PL no constituinte analisado, enquanto a sua não-marcação restringe a inserção do número. Como nas línguas que configuram transmissão linguística irregular a concordância de número é mais afectada do que a concordância de género, é possível postular que, no PA, a regra de concordância plena de género poderá não ser inserida num SDET em que não se aplica plenamente a concordância de número (p.e. *as bananeira bem alto*), enquanto que, num SDET em que esta regra é inserida na totalidade, é improvável não acontecer concordância plena de género (p.e. *as bananeiras bem altos*). Portanto, o primeiro exemplo será comum em Almojarife, ao passo que o segundo exemplo encontra poucas possibilidades de ocorrer. Destes aspectos resulta que, se o falante inserir a marcação de número plenamente, também acontecerá, categoricamente, marcação total de género (p.e. *as bananeiras bem altas*), sendo esta igualmente possível caso não aconteça marcação plena de número (p.e. *as bananeira bem alta*).

Relativamente à questão dos nódulos que estabelecem uma relação de especificador/núcleo, abster-nos-emos de a comentar, já que cai fora do âmbito do presente trabalho, ao adequar-se mais a uma análise do tipo sintagmático, e não do género atomístico.

### **2.13. Conclusões parciais**

Face aos vários pressupostos que expusemos, verifica-se que, para se determinar se uma língua configura processo de transmissão linguística irregular, é necessário ter em consideração dois aspectos fundamentais: “*por um lado, o item da estrutura lingüística que está sendo focalizado, a sua função mais ou menos abstrata, sua transparência semântica, e o valor mais ou menos marcado do parâmetro sintático em jogo; tendo, sempre, como referência os modelos transplantados da variedade lingüística européia; e, por outro lado, a comunidade de fala em que esse item é analisado, o que implica considerar o grau de contato lingüístico por que essa comunidade passou, a sua exposição aos modelos lingüísticos difundidos institucionalmente, etc.*”<sup>116</sup> Observando-se a importância destes aspectos na determinação de uma variedade emergente em contexto de transmissão linguística irregular, estar-se-á, tal como



acontece no caso dos pidgins e crioulos, não só a perceber como a confluência de múltiplos factores, que vão desde os efeitos produzidos pelo contacto de línguas e da variação daí resultante, até ao modo como dispositivos cognitivos entram em jogo para a modelação das línguas, em geral, mas também a ter em conta que essas mesmas variantes encerram em si riquezas sociolinguísticas únicas, que necessitam de ser preservadas com recurso a planificações e políticas de línguas adequadas, em particular.

Por fim, e tendo em conta que o ponto de partida de todo o processo de transmissão linguística irregular é a não-realização, em maior ou menor grau, de morfologia flexional e das regras de concordância nominal e verbal da língua lexificadora durante a aquisição inicial da LA, verifica-se que esse é o quadro que se adequa ao PA, tal como se ajusta a variedades reestruturadas do PB. Por outro lado, como o substrato directo do PA (o santomense) configura um crioulo basilectal, marcado por irregularidade e drástica redução de complexidade morfológica, podemos assumir que, nas análises das comparações detalhadas que levaremos a cabo tendo em conta a cadeia sociohistórico-linguística estabelecida entre o PB (nomeadamente a fala do município do Rio de Janeiro, a norma culta de Salvador e o dialecto de HEL-Ba), o PA, o substrato directo deste (o santomense) e também os seus substratos ancestrais (línguas do grupo níger-congo atlântico), os conceitos de transmissão linguística irregular e criouliização/pidgnização, a par de outros, como os que giram em torno da aquisição e nativização de L1, aquisição bilingue, ASL (individual e massiva) e fossilização, ocuparão posição central para ajudarem a determinar as nossas conclusões acerca dos fenómenos que motivam a CPL-var no interior do SN do PA. Paralelamente, e tendo em conta a concepção da variação linguística dentro da estrutura gramatical, não deixaremos também de recorrer aos pressupostos inerentes ao Princípio da Coesão Estrutural (Haiman, 1983:782), observando-os no seu sentido mais fraco (o probabilístico) e mensurando-os em concomitância com os efeitos de outros factores linguísticos.

Assim sendo, no presente trabalho assumiremos, essencialmente, que a marcação PL no SN do PA depende do jogo estabelecido entre contextos linguísticos (língua-I) e extralinguísticos (língua-E), conjugando análises teóricas e quantitativas. Ainda assim, levaremos em consideração uma abordagem ecléctica, que tenha em consideração os pressupostos das diferentes metodologias propostas no presente capítulo. Face ao que se

referiu acerca da ASL e aquisição/nativização de L1's, a orientação do nosso trabalho inferirá que a GU está, de uma forma ou de outra, disponível para os adultos que aprendem uma L2, já que mantém os seus princípios activados. Paralelamente, teremos em consideração que, com a aquisição intermédia de morfologia flexional relevante e de itens lexicais, é possível restabelecer parâmetros para se atingir um estágio final de aquisição. Contudo, a variação que surge fossilizada em fase tardia pode também ser entendida como falha no restabelecimento de alguns parâmetros (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clahsen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), visto as categorias funcionais deixarem de estar disponibilizadas depois do período crítico do desenvolvimento. Baxter (2004, 2009), em constatações acerca da CPL-var do PT, defende que as estruturas das L1 africanas exercem, conjuntamente com as do crioulo, importante influência no ponto de entrada do item com morfema de PL que leva à configuração do SN PL do PtgL1, adquirido em situação de transmissão linguística irregular. Figueiredo (2008) avançou a hipótese de a CPL-var do PA, tal como acontece com a CPL-var do PT, estar sujeita a condicionantes que ancoram na aquisição defectiva do PtgL2 e PtgL1, em situações de contacto que envolvem estruturas de PtgL2 influenciadas pela línguas do grupo níger-congo atlântico. Como tal, para o presente trabalho, partiremos igualmente do pressuposto que a aquisição da L1 se inicia no estado FL0, enquanto que o estágio da Gramática *n* é o estado inicial para a ASL por parte de aprendentes em idade pós-puberdade. Contudo, a variação da interlíngua, até ao seu estágio final de sistematicidade, não será resultado exclusivo da refixação de parâmetros guiada pela L1, uma vez que é também consequência da interferência de outros factores, sejam eles cognitivos (maturidade, quantidade e qualidade do *input*, frequência de uso da língua ou conhecimento de outras línguas) sejam eles extralinguísticos (motivação, idade, sexo, escolaridade, aspectos culturais, contexto sociohistórico da aquisição, etc.).

No próximo capítulo daremos conta da metodologia a que recorreremos para constituição dos nossos dados, considerando quer factores já abordados em outros trabalhos, e que considerámos terem influência na marcação PL, quer ainda factores por nós constituídos, os quais nos permitirão ajudar a observar, minuciosamente, quais os elementos que motivam a pluralização no SN do PA. Com base em todos estes factores,

levantaremos hipóteses que poderão conceder não só uma visão mais abrangente sobre a CPL-var no SN do PA mas também ajudar a perceber quais as condicionantes que a determinam. As hipóteses serão posteriormente confirmadas, ou não, nos capítulos 4 e 5.

## CAPÍTULO 3

*“Não ande apenas pelo caminho já percorrido, pois ele conduz somente até onde outros estiveram antes.”*

(Alexandre Graham Bell)

### A Metodologia

No presente capítulo iremos descrever: (i) a metodologia aplicada na recolha dos dados que constituem o *corpus* do presente trabalho; (ii) a variável dependente e os grupos de factores independentes que constituímos para efectuar a análise do fenómeno da CPL-var no SN do PA; (iii) os procedimentos adoptados para se codificarem os dados; (iv) o suporte computacional utilizado para tratamento dos dados e sua análise estatística; (v) a caracterização dos informantes entrevistados para constituição do *corpus*.

#### 3.1. Aplicação dos métodos quantitativos no estudo da aquisição de línguas

A variação linguística produzida por falantes de determinadas comunidades tem sido observada com recurso a métodos matemáticos, nomeadamente aquele que permite analisar a variação segundo o paradigma de Labov (1972a), isto é, o método conhecido como “*the variable rule analysis*” (VARBRUL). Especialmente usado no estudo da ASL, seja em contexto informal (L. Dickerson, 1974; W. Dickerson, 1976; Adamson & Kovak, 1981; Adamson, 1988; Wolfram, 1989; Young, 1991; Young & Bayley, 1996; etc.) seja em contexto formal de ensino (Godinho, 2005, 2008; Martins, 2007; Silva, 2010), o pacote de programas VARBRUL, em virtude de calcular as probabilidades dos factores linguísticos e extralinguísticos que motivam a variação linguística, tem-se revelado um precioso auxiliar no estudo da aquisição. Paralelamente, a ele recorreram também pesquisadores que analisaram a variação da fala quer de comunidades linguísticas urbanas (Cedergren, 1973; Sankoff & Sankoff, 1973; Naro & Lemle, 1976; Braga, 1977; Lemle & Naro, 1977; Poplack, 1980a; Guy, 1981a; Scherre, 1988; Fernandes, 1996; L. Andrade, 2003; Lopes, 2000; Jon-And, 2008, 2009) quer de grupos rurais isolados ou semi-isolados (Emmerich, 1984, 1992; Baxter, 1992, 2004, 2009; Lucchesi & Macedo, 1997; Lucchesi, 2000a; P. Andrade, 2003; Tieppo, 2003; Figueiredo, 2008, 2009a, 2009b).

Trabalhando com uma variável dependente dotada de dois ou mais valores representativos do fenómeno em análise (no nosso caso, a ausência ou marcação de número PL no SN) e um sistema de múltiplas variáveis linguísticas e extralinguísticas (as variáveis independentes), que incluem diversos factores relacionados com contextos linguísticos e sociais da comunidade em estudo, o programa VARBRUL combina relações entre a variável dependente e as variáveis independentes, a fim de determinar o modo como os factores das segundas motivam o fenómeno em análise e representado pela primeira variável. Para o programa poder realizar as combinações, deverão ser retirados os factores ou variáveis independentes que ele não considere estatisticamente relevantes, pois só assim se obterão valores fiáveis tanto para a frequência de uso conotada a cada um dos valores dos grupos de factores independentes como para o peso que cada um dos factores exerce na motivação do fenómeno em estudo. Os resultados, fornecidos, em pesos probabilísticos, quantificam a interferência exercida por cada variável independente na aplicação da regra em estudo, revelando, deste modo, quais são os grupos de factores que contribuem ou não para a variação (Preston, 1996b:239). Como tal, o grau de fiabilidade é superior aos dos valores representados através de meras percentagens, já que estes não permitem concluir se a variação em observação é sistemática. Paralelamente, o recurso às variáveis extralinguísticas deixa observar, diacronicamente, quando e como se incorpora, aplica e desenvolve determinada regra durante o processo de aquisição de uma língua.

A multiplicidade de recursos fornecida pela ferramenta VARBRUL faz dela um dos instrumentos ideais para se observar e explicar a sistematização da variação, daí termos recorrido a tal suporte computacional para efectuarmos a análise da CPL-var detectada nos SN's da fala da comunidade de Almojarife. Para Preston (1996b:253), todos os falantes de uma determinada L1 constituem a comunidade linguística desse dialecto, uma *sibling specie* de outras línguas irmãs que compartilham a mesma estrutura, mas que se desenvolvem independentemente (cf. Croft, 2000:13-15). Com efeito, esta pode possuir subgrupos distintos, filogeneticamente ligados entre si, já que compartilham não só os mesmos traços distintivos dessa mesma L1 mas também características sociais comuns a todos esses subgrupos. Deste modo, com recurso ao VARBRUL, é possível analisar os factos linguísticos presentes na produção linguística real dos falantes do PA,

uma variedade filogenética do PE, e tentar explicar a variação linguística patente nos seus SN's, relacionando-a com factores de natureza linguística e social. Simultaneamente, e numa perspectiva da aquisição, poder-se-á observar a integração e desenvolvimento de aspectos particulares não só da LA na fala da comunidade mas também dos fenómenos conotados aos substratos que poderão estar a motivar a variação em estudo. Assim sendo, a nossa pesquisa assentou nas seguintes etapas de trabalho: observação da comunidade linguística em causa, escolha criteriosa dos informantes, captação de dados, levantamento de hipóteses de trabalho e análise e interpretação dos dados captados. Como base de partida para a nossa pesquisa acerca da CPL-var no SN tivemos em linha de conta os trabalhos pioneiros de Naro & Lemle (1976) e Lemle & Naro (1977), para o PB. Em adição, e dado que neste momento é vasta a análise sobre a CPL-var, sobretudo no PB, adoptámos também os parâmetros e métodos subjacentes a outras investigações, nomeadamente os seguidos por Braga (1977), para o português mineiro do Rio de Janeiro, Guy (1981a) e Scherre (1988), para a fala do município do Rio de Janeiro, Lopes (2000), para a NURC, P. Andrade (2003), para o dialecto de HEL-Ba, Figueiredo (2008, 2009, 2009a), para o PA, Jon-And (2008) para o PMp, Baxter (2009), para o PT e dialecto de HEL-Ba, e Jon-And (2009), para o PCV. Nas análises por nós efectuadas são estabelecidos bastantes paralelos de comparação com todos estes trabalhos,<sup>117</sup> especialmente os respeitantes às variantes africanas e brasileiras de português. Porém, outros trabalhos foram ainda levados em linha de conta, a fim de observarmos a variação no SN, nomeadamente os produzidos para o estudo de variantes de português da região sul do Brasil (Fernandes, 1996; L. Andrade, 2003) ou falas caboclas brasileiras (Tieppo, 2003). Por último, também os recentes estudos sobre a CPL-var nos SN's produzidos por aprendentes chineses do PtgL2, em contexto formal de ensino, foram por nós considerados na presente análise (Godinho, 2005, 2008, 2009; Silva, 2010).

Não obstante, a explicação dos resultados do nosso estudo não se limitará à óptica comparada, uma vez que não pode ser desconsiderada a possibilidade de a CPL-var ter como motivações os princípios universais subjacentes à aquisição das línguas. De facto, sendo o PA fruto de transmissão linguística irregular e tendo como substrato o santomense, isto é, um CP resultante da aquisição defectiva de PtgL2, os referidos

resultados terão também que ser observados numa perspectiva psico-cognitiva e generativista, nomeadamente no que concerne à aquisição das categorias funcionais por falantes adultos do PtgL2 e, mais concretamente, no que diz respeito à captação dos traços de número nos elementos do SN.

Assim sendo, o modelo teórico construído para a análise quantitativa da CPL-var nos SN's dos nossos dados segue a proposta dos estudos de Guy (1981a), Scherre (1988), Lopes (2001), P. Andrade (2003), Baxter (2004, 2009), Godinho (2005) e Jon-And (2008, 2009). Por seu lado, a análise e interpretação dos resultados assenta não só nos mesmos estudos mas também em diversos trabalhos das áreas psico-cognitiva e generativista sobre aquisição de L1 e L2, aquisição monolíngue e bilingue, fossilizações e aquisição dos traços de número nos elementos do SN do PE e do PB, referidos no capítulo 2. De acordo com o modelo teórico construído e tendo como suporte para análise dos resultados os trabalhos mencionados, procuraremos, com recurso a dados linguísticos reais, fornecer o nosso contributo para acrescentar algo mais aos conhecimentos já constatados acerca dos dispositivos inerentes aos processos de ASL por adultos em situação de contacto massivo, bem como de L1's adquiridas por transmissão linguística irregular em contexto de bilinguismo.

### **3.1.1. Recolha e transcrição dos dados do *corpus***

A abordagem meramente formalista, que determinou os estudos em linguística até às propostas labovianas, não possibilitava que as observações de idiomas levassem em consideração outros factores que não fossem os intrínsecos à própria língua. No entanto, as línguas evidenciam variação lexical, sintáctica e fonética relativamente ao que as gramáticas tradicionais, detentoras da norma-padrão, entendem como correcto, sendo alguns fenómenos da variação determinados pelas relações que a língua estabelece com fenómenos externos. Por outro lado, embora as formas orais não contempladas pela norma (e pelo registo escrito) apontem, bastantes vezes, para mudanças em curso, não significa que estas tenham que, necessariamente, ser de imediato aceites pelos organismos reguladores da língua, como as instituições de ensino. Assim, compete aos especialistas em linguística determinarem se as variações detectadas apontam, de facto, no sentido da mudança. Para se conseguir tal, é necessário comparar diferentes estádios

de uma língua, com recurso a dados linguísticos. Ora, para a maior parte do tempo de vida do português, tal acesso só era possível observando textos escritos, que na maioria das vezes não correspondem ao registo oral espontâneo. Assim sendo, estes não constituirão os *corpora* ideais para determinado tipo de investigação científica em Linguística (Histórica ou Diacrónica), mormente para identificação de fenómenos que se prendem com as mudanças decorrentes do contacto entre línguas. Todavia, hoje em dia é possível recorrer a *corpora* orais para se comparar a forma de utilização da mesma com o que a gramática tradicional prescreve ou aquilo que a norma padrão considera correcto, detectando-se e observando-se, assim, não só as mudanças em curso mas também o sentido para o qual estas se direccionam.

Ellis (1994:144) e Romaine (2003:411) defendem que a variação deve ser encarada sob dois prismas: o livre e o sistemático. O processo é desencadeado por um reduzido número de falantes relativamente a um insignificante número de situações linguísticas, pelo que a nova forma que entra na *interlíngua* apresenta uma variação não sistemática, que não é limitada por factores observáveis e ocorre em período de tempo limitado. Contudo, o fenómeno vai-se alastrando de modo progressivo, até se generalizar e atingir toda a comunidade linguística, tornando-se sistemático. Iniciando-se por razões linguísticas, psicológicas e sociológicas, a variação tem a agir por detrás dela diversas motivações, pelo que poderão actuar, no seu condicionamento, quer fenómenos como a diversidade geográfica, a tensão entre arcaísmos e neologismos, as deficiências reveladas ao nível da aprendizagem das línguas ou a movimentação em direcção à aquisição de prestígio social, quer fenómenos decorrentes do contacto entre línguas, com os falantes buscando a eficácia do seu sistema comunicativo através da simplificação do mesmo.

Relativamente ao contacto entre línguas, este decreta a emergência de uma LA que, geracionalmente, pode mesmo levar ao desaparecimento da L1 do falante, por vezes em espaço de tempo abrupto.<sup>118</sup> Trata-se, pois, de um fenómeno de mudança que tem origem em influências exógenas ao sistema linguístico, mas que o afecta na totalidade, já que ele necessita de se estabilizar para corresponder eficazmente aos anseios e necessidades comunicativas dos seus falantes. Como tal, enquanto no fenómeno de mudança que afecta os subsistemas da língua os falantes desta se mantêm monolíngues,



no caso da mudança por contacto, os falantes da “comunidade que muda” (*shifting population*) tendem a ser bilingues, mas usando uma L2 que manifesta variações decorrentes das interferências da L1 ou das mudanças inerentes à própria variedade agora falada. As interferências da L1 enraízam em várias questões, dependendo, por exemplo, de aspectos como a maior ou menor dependência política, económica ou social da “comunidade que muda” face à comunidade da LA.

Como se mencionou no capítulo anterior, a aquisição de uma nova língua é também caracterizada pela criação de “terceiras hipóteses”, transitórias ou que se cristalizam, de funcionamento de determinados subsistemas da LA, não só por influência da L1 mas também por reanálises originais da gramática da LA em fase de apropriação. Este processo de aquisição é descritível em termos de *interlínguas*, isto é, de sucessivas fases de aproximação à LA. Por outro lado, estas evidências da L2 também poderão não reflectir nenhuma das gramáticas das línguas em contacto que originaram a nova língua, mas corresponder simplesmente a idiosincrasias do falar bilingue com base nos universais linguísticos, afinal em tudo semelhantes aos das L1’s adquiridas em situação que não configura transmissão linguística irregular.

Atentando aos aspectos referidos, facilmente se infere que a Linguística não se pode reduzir apenas ao estudo das manifestações contemporâneas dos dialectos a que se convencionou classificar de normas. E se tivermos em conta que, para o linguista, os conceitos de dialecto e norma não são opostos,<sup>119</sup> então, e embora a Dialectologia tenha grandes tradições em Portugal e mantenha relações estreitas com a Linguística Histórica, se compreenderá a necessidade de direccionar também os estudos linguísticos para os fenómenos de mudança decorrentes do contacto. De facto, e embora este seja um fenómeno cujo processo é do mesmo tipo daquele que é inerente à própria língua, tem origem diferente e, conseqüentemente, desfechos diferentes.

Nesta perspectiva, impõe-se o estudo de dados linguísticos numa óptica sociolinguística, o que “*é normalmente entendido como buscar razões, no contexto social, para os diversos usos linguísticos.*” (Lopes, 2001:105). No nosso caso concreto, os dados que constituem o *corpus* da presente análise foram recolhidos com recurso a 18 entrevistas do tipo sociolinguístico-laboviano (Labov, 1972a, 1983), entre Fevereiro e Maio de 1998, no âmbito do projecto “*Semi-creolization: testing the hypothesis against*

*data from Portuguese-derived languages of São Tomé (Africa)*”, financiado pelo Australian Reserch Council e dirigido por Alan N. Baxter, com a colaboração de Dante Lucchesi. As gravações das entrevistas decorreram sob a responsabilidade do segundo e, tendo em vista a natureza das variáveis linguísticas e extralinguísticas a constituir para observação dos dados linguísticos produzidos, cada uma das conversas teve a duração de cerca de uma hora (Tabela 8.1).

Finda a recolha das entrevistas, as mesmas foram transcritas ainda em Almojarife por Dante Lucchesi. Nessa tarefa, colaborou o assistente de campo local, Francisco Paulino, que forneceu informações de valor extremo para a interpretação das produções de fala. Posteriormente, os dados auditivos e as transcrições foram entregues ao autor do presente trabalho, que recopiou os primeiros para formato digital, reviu totalmente as segundas, procedeu à sua reorganização e colectou os SN's para análise.

Para a transcrição dos dados, foi seguida a linha metodológica da representação autêntica das realizações linguísticas proposta por Poplack (1989), dando-se conta de aspectos como os falsos inícios de discurso, as palavras interrompidas, as hesitações discursivas, as intervenções paralinguísticas, a reformulação de sentenças ou a pronúncia de palavras com características dialectais distintas da forma padronizada. Para tanto, recorreu-se à ortografia estandardizada, em detrimento da transcrição fonética (Tagliamonte, 2006:55-56), uma vez que a primeira permite uma melhor visualização do texto e, conseqüentemente, dos fenómenos que se pretendem analisar (cf. Projecto Vertentes do Português Rural do Estado da Bahia).<sup>120</sup> Contudo, a transcrição ortográfica procura reproduzir, o mais fielmente possível, as características da produção oral.

Atribuiu-se também um código identificativo a cada um dos informantes (Tabela 8.1), que surge entre parêntesis rectos após os exemplos apresentados durante o presente trabalho. Seguidamente, exemplifica-se com um pequeno excerto do diálogo entre o documentalista (DOC) e a informante (INF) [MANEVEM2], a aplicação de algumas das convenções ortográficas adoptadas na transcrição da totalidade dos registos áudio do nosso *corpus* (cf. Tagliamonte, 2006:56-64).<sup>121</sup> No referido diálogo, os SN's por nós coligidos encontram-se grafados a negrito:

DOC: *Seora como se chama?*  
 INF: *Ma— chama Maria Pedro—**das**—Neve.*

DOC: *Hum...*  
 INF: *Quê nome co a dá... a lá é? Quê nome co a dá é? Maria—Pedro—**das**—Neve. Nã Stáva... nã Stáva. Mê nome em casa é Stáva. Mê batisado é Maria—Pedro—**das**—Neve.*

DOC: *Maria—Pedro—das—Neves?*  
 INF: *É. Maria—Pedro—**das**—Neve.*

DOC: *E nome casa é Stava?*  
 INF: *É. É Stáva.*

DOC: *Aqui as pessoas tem dois nomes normalmente, né?*  
 INF: *É. Sempre gen São Tomé tem **duas** nome. Tê nome pa chamá solteiro e casá... agora tem nome quê naturá. Naturá é quê Maria—**das**—Neve, que Stáva chamá só a toa, casa.*

DOC: *E porque derão esse nome Stava?*  
 INF: *É porque mia... mia velha é quê pôso Stáva. Que mia mãe nã pariu... pariu ho— — macho. Mas três, quato pariu homé homé homé, **quato macho**. Depois é veio de trás, quê [ININT] onde quê stáva? [ININT] stáva.*

DOC: *Hum...*  
 INF: *Onde é quê stáva?*

DOC: *Hum... agora entendi. Antão era a— porque perguntaram onde é que estava?*  
 INF: *É. Onde quê stáva quê nã vi a muito tempo, que homé veio quato, pô veio agora é quinto.*

DOC: *E o marido da senhora?*  
 INF: *Mê marido nã stá.*

DOC: *Hum... ele— como ele chama?*  
 INF: *Mê marido?*

DOC: *Hum.*  
 INF: *Chama Lázaro.*

Como se pode confirmar, no presente excerto foram utilizadas as seguintes convenções ortográficas:

- (i) Falsos começos representados com um hífen:
  - (1) *ele— como ele chama?* [DOC]
- (ii) Pausas audíveis sinalizadas com reticências:
  - (2) *É porque mia... mia velha* [MANEVEM2]
- (iii) Palavras parciais marcadas com duplo hífen:
  - (3) *pariu ho—— macho.* [MANEVEM2]

(iv) Nomes de pessoas, lugares, músicas, jogos, etc., que representam ocorrências lexicalizadas, grafados com hífen entre os seus elementos:

(4) chama Maria–Pedro–das–Neve. [MANEVEM2]

(v) Partes incompreensíveis do discurso indicadas com recurso a parêntesis:

(5) *Depois é veio de trás, quê (ININT) onde quê stáva?* [MANEVEM2]

(vi) Intervenções paralinguísticas transcritas com recurso à ortografia padronizada:

(6) Hum... agora entendi. [DOC]

(vii) Fenómenos linguísticos que configuram formas dialectais das palavras padronizadas ou das formas fonológicas *standard*, grafados com ortografia padrão e de modo a reproduzirem o enunciado real:

(7) Mê nome um casa é Stáva [MANEVEM2]  
“O meu nome em casa é Estava”

### 3.1.2. Tratamento dos dados

O dados colectados para análise foram codificados de acordo com as categorias e subcategorias morfossintácticas seleccionadas como variáveis-alvo para o nosso estudo, as quais se encontram descritas no ponto 3.3. Para possibilitar a análise dos dados áudio-transcritos e comparar os resultados obtidos com os de outros estudos levados a cabo para diferentes variedades reestruturadas de português, foi necessário submeter as unidades e categorias gramaticais a um controlo de operacionalidade descritiva, de acordo com o conjunto de noções e modelos teóricos já avançados por outros linguistas, nomeadamente em análises efectuadas para o PVB (Braga, 1977; Guy, 1981a; Scherre, 1988; Lopes, 2001; P. Andrade, 2003), para o PT (Baxter, 2004; 2009), para o PVA (Inverno, 2005), para o PMp (Jon-And, 2008),<sup>122</sup> para o PCV (Jon-And, 2009)<sup>123</sup> e para o próprio PA (Figueiredo, 2008, 2009a, 2009b). Estes passos do nosso trabalho encontram-se tratados nos capítulos 4 e 5.

### 3.1.3. Os dados dos informantes

O propósito inicial das entrevistas foi de gravar um *corpus* estratificado por sexo e idade, tendo os factores escolarização e estadia fora da comunidade sido também

contemplados, em virtude de se terem mostrado relevantes em diversos estudos de sociolinguística variacionista levados a cabo no Brasil. Assim sendo, entrevistaram-se informantes de ambos os sexos, repartidos por três faixas etárias distintas<sup>124</sup> e nascidos em Almojarife ou arredores, mas a residir nesta comunidade há longo tempo. Os informantes exercem ou exerceram também variados tipos de profissão, sendo a sua escolarização diversificada, indo desde a ausência total da mesma até distintos graus de aprendizagem. Com tal metodologia, tentámos evitar que os dados confluíssem na direcção de favorecer determinadas características linguísticas ou sociais.

No final de cada entrevista, a ficha de identificação do informante contemplava, pelo menos, os seguintes aspectos: (1) nome; (2) sexo; (3) idade; (4) línguas faladas; (5) local de nascimento; (6) profissão; (7) estado civil; (8) nível de escolaridade; (9) estadia fora da comunidade.

Refira-se ainda que todos os informantes são bilingues, uma vez que falam o santomense e o português reestruturado de aquisição recente, cuja variação no SN constitui o objecto do presente estudo. Como referimos, se aliarmos este aspecto ao facto de as características sociais dos entrevistados serem heterogéneas, entendemos como é grande o grau de variação que nos propomos analisar. A Tabela 8.1 apresenta os dados sociais dos informantes do PA.

Após efectuada a recolha e tratamento dos dados, interessa estabelecer o objecto de estudo e as etapas de trabalho que permitirão a sua análise. Para tanto, seguimos a metodologia adoptada por Godinho:

- “(i) *A determinação do fenómeno objecto de estudo, a concordância de plural no sintagma nominal (a variável dependente);*
- (ii) A determinação das variáveis independentes – os grupos de factores que estão relacionados com o fenómeno em estudo;*
- (iii) A codificação, que constitui a etapa preparatória para utilização do modelo linguístico;*
- (iv) A quantificação dos dados;*
- (v) A interpretação dos dados.” (Godinho, 2005:242).*

**Tabela 8.1.** Informantes de Almoxarife: faixas etárias e seus aspectos sociais.

Nr.	Código do informante	Escalão etário	Sexo	Ensino	Estadia (Permanência fora da comunidade)	Duração da entrevista
01	[OSVALH1]	FE1 (20-40)	Masc.	4ª. Classe	Não	01:00:28
02	[ZECAH1]	FE1 (20-40)	Masc.	4ª. Classe	Não	01:10:00
03	[CASTEH1]	FE1 (20-40)	Masc.	8º. Ano	Não	00:59:14
04	[ANTOM1]	FE1 (20-40)	Fem.	8º. Ano	Não	01:04:45
05	[DULCEM1]	FE1 (20-40)	Fem.	3ª. Classe	Não	00:56:20
06	[CARMOM1]	FE1 (20-40)	Fem.	8º. Ano	Não	01:06:19
07	[ABILH2]	FE2 (41-60)	Masc.	4ª. Classe	Sim (Gabão e Libreville)	01:08:59
08	[LUIH2]	FE2 (41-60)	Masc.	Não	Sim (Príncipe)	01:25:57
09	[MAURIH2]	FE2 (41-60)	Masc.	3ª. Classe	Sim (Libreville)	01:03:28
10	[CLOTIM2]	FE2 (41-60)	Fem.	4ª. Classe	Não	01:02:36
11	[FRANCM2]	FE2 (41-60)	Fem.	Não	Não	00:56:28
12	[MANEVEM2]	FE2 (41-60)	Fem.	Não	Sim (Gabão)	01:01:10
13	[CELESH3]	FE3 (+60)	Masc.	3ª. Classe	Não	01:04:44
14	[MANOH3]	FE3 (+60)	Masc.	Não	Não	01:02:55
15	[SALVH3]	FE3 (+60)	Masc.	Não	Não	01:10:53
16	[PRINCEM3]	FE3 (+60)	Fem.	Não	Não	01:15:05
17	[MMDEUSM3]	FE3 (+60)	Fem.	Não	Sim (Angola)	01:11:51
18	[MINISM3]	FE3 (+60)	Fem.	1ª. Classe	Não	01:08:16

**Obs.:** Os nomes dos informantes não são fornecidos para salvaguarda da identidade dos mesmos.

### 3.1.4. Caracterização dos dados

Definidas as etapas do nosso estudo, começámos por colectar todos os SN's de número PL constantes nas entrevistas e procedeu-se à codificação dos dados do *corpus* que configuram SN's pluralizáveis, constituídos por pelo menos dois elementos. Refira-se que para o modelo de codificação não foram tidos em conta os SN's que apenas apresentam itens singulares (p.e. DET<sub>[-PL]</sub> N<sub>[-PL]</sub> – ponto 2.11.3.3 do presente trabalho), ainda que algum deles devesse representar o PL. Tal opção fica a dever-se ao facto de, por norma, a gramática tradicional não considerar “*como desvio a não-marcação de*

*plural nos dois últimos itens*” (P. Andrade, 2003:88). Assim sendo, SN’s como o constante nos exemplos [227c] ou [238] foram excluídos da nossa codificação:

[238] PA:  
*era forte e **branco tudo** gostava de mi* [CELESH3]  
“era forte e todos os brancos gostavam de mim”

Contrariamente, foram levados em conta, nas nossas codificações, os elementos que exibem plurais mas não são marcados na norma padrão, uma vez que tal aspecto é entendido pelos gramáticos como desvio. Portanto, itens como o numeral referido no exemplo [239] foram devidamente codificados, a fim de possibilitarem o seu estudo:

[239] PA:  
*Que morde-mo pode chegá **uns cincos abelha*** [DULCEM1]

O levantamento dos SN’s de número PL coloca ainda outras questões pertinentes, como a de eliminar SN’s que não se constituem como elementos válidos para estudos de variação linguística. Assim, foi preciso fazer valer o princípio da *Lex Parsimoniae* (Lei da Parcimónia)<sup>125</sup> em itens do tipo “*Palácio dos Congresso*” [CASTEH1], isto é, nomes próprios que funcionam como itens lexicalizados na forma PL. De facto, nestes casos a estrutura de número PL não é gerada por qualquer dos processos sintácticos que originam os plurais comuns da língua portuguesa. Igual critério tivemos em relação, por exemplo, aos SN’s que configuram locuções adverbiais envolvendo o nome “vez”, como acontece com “às vezes”, “umas vezes” ou “algumas vezes”, ainda que as mesmas registem um elevadíssimo e diversificado número de realizações no nosso *corpus*, como seguidamente se exemplifica:

[240] PA:  
*são coisa que às VEZE são coesa acontecida* [CASTEH1]

[241] PA:  
*às VEZ gen vê pexe* [OSVALH1]

[242] PA:  
*às VEZES eu vô pá bar trabalho* [CARMOM1]

[243] PA:  
*à VEZE também* [CARMOM1]

[244] PA:  
*Só é sozinho, à VEZ dois* [SALVH3]

Guy (1981a:124), ao analisar o efeito da *saliência fónica* na marcação PL, estabeleceu uma escala de saliências. Reconhecendo que ocorrências do tipo de “às vezes” podem ferir resultados finais, destinou dois factores para o elemento “vez”: um para a expressão “às vezes” e outro para as demais ocorrências em que intervém o referido item. De facto, deve ser levado em conta que a selecção de ocorrências que configurem variáveis linguísticas terá de assentar na robustez que estas possam conferir à análise do fenómeno da variação, e não apenas no critério da frequência. No caso presente, as variações à forma lexicalizada são de facto mínimas quando comparadas com a realização padrão “às vezes”, o que afecta de forma viciosa os resultados finais acerca da marcação do número.<sup>126</sup> Além do mais, é também frequente a produção de repetições sistemáticas após uma primeira realização, configurando o que se convencionou classificar de falsos inícios de discurso:

[245] PA:  
DOC: *Tem muita gente que passa mal na canoa, né?*  
INF: *Sim, várias VEZES, várias VEZES.* [DULCEM1]

Como se depreende, a codificação plena deste tipo de enunciados e sua posterior inserção nos dados para análise iria produzir um número desproporcionado de instâncias com o mesmo tipo de realização, acabando por viciar os resultados finais (Tagliamonte, 2006:94). Assim o entendeu também Scherre (1988:75), que “*desconsiderou*” igualmente a expressão “às vezes” nas suas análises.

No que diz respeito ainda à eliminação de repetições, o efeito da Lei da Parcimónia estendeu-se também às frases do seguinte tipo:

[246] PA:  
*Esses peixe piquinino piquinino piquinino.* [ZECAH1]

Depois de levantados todos os dados do *corpus relevantes para o nosso estudo*, contabilizaram-se 2.340 itens passíveis de marcação PL.



### 3.2. O suporte computacional

A análise da variação linguística inerente a determinada comunidade só é possível com recurso a metodologias que permitam quantificar as possibilidades que determinem essa variação, estando as mesmas dependentes dos contextos estruturais e das situações em que a língua é usada (Lucchesi, 2000a:146).

O uso da língua pode revelar variações em função de fenómenos como as limitações da memória, as distrações e até o registo específico de cada falante, isto é, o *idiolecto*. Contudo, este tipo de variação esporádica não se presta aos estudos em sociolinguística variacionista preconizados por Labov (1969, 1972a), os quais se debruçam sobre factores linguísticos e extralinguísticos que determinam um padrão de variação sistemático (Paolillo, 2002:2). Deste modo, foi com vista a observar este que Labov (1969) propôs a noção de regra variável, que surge condicionada pelo peso que cada factor, linguístico ou social, exerce sobre a aplicação da regra. Para tanto, Cedergren & Sankoff (1974) desenvolveram um modelo de regressão logarítmica múltipla, num pacote estatístico denominado VARBRUL, que deixa ver com precisão a influência que cada factor exerce sobre o uso da regra. Posteriormente, Pintzuk (1988) adaptou o programa ao uso em PC's.

Visto que permite não só trabalhar com dados quantificados e apreciar a distribuição geral da variável em estudo, mas também calcular o peso que cada factor exerce sobre a regra da aplicação da concordância, esta ferramenta tem sido frequentemente utilizada por linguistas em diversos estudos da CPL-var de variedades brasileiras (Braga, 1977; Guy, 1981a; Scherre, 1988; Fernandes, 1996; Lopes, 2001; L. Andrade, 2003; P. Andrade, 2003; Tieppo, 2003; Baxter, 2009) ou africanas de português, como o PT (Baxter, 2004, 2009), o PA (Figueiredo, 2008, 2009a, 2009b), o PMp (Jon-And, 2008) ou o PCV (Jon-And, 2009), dialectos que forneceram ou estão a fornecer resultados que poderão ser comparados com os do presente trabalho. A opção por este tipo de ferramenta tem a ver com o facto de ela conceder a possibilidade de se avaliar a quantidade ou a força da relação entre os elementos submetidos à análise, já que esta é representada por valores incorporados numa fórmula que, por seu lado, permite prever o que pode acontecer num determinado conjunto de situações. No que concerne à operacionalidade e funcionalidade do VARBRUL, interessa referir que as mesmas já

foram exaustivamente descritas quer na literatura da especialidade (p.e. Guy & Zilles, 2007) quer em trabalhos anteriores que abordam a questão da concordância variável (Scherre, 1988; Lopes, 2001; P. Andrade, 2003; Godinho, 2005), pelo que remetemos para os mesmos as explicações detalhadas sobre aquelas. Poder-se-á, apesar de tudo, avançar, de modo sintético, que o pacote de programas VARBRUL “*combina relações complexas entre variáveis múltiplas, e a variável em análise – a variável dependente – pode ter dois ou mais valores*” (Godinho, 2005:66), por forma a conceder “*valores percentuais relativos às frequências dos diversos fatores estabelecidos para cada variável, além de fornecer o peso relativo de cada um dos fatores, informando sua relevância no comportamento do fenómeno sob estudo*” (P. Andrade, 2003:89). Relativamente ao “*ponto de vista da aquisição, o método permite aferir, com uma relativa margem de precisão, a tendência para o desenvolvimento, incorporação e aplicação de uma determinada regra gramatical ao longo do processo de aquisição de uma L2*” (Godinho, 2005:68). Cabe aqui referir, ainda, que as variáveis linguísticas submetidas a análise podem ser de categoria fonológica, morfológica, sintáctica ou lexical, constituindo grupos de factores compostos por variantes (os valores possíveis das referidas variáveis) nas quais pode ocorrer, ou não, variação. Deste modo, o potencial fornecido pela ferramenta VARBRUL, associado a constatações por parte do pesquisador sobre o uso da linguagem em múltiplos contextos, permite ir para lá da fronteira do mero juízo da “gramaticalidade” ou “agramaticalidade” (Paolillo, 2002), isto é, deixa antever se o uso da variação é sistemático, no sentido de configurar mudança em curso.

Não obstante a fiabilidade do pacote de programas VARBRUL na previsão dos efeitos dos contextos sobre a variável dependente em observação e dos padrões de variação da comunidade linguística em estudo, o entendimento dos resultados está directamente relacionado com a quantidade de dados submetidos a análise. De facto, o método matemático com que a ferramenta trabalha determina que, quanto mais numerosos e sólidos forem os dados, mais rigorosa se torna a fiabilidade dos resultados finais (Young & Bayley, 1996:301). Para além destes aspectos, a ferramenta permite ainda que o seu utilizador modifique hipóteses iniciais e reanalise os dados de modo fácil (Young & Bayley, 1996; Paolillo, 2002).

A análise da variação linguística com recurso ao VARBRUL implícita, por parte do pesquisador, uma progressão metodológica que obedeça às seguintes etapas:

“(1) A colocação das hipóteses de natureza linguística e social. ... na colocação das hipóteses é preferível ser mais liberal do que conservador, uma vez que se torna mais fácil retirar factores ou grupos de factores que revelaram não afectar a variação na interlíngua do que incluir factores já depois de codificados os dados (Young & Bayley, 1996:57); (2) a codificação dos dados; (3) a verificação da codificação; (4) a recodificação de acordo com os primeiros resultados; (5) a avaliação dos factores e grupos de factores significativos para a análise e (6) a interpretação dos dados de acordo com as hipóteses colocadas” (Godinho, 2005:230).

Tecnicamente, o VARBRUL fornece a possibilidade de se efectuarem análises de uma variável dependente que tem valores nominais, ou seja, categorias que configuram as realizações possíveis dessa variável. Esta variável dependente pode ser binária (uma escolha entre duas alternativas), ou enária (uma selecção entre  $n$  alternativas). Assim sendo, o modelo pressupõe não só a existência de influências de diverso tipo (as variáveis independentes) que atingem, probabilisticamente, um ponto único de “decisão”, “opção” ou “escolha”, mas também a possibilidade de ocorrerem diversas observações, independentes do resultado da “decisão”. As variáveis independentes constituem, portanto, grupos de factores, enquanto as observações representam os dados do *corpus*, devendo estes, para tanto, ser codificados de modo a representarem uma determinada combinação de factores.

O VARBRUL trabalha na base de probabilidades, executando um modelo de regressão múltipla, fundamental para verificar o efeito que um conjunto de variáveis independentes exerce sobre a variável dependente. Deste modo, pode observar-se quais são as variáveis independentes que afectam, ou não, a regra da concordância em número. Este aspecto só é possível porque a ferramenta permite testar o grau de significância de cada grupo de factores, processo que se inicia com a rodagem de apenas um grupo de factores e ao qual o programa vai, sucessivamente, adicionando outros grupos de factores, até que todos estes tenham sido levados à análise computacional. Este processo, conhecido como análise *step-up/step-down*, fornece, no final, os valores para os pesos relativos e para o *log-likelihood* (cálculo da verosimilhança máxima),<sup>127</sup> um parâmetro estatístico que mede com elevado grau de confiança as frequências de uso de determinado registo linguístico, por forma a que este possa ser projectado como

estatisticamente representativo da comunidade. As formas como se processam quer as análises *step-up* e *step-down* quer a leitura e descodificação dos resultados fornecidos por estas encontram-se detalhadamente descritas em trabalhos específicos sobre o modo como operar com a ferramenta VARBRUL (Johnson, s.d.; Paolillo, s.d., 2002; Pintzuk, 1988; Guy & Zilles, 2007), pelo que remetemos para os mesmos a leitura sobre estas questões. Resta apenas referir que os resultados das análises, para além, de possibilitarem a confirmação, ou não, das hipóteses iniciais, também “*precisam de ser interpretados de acordo com as hipóteses colocadas: (i) um grupo de factores não é relevante e, como tal, a hipótese não se confirma; (ii) um grupo de factores é relevante, mas a influência dos factores não tem o peso inicialmente atribuído e, como tal a hipótese não se concretiza e (iii) o grupo de factores é saliente e a influência dos factores confirma a hipótese*” (Godinho, 2005:232).

Para realizar a quantificação dos dados no presente trabalho, recorreremos ao pacote de programas GOLDVARB X (Sankoff, Tagliamonte & Smith, 2005), a versão mais recente do programa VARBRUL (Pintzuk, 1988) para ser aplicada ao *software* Windows. Todavia, importa referir que lançámos também mão de outras versões do programa, quando entendemos que as mesmas forneciam melhor desempenho para obtenção de resultados a observar na análise. Foi o caso da versão GOLDVARB 2001 (Robinson, Lawrence & Tagliamonte, 2001), que produz tabelas coloridas que facilitam a detecção de interferências estatísticas entre grupos de factores. Também foi de grande utilidade o programa TSORT, do pacote VARBRUL-2, que permite isolar os dados relevantes a qualquer factor.

### **3.3. Descrição da variável dependente e das variáveis independentes**

As variáveis linguísticas do presente trabalho, independentes e dependente, adstritas a grupos de factores, foram constituídas partindo do pressuposto que, ao intervirem na marcação PL dos itens que compõem o SN actuam também na configuração da concordância variável do SN. Observando-se a influência das variáveis sobre os itens do SN, obtém-se, então, um entendimento mais sólido acerca dos fenómenos que inibem ou não a marcação PL no PA.

A exemplo da metodologia levada a cabo por Guy (1981a), Scherre (1988), Lopes (2001), P. Andrade (2003) e Baxter (2009), para estudos sobre a concordância nominal de variedades de PB, e Baxter (2004, 2009), Figueiredo (2008, 2009b) e Jon-And (2008, 2009), para as variedades africanas de português, submetemos os dados codificados (veja-se ponto 3.4) a análises, recorrendo à abordagem do tipo atomístico. Deste modo, o nosso estudo fará a observação do SN sob uma perspectiva ontogénica, se assim se pode dizer, uma vez que o analisa à escala individual, ou seja, cada elemento do SN é entendido como um dado de análise (uma ocorrência da variável dependente), definido pela presença ou ausência de marca de PL. Como tal, atribuiu-se um valor positivo aos SN's que tenham todos os seus constituintes marcados e um valor negativo aos que contenham pelo menos um elemento sem marcação de PL. Por seu lado, os dados foram codificados de acordo com quinze variáveis independentes do tipo linguístico-explanatório para serem analisadas na sua interação com duas variantes (*SN Plural com Concordância Total* e *SN Plural sem Concordância Total*) da mesma variável dependente. Pesquisas sobre a CPL-var no SN do PA, com recurso à abordagem sintagmática, serão apresentados em trabalhos posteriores, que pretendemos levar a cabo para complementar esta que agora se analisa.

Das variáveis que nos propusemos contemplar na presente análise, algumas foram já observadas em estudos sobre CPL-var (Guy, 1981a; Scherre, 1988; Lopes, 2001; P. Andrade, 2003; Baxter, 2004, 2009; Godinho, 2005; Figueiredo, 2008, 2009b; Jon-And, 2008, 2009), concordância variável de género (Lucchesi, 2000a; Figueiredo, 2009a) ou concordância verbal variável (Martins, 2007; Silva, 2010). Porém, também constituímos factores diferentes dos observados anteriormente em análises do tipo atomístico.

### **3.3.1. Variável dependente**

As variáveis dependentes constituem categorias particulares que se distinguem qualitativamente das outras variáveis também presentes na análise. Uma das características que as diferencia particularmente é o facto de serem descontínuas, isto é, não admitirem valores intermédios. Como tal, podem configurar-se como entidades binárias, ternárias ou enárias.

Importa também referir aqui que, para o estudo da *aplicação* ou *não-aplicação da regra da variação*, deverá ser descartado o pressuposto de que *aplicação* é sinónimo de desvio à norma padrão, enquanto *não-aplicação* significa o uso desta. Efectivamente, em algumas ocorrências, a própria norma padrão aceita mais do que uma forma,<sup>128</sup> defendendo Labov (1982) que a forma mais sistemática não é a prescritiva, mas sim a mais informal, isto é, a mais utilizada pelo falante comum. Nesta perspectiva, a discussão em torno da variação não deve ser subserviente à forma padrão, muitas vezes erroneamente entendida como o alicerce incontornável das ocorrências de um determinado dialecto.

A análise atomística da presente investigação contempla uma variável dependente binária (Tabela 9.1), que permite a obtenção de resultados indicativo-percentuais ou indicativo-probabilísticos para dois tipos de ocorrências: (i) a presença de morfema de PL no item analisado; (ii) em alternativa, a ausência de morfema de PL no item analisado. Este tipo de modelo binário radica nos princípios generativo-transformacionais aplicados na determinação da *aplicação* ou *não-aplicação da regra variável*,<sup>129</sup> e implica que a existência de um factor favorecedor da *aplicação* ocasiona, necessariamente, a proporcional *não-aplicação* do outro factor.

**Tabela 9.1.** A variável dependente (análise atomística).

<b>Variável Dependente</b>
1. Presença de morfema de plural no item analisado
2. Ausência de morfema de plural no item analisado

Com base na metodologia dos trabalhos sobre CPL-var mencionados no ponto 3.3 (Descrição da variável dependente e das variáveis independentes), partimos também do pressuposto que a marcação PL não acontece de forma casual, sendo condicionada pelas variáveis independentes. Considerandos mais detalhados acerca destas serão fornecidos no ponto seguinte deste trabalho.

### **3.3.2. Variáveis independentes**

Com base nos referidos estudos, Figueiredo (2008, 2009b) procedera já à constituição de algumas variáveis independentes linguísticas e sociais para uma observação

atomística preliminar acerca do modo como as mesmas poderão exercer influência na marcação de número dos itens do SN do PA. As variáveis constituídas foram: *idade* (variável social), *posição do item analisado relativamente ao núcleo do sintagma nominal*, *posição linear* (i.e. *posição do item analisado no sintagma nominal*), *marcas precedentes*, *contexto fonológico posterior* e *saliência fônica* (variáveis linguísticas). No presente estudo, tais variáveis foram complementadas com outras nove, de forma a obter-se um quadro mais completo sobre os fenómenos que subjazem à presença/ausência de marcação PL no SN do PA. Deste modo, numa perspectiva mórfica, acabámos por avaliar a possível influência na marcação exercida por quinze variáveis independentes, onze do tipo linguístico e quatro do tipo social. Estas variáveis estão patentes na Tabela 10.1 e serão detalhadamente descritas no ponto 3.6.

Como se pode ver na tabela, as variáveis independentes do tipo linguístico constituídas para o nosso estudo atomístico acerca da CPL-var no SN do PA configuram ainda diferentes subcategorias:

- (i) 3 variáveis fonológicas: *processos morfofonológicos de formação de plural* (doravante designada *saliência fônica*); *tonicidade dos itens lexicais singulares* (doravante designada *tonicidade*); *contexto fonológico posterior*.
- (ii) 7 variáveis morfossintáticas: *marcas precedentes*; *classe gramatical*; *posição do item analisado em relação ao núcleo do SN*; *posição linear*; *ordem do item na cadeia dos constituintes flexionáveis*; *grau de concordância de número no SN*; *marcação de género*.
- (iii) 1 variável lexical: *categoria semântica do nome*.

Após a constituição das variáveis que se pretendem observar, é necessário passar à codificação dos dados para serem introduzidos no suporte computacional, a fim de serem devidamente observados. Nesta conformidade, e com base nos estudos indicados no ponto 3.3, a observação dos dados envolveu as seguintes etapas: “(1) a definição do fenómeno em estudo (a variável dependente), isto é, a concordância de plural no sintagma nominal, (2) a definição das variáveis independentes – grupos de factores que são considerados como factores condicionantes da variável dependente em estudo, (3) a codificação dos dados – etapa que

*prepara os dados para serem tratados pelo suporte computacional, (4) a quantificação dos dados e (5) a análise da quantificação.*” (Godinho, 2005:188).

**Tabela 10.1.** Variáveis independentes (análise atomística).

<b>Variáveis independentes</b>	
<b>Variáveis independentes linguísticas</b>	
1.	Saliência fónica:
1a.	Processos morfofonológicos de formação de plural
1b.	Tonicidade dos itens lexicais singulares
2.	Marcas precedentes ao item analisado
3.	Contexto fonológico posterior ao item analisado
4.	Classe gramatical do item analisado
5.	Posição do item em relação ao núcleo do sintagma nominal
6.	Posição do item no sintagma nominal (Posição linear)
7.	Marcação de género
8.	Ordem do item analisado na cadeia dos constituintes flexionáveis em número
9.	Grau de concordância de número no sintagma nominal
10.	Animacidade do nome
<b>Variáveis independentes extralinguísticas</b>	
11.	Sexo
12.	Faixa etária
13.	Estadia (Permanência fora da comunidade)
14.	Escolaridade

### **3.4. Codificação dos dados**

Finda a constituição das variáveis dependente e independentes, foi necessário elaborar os factores que ficariam envasados nestas, com vista a podermos seleccionar os dados e preparar o arquivo de ocorrências que envolvem SN's PL, antes de passarmos ao tratamento do mesmo. Cabe aqui referir que, por vezes, a constituição dos factores coloca o investigador perante questões que exigem atenção e sagacidade cuidada, sob



pena de tratar incorrectamente os dados e, com tal, viciar os resultados finais. De facto, se bem que os factores das variáveis fonológicas se revelem relativamente lineares quanto ao tratamento (a alternância pode ancorar, por exemplo, na simples diferença de um ou dois traços fonológicos ou na realização ou não de algumas consoantes em final de palavra), o mesmo não se pode dizer dos factores das variáveis morfossintácticas e semânticas. Estes podem envolver, por exemplo, variação nas flexões, alternância de itens lexicais ou diferenças sintácticas produzidas pela derivação frásica. Perante realizações como a do exemplo [247], ao codificar-se atomisticamente o item “várias”, qual o critério a seguir: considerá-lo um adjectivo, com o significado de “diferentes”, ou, pelo contrário, entendê-lo como um quantificador, com o significado de “muitas”?

[247] PA:

*Mêmo que ele pode tê várias mulher [ANTOM]*

Este tipo de questões coloca o investigador, por vezes, perante o labirinto da ambiguidade, ao ter que lidar com as problemáticas relações entre forma linguística e função linguística, e que têm gerado alguma falta de convergência entre conceitos linguísticos e sociolinguísticos, começando pela própria tentativa de tratar o termo “significado”. Como tal, o pesquisador tem que desconsiderar o dado ambíguo ou, em contrapartida, optar por um critério único de classificação. No caso presente acreditamos que, na competência do falante almojarifano, actua em primeiro lugar o conceito do termo ligado à noção do concreto, isto é, de quantificador, e não tanto o conceito que remete para as conotações abstractas. Por conseguinte, pensámos atribuir aos itens “vários/várias” a categoria morfológica de quantificador, mas levantou-se aqui novo problema, já que aos quantificadores aparece também ligada a noção de indefinido. Voltaremos a esta questão no ponto 3.6.4 do presente trabalho, dedicado à constituição e codificação dos factores atribuídos à variável independente classe gramatical.

Por outro lado, a dificuldade do empreendimento cresce consideravelmente quando se lida com dialectos vernáculos, cujo código linguístico conhece variações consideráveis, quando comparado com o das formas padronizadas. A alternância das palavras na ordem linear da frase, associada à ausência de preposições, pode, em certos casos, colocar-nos perante realizações que exigem uma análise cuidada, quando se trabalha com a codificação de dados. Note-se, por exemplo, como a ausência de preposição atribui ao

SN preposicionado a função de OD na ocorrência seguinte, se observado à luz dos conceitos sintáticos da gramática tradicional:

[248] PA:  
*Ê vende **mil dobra** cada um peixe* [CLOTIM2]  
 “Eu vendo a mil dobras cada peixe”.

Numa perspectiva geral, a atribuição de factores a todas as variáveis ou contextos linguísticos e extralinguísticos que se pressupôs condicionarem a variável dependente é, também ela, parte do processo de codificação. Contudo, como a referida atribuição é regida por questões metodológicas bastante específicas, entendemos por bem expor os passos que seguimos na selecção dos factores que incluímos em cada uma das variáveis independentes. As tabelas que acompanham a explicação do critério de selecção têm o propósito de fornecer uma visão global sobre a forma como ficou constituído cada grupo de factores, com os seus respectivos factores.

### 3.5. Variável dependente constituída

A variável dependente que se observa no presente trabalho constitui a presença ou ausência de marca PL nos elementos do SN do PA. Principalmente com base nos trabalhos sobre CPL-var levados a cabo por Guy (1981a), Scherre (1988), Lopes (2001), P. Andrade (2003) e Baxter (2004, 2009), parte-se do pressuposto que a presença da marca de número PL não acontece por mero acaso, sendo sim influenciada pelos fenómenos que contemplámos nas variáveis independentes.

**Tabela 11.1.** Factores constituídos para a variável dependente.

1) Variável dependente				
Cód.	Factor	Exemplo	Nr. de Ocorrências	%
<b>S</b>	Presença de morfema de plural no item analisado	<b>JOVENS</b> habilitoso [CASTEH1]	1.202/2.340	49
<b>0</b>	Ausência de morfema de plural no item analisado	<b>jovens</b> HABILIDOSO [CASTEH1]	1.138/2.340	51

### 3.6. Variáveis independentes constituídas

#### 3.6.1. Variável independente *saliência fónica* e suas subvariáveis

Como podemos observar na Tabela 10.1, a variável independente *saliência fónica* contempla duas subvariáveis: *processos morfofonológicos de formação de plural e tonicidade dos itens lexicais singulares*. Na linha dos trabalhos que orientam a nossa pesquisa, iremos tratar estas individualmente, atribuindo à primeira a designação de *saliência fónica* e à segunda a denominação de *tonicidade*.

##### 3.6.1.1. *Saliência fónica*

O grupo de factores *saliência fónica* tem marcado presença constante nas pesquisas sobre concordância variável de diferentes tipos, desde os trabalhos primordiais de Naro & Lemle (1976) e Lemle & Naro (1977) sobre concordância verbal. Pela mesma altura, Braga & Scherre (1976), tendo como base o trabalho de Naro & Lemle (1976), aplicaram a metodologia aos estudos da concordância PL, postulando que existe uma relação directa entre maior frequência de marcação em número e maior *saliência fónica*, com a segunda gerando a primeira. A escala estabelecida por Braga & Scherre (1976:465-477) para avaliar o grau de diferenciação fónica na relação singular/plural propunha os seguintes níveis: (1) inserção de *-s* final e abertura vocálica ou PL metafónico (p.e. olho > olhos/); (2) inserção de *-s* final e mudança silábica (p.e. milhão > milhões; imóvel > imóveis); (3) inserção de *-es* final em palavras terminadas em *-r* no singular (p.e. flor > flores); (4) inserção de *-s* final em palavras que formam o PL de modo regular (p.e. livro > livros); (5) inserção de *-es* final em palavras que acabam em *-s* no singular (p.e. país > países; rapaz > rapazes). Os resultados levaram as autoras a concluir que as formas mais salientes favorecem mais a inserção de PL do que as formas menos salientes. Paralelamente, observaram também que a escala de *saliência fónica* não só ganhava mais evidência nos falantes de classe média e média alta do que nos de classe baixa como também apresentava tendência para inibir a marcação PL dos itens terminados em *-s* no singular e que pluralizam com a inserção de *-es* final. O fenómeno, por nós descrito no ponto 1.1 (Português do Brasil e concordância nominal variável de número), foi justificado em termos de analogia (Braga & Scherre, 1976:474).

A pesquisa de Braga (1977), na linha das anteriores, aplica a mesma escala de diferenciação fónica e volta a referir que a *saliência fónica* detém importância na marcação PL da classe média. Por seu lado, a classe baixa apenas inibe a pluralização dos itens terminados em *-s*, favorecendo-a em todos os outros factores. Posteriormente, a análise efectuada por Scherre (1978:79-82), observando factores sociais para o MRJ, determinaria uma escala de três níveis para verificar a influência da *saliência fónica* na marcação PL.<sup>130</sup> A autora detectou então que os falantes escolarizados aplicam mais a pluralização do que os semi-escolarizados, apesar de ambos favorecerem a marcação. Simultaneamente, Scherre (1978:82) verificou que os itens em *-s* no singular apresentam menor probabilidade de concordância do que os elementos regulares.

Ponte (1979:164-166), observando o PPA, recorreu a uma escala idêntica à de Scherre (1978), tendo constatado uma escala hierárquica de marcação semelhante à dos informantes semi-escolarizados do MRJ, mas com a diferença de os seus falantes favorecerem mais a marcação PL dos itens em *-s* no singular do que a dos elementos regulares. Por seu lado, Nina (1980), com recurso a uma escala de diferenciação fónica de quatro níveis,<sup>131</sup> analisou a pluralização no SN da fala da micro-região bragantina, Pará. Apesar de os resultados apontarem para um fraca motivação da *saliência fónica* na pluralização, os itens que pluralizam com adição de *-s* final e mudança silábica (p.e. *casal/casais*) apresentaram uma taxa de inibição total (100%). Nina (1980) indicou ainda uma escala de hierarquia fónica com influência na marcação PL, mas Scherre (1988:70) chamou a atenção para o facto de a mesma não ter sido correctamente interpretada, até porque Nina (1980), ao trabalhar com percentuais, não mediu o peso real das marcações. Assim sendo, e por estarmos de acordo com a observação de Scherre (1988:70), abster-nos-emos de efectuar quaisquer comparações entre os dialectos bragantino e de Almojarife.

Guy (1981a:181-189), aplicando a perspectiva proposta por Naro (1981) de se observar a relação entre maior frequência de marcação em número e maior *saliência fónica* tendo em conta a diferença fónica e a tonicidade, analisou a marcação PL na fala de informantes semi-escolarizados do Rio de Janeiro, mas aumentou a escala de diferenciação fónica para sete níveis.<sup>132</sup> Com tal metodologia, o autor pretendia verificar também se se estabelece uma relação directa entre o grau de diferenciação fónica e a

percepção do ouvinte no momento da aquisição, levando este a apreender mais facilmente as formas que melhor percebesse. Nesta conformidade, a variável independente *saliência fônica* prestaria uma contribuição importante para se compreenderem dois fenômenos opostos no momento da aquisição, a não-realização e a apreensão, já que será “*a partir dos itens com concordância menos saliente que a variação se inicia; e é a partir dos itens com concordância mais saliente que se dá a percepção, na fase de aquisição do mesmo fenômeno: a aquisição do fenômeno da concordância se inicia pelos itens com maior oposição singular/plural*” (Lopes, 2001:132-133). Assim sendo, a variável que aqui se aborda presta-se ao estudo quer da variação da concordância quer da aquisição de L1 e de L2, já que não existe paralelismo quanto ao efeito da *saliência fônica* quer na aquisição de diferentes variedades do mesmo dialecto (Castro & Ferrari-Neto, 2007:68) quer na aquisição registada em distintas fases etárias (Correia, 2005; Baia, 2008). Contudo, a este propósito, lembre-se que, no caso específico da ASL (pidgins e crioulos), os falantes adultos poderão dar primazia às estruturas do tipo CV em virtude da sua força lexical (Clements, 2009:22), adquirindo-as em vez de estruturas que apresentam maior *saliência fônica*. Portanto, dificilmente se confirmará o postulado de Naro & Lemle (1976:240-241) e Lemle & Naro (1977:47) de que a mudança é determinada pela deriva natural da língua, iniciando-se nos contextos menos salientes e atingindo, depois, as formas mais salientes, por “imitação”.

No caso específico do estudo de Guy (1981a), estes aspectos parecem ajustar-se também, já que os resultados finais evidenciaram que não existe uma diferença relevante na influência da marcação PL motivada pelos elementos regulares e finalizados no singular em *-ão* e *-l*, com todos eles inibindo a pluralização. Por outro lado, o item “vez”, assim como os que finalizam no singular em *-r* e *-s*, isto é, os mais salientes, exibem tendência para favorecerem a marcação PL. O pesquisador chamou ainda a atenção para a forte propensão de os elementos que pluralizam metafonicamente se apresentarem marcados, ainda que, no seu *corpus*, o seu número fosse bastante reduzido. Como tal, Guy (1981a:189) entendeu que a variável independente *saliência fônica* pode mesmo ser reduzida a uma oposição binária em termos fonéticos: itens que terminam em consoante, com tendência para favorecerem a marcação, *versus* itens que terminam em

vogal, com propensão a inibirem a pluralização. Ainda a reter da análise levada a cabo pelo autor, são os seguintes postulados: (i) não é possível comprovar conclusivamente que a *saliência fônica* exerça efeito na marcação PL; (ii) a influência da tonicidade na marcação é aparente, visto ser idêntica à exercida pela mesma variável sobre o cancelamento do –s final não-morfémico.

Buscando confirmar as premissas advogadas por Guy (1981a), Scherre (1988) reanalisa os dados do MRJ e observa a possibilidade de o eixo da *saliência fônica* se estender igualmente ao número de sílabas. Com esta reanálise, a pesquisadora procura também estudar, em pormenor, a possível sobreposição entre as três dimensões referidas (diferença fônica, tonicidade e número de sílabas) na marcação do PL e conclui que a escala da *saliência fônica* não pode ser reduzida à oposição binária proposta por Guy (1981a), já que esta variável estará correlacionada quer com as características sociais dos falantes (actua a nível do funcionamento sincrónico da gramática, estabelecendo diferenças em função dos anos de escolarização dos falantes) quer com a Condição de Distintividade da hipótese funcionalista das Condições de Distintividade (Kiparsky, 1972:195), que postula existir uma tendência para a informação semântica relevante ser retida na estrutura de superfície (as subcategorias da *saliência fônica* estabelecem diferenças entre si na forma de motivarem a marcação PL). Nesta conformidade, Scherre (1988:78) concluiu que, por um lado, “os processos de formação do plural e a tonicidade dos itens lexicais singulares influenciam a concordância de número entre os elementos do SN”, com os itens que apresentam maior distinção na oposição singular/plural a favorecerem mais a marcação, mas, por outro lado, que “o número de sílabas dos itens lexicais singulares não exerce influência sobre a concordância de número dos elementos do SN”. Em aberto, a autora deixou ainda a possibilidade de a *saliência fônica* estabelecer ligação com a relação entre variação e mudança, que viria a ser confirmada quando ela, no mesmo trabalho, se debruçou sobre a forma como as variáveis sociais intervêm no processo de marcação de número do MRJ.

Entretanto, Lucchesi (2000a:273), observando a concordância variável de género de HEL-Ba, conclui também que a *saliência fônica* influencia a marcação, já que esta é mais evidente nos itens que formam o feminino com alteração do radical da palavra, isto é, morfologicamente mais salientes.<sup>133</sup>

Lopes (2001:137), utilizando a uma escala de diferenciação fónica de sete níveis, idêntica à de Guy (1981a), recorreu igualmente às dimensões propostas por Scherre (1988), isto é, à diferença fónica, tonicidade e número de sílabas, estudando a influência da variável independente *saliência fónica* na marcação PL de dois públicos da NURC: o grupo popular e o grupo universitário. Paralelamente, e na linha do efectuado por Scherre (1988), observa também como a mesma variável actua sobre os diversos níveis de escolaridade (Fundamental, Colegial e Universitário), bem como sobre as diferentes categorias gramaticais. Outra particularidade do estudo de Lopes (2001:155-164) é a análise do modo como o grau *saliência fónica* actua ao nível da marcação PL em falantes de descendência afrobrasileira e não-afrobrasileira.<sup>134</sup> Lopes (2001:164) conclui que, no processo de aquisição, os itens mais salientes são os que apresentam o maior grau de concordância, enquanto os elementos menos salientes são alvo de variação. Este aspecto acaba por ter reflexo na concordância revelada pelo grupo afrobrasileiro, que denota maior índice de variação pelo facto de a sua história assentar na ancestralidade escrava, isto é, em falantes cuja aquisição do português se deu a partir de contactos linguísticos diversos, marcados por grande divergência nos DLP's.

P. Andrade (2003:100-104), por seu lado, estabeleceu uma escala de diferenciação fónica para oito níveis,<sup>135</sup> tendo os resultados iniciais apontado para a possibilidade de o efeito da *saliência fónica* não se fazer sentir na marcação PL do dialecto de HEL-Ba. De facto, parecia ocorrer, neste, a tendência para uma maior incidência de marcação nos itens de alternância regular, o que contrariava a teoria de as formas mais salientes serem mais marcadas do que as menos salientes. Esta particularidade levava a ferramenta VARBRUL a rejeitar o grupo de factores *saliência fónica*, não o considerando como motivador da variação registada na pluralização do SN. Contudo, amalgamando factores que compartilham traços linguísticos,<sup>136</sup> a linguista elevou os dados de ocorrências incluídas nas novas supercategorias de factores e fez com que o grupo de factores ficasse reduzido a três factores, tornando-se relevante para a marcação e fornecendo dados sobre índices de frequência. Os achados de P. Andrade (2003) levaram a pesquisadora a confirmar outros postulados anteriores (Guy, 1981a; Lopes, 2001), que chegaram a conclusões aproximadas, apesar de terem trabalhado com factores diferentes. A

conclusão final de P. Andrade (2003:104) foi que a *saliência fónica* tem influência na aplicação da regra da marcação PL.

Baxter (2004:107-113), analisando o PT, chama a atenção para o facto de a *saliência fónica* ter vindo a ser também observada quer em estudos sobre processos de aquisição em despidginização (Emmerich, 1984) e descrioulização (Bickerton, 1975) quer em trabalhos sobre aquisição de L1 (Ingram, 1989) e L2 (Bayley, 1994; Wolfram, 1985). Nas análises efectuadas, o autor optou por estudar o comportamento da *saliência fónica* quer quanto à alternância singular/plural quer na relação que estabelece com a tonicidade. Para tanto, constituiu uma escala de diferenciação fónica com seis graus.<sup>137</sup>

Cruzando as variáveis independentes *saliência fónica* e *tonicidade*, Baxter (2004:112-113) apresenta resultados a revelaram padrões gramaticais aproximados entre a geração mais nova e a geração intermédia dos tongas, mas que se afastam substancialmente dos da faixa mais idosa da mesma comunidade. À luz dos factos, este aspecto é perfeitamente entendível, uma vez que a geração mais velha esteve sujeita a forte *input* de DLP's L2 proveniente de falantes adultos cuja L1 é africana. Paralelamente, o efeito da *saliência* que foi detectado na marcação PL do PT vai ao encontro dos achados em estudos de aquisição, os quais apontam para um desenvolvimento precoce da morfologia flexional envolvendo as oposições mais salientes (Bayley, 1994:167), por contraste às oposições que atingem as *saliências* mais fracas, adquiridas mais tardiamente. Assim sendo, o efeito da *saliência* na marcação PL é fortemente favorecido pelas distinções morfofonológicas das oposições mais marcadas, pelas combinações de sílabas plurais tónicas e pelas mudanças que afectam o radical da palavra.

Para o PMp (Jon-And, 2008) e o PCV (Jon-And, 2009), Jon-And apresentou igualmente resultados provenientes do cruzamento entre os grupos de factores *saliência fónica* e *tonicidade*. No PMp, a marcação PL revelou tendência para incidir fortemente sobre as oposições mais salientes (p.e. *casal* > *casais*; *avião* > *aviões*; *pão* > *pães*), mas com os itens terminados em *-s* ou *-r* no singular, isto é, que fazem o PL acrescentando *-es* ao radical (p.e. *mês* > *meses*; *mulher* > *mulheres*), inibindo fortemente a marcação. Quanto ao PCV, todos os factores constituídos marcam positivamente a pluralização, à excepção dos elementos em *-s* no singular, que a inibem fortemente, e dos *oxítonos ou monossílabos tónicos regulares* (p.e. *café* > *cafés*; *pai* > *pais*), com um peso próximo da



neutralidade, mas a tender ligeiramente para a inibição. Deste modo, e relativamente a esta variedade de português, a linguista considera que “*a variável morfofonológica saliência fónica tem um efeito menos forte no PCV do que em outras variedades africanas e variedades brasileiras de português*” (Jon-And, 2009:7).

Tendo como base as metodologias de trabalhos anteriores, aplicámos inicialmente aos nossos dados uma escala de diferenciação fónica para oito níveis, de acordo com a Tabela 12.1. Assim, no que diz respeito aos itens que pluralizam com inserção de *-s* e mudança silábica, seguimos o proposto por Guy (1981a), mas ampliámos para três os nossos factores: (i) um para englobar os itens finalizados em *-l*; (ii) outro para abarcar os nasais em *-ão* que pluralizam em *-ões*; (iii) e um terceiro para o mesmo tipo de nasais, mas que formam o PL em *-ães*. A separação dos itens em *-ão* que pluralizam com inserção de *-s* e alteração silábica foi primeiramente testada por P. Andrade (2003:102) no dialecto de HEL-Ba, tendo os dados apresentado percentuais de marcação distintos (*-ão/-ões* = 16%; *-ão/-ães* = 0%). Recorrendo a idêntica distinção, pretendíamos confirmar se os falantes do PA estabeleciam também escalas diferentes de marcação para os itens e se essa distinção apresentava paralelismos com os dados de P. Andrade (2003). Contudo, não foi possível observar este efeito, já que no nosso *corpus* são nulas as ocorrência de elementos em *-ão* que pluralizam em *-ães*.

Quanto aos itens nasais em *-ão* que pluralizam de forma regular, isto é, com inserção de *-s* final (p.e. irmão > irmãos), fundimos o seu factor com os dos outros elementos nasais que formam também o PL de modo regular, ou seja, com os itens que finalizam em *-em* e *-um* (p.e. homem > homens; atum > atuns).

Para a constituição dos restantes factores que compõem a nossa escala de diferenciação fónica, adoptou-se a metodologia de Scherre e Braga (1976): (i) itens com inserção de *-s* em palavras de PL regular (p.e. primeiro > primeiros); (ii) itens com inserção de *-s* e abertura vocálica, ou seja, plurais metafónicos (p.e. jogo > jogos)<sup>138</sup>; (iii) itens terminados em *-r* e que pluralizam com inserção de *-es* (p.e. jogador > jogadores); (iv) e itens terminados em *-s* e que fazem o PL com inserção de *-es* (p.e. mês > meses). Relembremos ainda mais dois aspectos relativamente à inclusão de determinados casos em alguns dos factores por nós constituídos: (1) os itens com realização fonológica parcial de PL (p.e. trabalhador > trabalhadore; mês > mese) foram

codificados como possuindo marca de PL (Scherre, 1978), sendo incluídos nos factores dos elementos terminados no singular em *-r* ou *-s*, respectivamente; (2) as expressões do tipo de “às vezes” foram excluídas das nossas codificações. Scherre (1988:75) incluiu ainda no factor dos metafónicos as ocorrências com duas marcas de PL (p.e. *papelzinho* > *papezinhos*), que no nosso *corpus* são nulas.

Na tabela que se segue, podem ser vistos os factores que compõe a nossa escala de diferenciação material fónica, bem como o seu número de ocorrências e respectivos percentuais de marcação PL:

**Tabela 12.1.** Factores constituídos para a variável independente *saliência fónica*.

2) Variável independente <i>saliência fónica</i>				
Cód.	Factor	Exemplo	Nr. de Ocorrências	%
<b>R</b>	Plurais regulares	nos PRIMEIROS jogos [CASTEH1]	995/1.973	50
<b>D</b>	Plurais metafónicos	nos primeiros JOGOS [CASTEH1]	(10/15)	67
<b>Y</b>	Plurais nasais regulares: -ão/-ãos; -em/-ens; -um/-uns	<b>HOMENS</b> fiderado [CASTEH1]	69/145	48
<b>n</b>	Singular -ão/-ães	[-]	-	-
<b>B</b>	Singular -ão/-ões	a <b>RAZÕES</b> [CASTEH1]	(21/27)	78
<b>L</b>	Singular -l	<b>bancos</b> MUNDIAIS [ABILH2]	(3/10)	30
<b>;</b>	Singular -r	<b>JOGADORES</b> mais interessado [CASTEH1]	42/79	53
<b>s</b>	Singular -s (-z)	uns MESES [CASTEH1]	61/86	71
<b>Totais:</b>			1.201/2.335	51

Com recurso à variável independente *saliência fónica* formulámos a seguinte hipótese:

**Hipótese 1 – Saliência fónica (Processos de formação de plural):** o maior peso de marcação PL incide sobre os itens que apresentam maior *saliência fónica*. Nesta conformidade, são mais marcados os plurais metafónicos, que pressupõem dupla distinção (fónica e morfológica) e os itens que pluralizam com ocorrência de mudança

silábica (palavras terminadas em *-l*, em *-r*, em *-s* e em *-ão*). Quanto aos plurais menos marcados, serão constituídos pelos itens regulares.

### **3.6.1.2. Tonicidade (Tonicidade dos itens lexicais singulares)**

A variável independente *tonicidade* estabelece ligações intrínsecas com a *saliência fónica*, uma vez que aborda questões relacionadas com a maior percepção da sílaba tónica, isto é, da sílaba que se destaca fonicamente. Como se viu, a propósito das observações acerca do grupo de factores *saliência fónica* (ponto 3.6.1.1), alguns estudos em aquisição têm apresentado evidências que esta se inicia a partir das formas silábicas mais fortes. De um modo geral, os falantes de uma L1 têm consciência das sílabas que compõem as palavras pertencentes ao seu código linguístico natural. Nesta conformidade, pode considerar-se que as sílabas representam unidades de pronúncia intuitivamente reconhecidas e que qualquer falante, mesmo que seja analfabeto, saberá não só diferenciar as sílabas constantes, por exemplo, na palavra “janela” (*ja-ne-la*), como perceber também a sílaba tónica da mesma. De facto, quando articulamos dissílabos ou polissílabos, distinguimos, normalmente, uma das sílabas (a sílaba tónica), que se destaca em relação às outras (as sílabas átonas), sendo percebida como mais “forte” devido à intensidade com que é produzida a vogal que nela ocorre.

Não obstante, no caso do português, quando ocorre um registo oral mais rápido, mesmo os falantes maternos podem experimentar alguma dificuldade em fazer a distinção silábica, especialmente em sequências fonológicas que envolvem ditongos. Tomando como exemplo a palavra “chapéu”, isto é, um termo que envolve a ocorrência de um ditongo decrescente (sequência de uma vogal e uma semivogal), constatamos que ele se revela estável, pronunciando-se sempre como um ditongo e nunca como uma sequência de duas vogais distintas. Já em situações que pressupõem o uso de termos como “seara” ou “Manuel”, ou seja, que abrangem ditongos crescentes (sequência de uma glide e uma vogal), constata-se que estes poderão ser pronunciados como ditongos, numa fala mais rápida, ou como hiatos, quando o registo oral é mais pausado.

Também a percepção das vogais (e sílabas) tónicas e átonas poderá não ser tão evidente como à primeira vista parece, porque, em algumas realizações, o falante/ouvinte depara-se com ocorrências que pressupõem quer a captação das

chamadas vogais e sílabas subtónicas quer a não-distinção da fronteira de morfema entre segmentos contíguos. Com efeito, algumas palavras derivadas, como “cafezinho” ou “facilmente”, além da acentuação fónica principal, que incide sobre a sua vogal tónica, possuem também uma acentuação fónica secundária, que recai, por norma, sobre a vogal tónica da palavra primitiva. Por outro lado, alguns segmentos átonos dependem, em termos de acentuação fónica, de um segmento contíguo, que pode ser precedente (p.e. calou-se) ou posterior (p.e. a bata).

Durante o processo de aquisição da L1, a criança vai, paulatina e inconscientemente, adquirindo o sistema fonológico do seu idioma, isto é, não só vai apreendendo as relações que se estabelecem entre os sons como também vai delimitando o campo de funcionamento de cada um deles. Todavia, os próprios falantes de uma L1 podem enfrentar dificuldades na interpretação de determinados contínuos sonoros do seu idioma natural, mesmo em fase tardia. Assim sendo, e como referimos no ponto 2.7 (Aquisição das categorias funcionais da L2: os casos do santomense e português de Almojarife), a ausência de conhecimento da LA pode conduzir à não-interpretação de articulações sonoras desta, como acontece com a não-deteção da sequência DET+N do português por falantes de L1's do grupo níger-congo atlântico, que será fonologicamente interpretada como um todo indistinto (cf. DeGraff, 1999b:482). Portanto, não surpreenderá que as formas silábicas mais salientes sejam apreendidas em primeiro lugar (J. Silva, 2005:65), contribuindo, em muitos casos, para a emersão de fenómenos de variação, que acabam cristalizados por *transmissão irregular*. Face a estes aspectos, a variável independente *tonicidade* poderá configurar-se como um elemento de importância extrema para o estudo de imensos fenómenos linguísticos, sejam eles em situação de aquisição sejam em contextos sociais de variação.

Guy (1981a:138), ao observar o efeito da tonicidade no apagamento do –s dos falantes cariocas semi-analfabetos, constatou que a variável tonicidade apresenta forte efeito sobre o fenómeno, com os itens monossilábicos e polissilábicos tónicos a resistirem acentuadamente ao referido apagamento (mais os primeiros do que os segundos), e os elementos átonos a favorecerem-no bastante. A razão para as sílabas tónicas favorecerem mais a inserção do –s do que as sílabas átonas reside, segundo o autor, no facto de as primeiras possuírem forte traço prosódico, isto é, propriedades de

intensidade, duração e altura em graus superiores aos das segundas. Como tal, são perceptualmente mais distintivas, logo mais resistentes aos apagamentos, como o demonstram ocorrências várias de manutenção e apagamentos registadas não só na história de línguas como o latim e o português mas também de CP's, como acontece com a fala caboverdiana.<sup>139</sup> Quanto à distinção entre monossílabos e polissílabos, o fenómeno não é mais do que uma extensão do anterior, já que os primeiros são acusticamente mais longos e proeminentes do que os segundos.

Guy (1981a:137) observou ainda que a inserção do *-s* final é também moderadamente favorecida pela presença da sílaba tónica na primeira posição da palavra em contexto seguinte (quando esta ocorre), um fenómeno que encontra respaldo em diferentes línguas hispânicas, casos do espanhol panamiano (Cedergren, 1973), espanhol cubano (Terrell, 1979) e EPR (Poplack, 1992 [1979]). Face às evidências sincrónicas e diacrónicas reveladas pela resistência das sílabas tónicas aos fenómenos de redução em vários dialectos, Guy (1981a:292) entende que se estará na presença de um universal linguístico, já que estes operam a nível de qualquer idioma humano. Assim sendo, não surpreende que a análise da influência do fenómeno do traço prosódico das sílabas na marcação PL venha revelando presença constante em trabalhos sobre variação.

Scherre (1988:77), para observar o efeito da variável independente *tonicidade* na inserção de marcas do MRJ, tomou como ponto de partida o item singular, estabelecendo três subdivisões prosódicas para o mesmo: (i) oxítonos e monossílabos tónicos; (ii) paroxítonos e monossílabos átonos; (iii) proparoxítonos. A constituição de um factor único que inclui paroxítonos e monossílabos átonos ficou a dever-se ao facto de a maioria dos monossílabos átonos surgir realizada sob a forma de itens proclíticos dos substantivos, o que os leva, muitas vezes, a serem interpretados como sílabas iniciais das palavras seguintes. Assim, perde-se tanto a noção de definitude como a informação de PL, embora se preserve o /s/ como partícula proclítica. Consequentemente, este elementos apresentam um elevado índice de marcação PL que, na opinião de Scherre (1988:113), e corroborada por Lopes (2001:142), é muitas vezes virtual.<sup>140</sup>

As conclusões de Scherre (1988:135) sobre o peso da tonicidade da sílaba do item singular na marcação PL são que o mesmo influencia a concordância, ocorrendo esta com mais propensão nos elementos oxítonos e menos incidências nos casos em que a

sílaba final não é tónica (paroxítonos e proparoxítonos). No entanto, convém relembrar que os resultados da autora tiveram em conta apenas as classes gramaticais em que ocorre o maior grau de diferenciação material fónica, ou seja, os substantivos, as categorias substantivadas e os adjectivos.

Lopes (2001:144-145), apesar de concordar com Scherre (1988:113) quanto à questão do mascarar da pluralização dos monossílabos átonos, separou-os dos paroxítonos em virtude de, ao contrário destes, apenas realizarem o PL de forma regular. Assim, optou por constituir factores individuais para ambos. Após trabalhar os seus dados, a autora verificou que, tal como o MRJ, a NURC revela tendência para marcar mais o PL em oxítonos e monossílabos tónicos, logo seguidos dos monossílabos átonos. A inibir a marcação surgem os proparoxítonos (os mais inibidores) e os paroxítonos (os menos inibidores). Dado que a oposição entre singular e PL mostra tendência a recair sobre a sílaba mais saliente, Lopes (2001:141) entende que tal fenómeno reflecte duas situações: “1) uma perda da concordância, inicialmente nas formas menos salientes, pois se dá mais facilmente a conservação da regra em sílabas salientes; 2) um processo de aquisição do fenómeno, ou parte dele, resultante de percepção parcial, em que formas oxítonas e monossílabos tónicos, por terem o morfema de plural mais saliente, foram as primeiras a serem percebidas e serem alvo de concordância”.

Quanto a Baxter (2004:109), cruzou as variáveis independentes *saliência fónica* e *tonicidade* para observação da influência das mesmas na marcação PL do PT, metodologia igualmente aplicada por Jon-And (2008:4, 2009:5) para estudo tanto da CPL-var do PMp como do PCV.

**Tabela 12.2.** Factores constituídos para a variável independente *tonicidade*.

<b>3) Variável independente <i>tonicidade</i></b>				
<b>Cód.</b>	<b>Factor</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
<b>X</b>	Oxítonos/Monossílabos tónicos	minhas <b>PREOCUPAÇÕES</b> [CASTEH1]	179/282	64
<b>M</b>	Monossílabos átonos	dAS <b>ondas</b> [CASTEH1]	283/320	88
<b>P</b>	Paroxítonos	<b>JOVENS</b> habilitoso [CASTEH1]	735/1.726	43
<b>&amp;</b>	Proparoxítonos	alguns <b>MÉDICOS</b> [CASTEH1]	(5/12)	42
<b>Totais:</b>			1.202/2.340	51

A Tabela apresenta os factores da variável independente tonicidade, tendo sido aplicada, para elaboração dos mesmos, a metodologia de Lopes (2001:144-145). Assim, e observando os fenómenos de marcação PL na linha do Princípio da Saliência Fónica (Naro & Lemle, 1976:240-241; Lemle & Naro, 1977:47), o que se pretende confirmar, tal como em todos os trabalhos anteriores, é se:

**Hipótese 2 – Tonicidade:** os itens do factor *oxítonos e monossílabos de uso tónico*, por apresentarem maior *saliência fónica*, são primeiramente apreendidos e revelam maior índice de concordância. Seguidamente, procura verificar-se se são mais marcadas as palavras proparoxítonas e, por fim, as paroxítonas e os monossílabos átonos.

### 3.6.2. Variável independente *marcas precedentes ao item analisado*

Os itens em análise, exibindo ou não marcas de PL, são também antecidos por elementos que podem estar ou não pluralizados. Deste modo, para se verificar se a presença de marcas anteriores impede ou beneficia a marcação no elemento seguinte, é necessário recorrer à variável independente *marcas precedentes ao item analisado*.

Esta variável independente é clássica em estudos sobre concordância de número,<sup>141</sup> visto permitir o vislumbramento da influência que o item ou itens situados anteriormente ao elemento analisado poderão ter na marcação deste. Os estudos sobre este grupo de factores têm optado por considerá-lo quer de modo distinto quer correlacionado com outras variáveis independentes, nomeadamente a *posição linear*. Cruzando *marcas precedentes, posição linear e classe gramatical*, Scherre (1988:183) efectuou análises aplicando a metodologia de Poplack (1980a), que observou o apagamento do *-s* final, marca de PL, no EPR<sup>142</sup>. Os resultados de Scherre (1988:184) indicaram que os itens flexionados em número conduzem também à flexão do elemento ulterior. Por seu lado, as ocorrências com apagamento favorecem a inibição no item que lhes é posterior. Scherre (1988:238) justificou os resultados recorrendo ao Princípio do Processamento Paralelo, baseado no Princípio da Iconicidade (Haiman, 1983:782), que determina que o falante constitui, mentalmente, grupos que compartilham traços. Por outras palavras: Scherre (1988) concluiu que são as marcas que levam a marcas, por oposição a zeros, que implicam zeros.<sup>143</sup> Nesta conformidade, contesta a teoria funcionalista das Condições de Distintividade (Kiparsky, 1972:195), defensora de que a língua exhibe

tendência para reter a informação relevante na estrutura superficial e descartar aquilo que for dispensável, isto é, inserir o –s marca de PL apenas quando ele é fundamental para indicar a ideia de pluralidade. Como tal, caso ocorra marcação no primeiro item do SN, e segundo Scherre (1988), a tendência não será para a apagar nos elementos seguintes, a fim de se evitar a redundância de informação, mas sim de inserir mais marcas, regularizando-se o padrão da CPL-var conforme o padrão icónico das línguas: quanto mais coesão, mais marcas; quanto menos coesão, menos marcas. Todavia, Lopes (2001:88-89) chama a atenção para o facto de Scherre (1988) ter incidido a atenção sobre *“estruturas do tipo ‘os meus pais’, que são vistos como casos específicos, pois a presença de marca no possessivo independe da marca no artigo ou demonstrativo que o antecede. Outro aspecto visto por Scherre é que duas ou mais marcas precedentes é fator favorecedor e não o contrário, entrando em desacordo com o previsto pela hipótese funcionalista”*. Estes aspectos levariam Lopes (2001:202) a debruçar-se sobre a variável *marcas precedentes* de modo distinto do efectuado por Scherre (1988), mas, apesar das diferenças de análises entre ambas as autoras, as conclusões seriam idênticas quanto à questão do processamento paralelo: na NURC, o *“contexto antecedente com marcas leva a mais marcas subsequentes; e, o contrário, contextos com Zero imediatamente antecedente produzem uma inibição da regra”* (Lopes, 2001:261).

Por seu lado, P. Andrade (2003:100) e Baxter (2009:282) advogariam que os seus achados, apesar de não contrariarem os de Scherre (1988) e Lopes (2001), apontam para que a incidência de marcação PL esteja conotada não a factores linguísticos, mas sim a factores de ordem social. Em HEL-Ba, e segundo P. Andrade (2003:99), os itens antecidos por elemento com marcação formal de PL revelam tendência para inibição da marcação, por oposição aos que são precedidos de realizações sem marca formal de PL, que a favorecem. Em adição, os itens que apresentam a mais elevada incidência para a incorporarem a marcação são os antecidos por ocorrências que detêm marca semântica de PL, caso dos numerais. Se bem que a hipótese funcionalista das Condições de Distintividade (Kiparsky, 1972:195), que advoga que a informação semanticamente relevante é retida na estrutura de superfície, encontre aqui algum acolhimento no que concerne à questão dos numerais, a verdade é que a autora ancorou as suas explicações no pressuposto avançado por Guy (1981a). Com efeito, este entende que a baixa



marcação no item seguinte ao elemento portador de marca de PL tem a ver com a influência das línguas africanas, nas quais “*predomina o uso de marcadores iniciais, em forma de prefixo ou clítico*” (P. Andrade, 2003:100). Assim sendo, confirma-se a tendência para, no PVB, o PL ser indicado na primeira oportunidade (Guy, 1981a:301-302), por norma o elemento que antecede o nome núcleo do SN.

Baxter (2009:280-282), por sua vez, comparando os resultados da variável independente *marcas precedentes* do PT e do dialecto de HEL-Ba, chegou à conclusão que o princípio de que *marcas levam a marcas* é marginal à atribuição de PL nos dois dialectos, os quais exibem estruturas de SN e padrões de marcação PL bastante simples. Como tal, e de acordo com o aventado por P. Andrade (2003:100), Baxter (2009:282) acredita que a marcação PL se deve não apenas a factores eminentemente linguísticos mas também à influência quer dos substratos quer dos contextos sociolinguísticos em que ambos os sistemas linguísticos se desenvolveram.

Relativamente aos estudos de Jon-And (2008, 2009) sobre o PMp e o PCV, a autora não apresenta quaisquer valores para o grupo de factores *marcas precedentes*, em virtude de o mesmo revelar interferências com outras variáveis e, conseqüentemente, ter sido excluído pela ferramenta VARBRUL.

Quanto a nós, a fim de verificarmos como se comporta a concordância PL no SN do PA em termos de marcas precedentes, fomos constituindo factores para a variável independente *marcas precedentes* à medida que íamos coligindo os SN's do nosso *corpus*, tendo como base orientadora a metodologia aplicada em trabalhos antecedentes (Braga, 1977; Guy, 1981a; Scherre, 1988, 2001; Lopes, 2001; P. Andrade, 2003; Baxter, 2009). Para tanto, e dado que o que importa verificar é o modo como o item precedente, com ou sem marcação PL (formal ou semântica), afecta a pluralização no elemento seguinte, tivemos como ponto de partida o elemento (ou elementos) que marca(m) o início da cadeia sintagmática, pelo que os nossos diferentes factores ficaram distribuídos por três subgrupos, envasados no grupo de factores *marcas precedentes*: (i) o subgrupo das cadeias sintagmáticas iniciadas por item com marcação formal de PL; (ii) o subgrupo das cadeias sintagmáticas iniciadas por item sem marcação formal de PL; (iii) o subgrupo das cadeias sintagmáticas iniciadas por item com marcação semântica de PL (numeral). No final, levaram-se à codificação os 30 apresentados na Tabela 12.3.

**Tabela 12.3.** Factores constituídos para a variável independente *marcas precedentes ao item analisado*.

<b>4) Variável independente <i>marcas precedentes ao item analisado</i></b>				
<b>Cód.</b>	<b>Factor</b>	<b>Exemplo</b>	<b>Nr. de ocorrências</b>	<b>%</b>
/	1 <sup>a</sup> Posição/Ausência de marca precedente:			
/		ALGUNS <b>médicos</b> [CASTEH1]	-	-
<b>Cadeias sintagmáticas iniciadas por item com marcação formal de plural</b>				
<b>P</b>	2 <sup>a</sup> posição, marca formal na 1 <sup>a</sup> posição: S	vários <b>PONTOS</b> [CASTEH1]	222/640	35
<b>Z</b>	3 <sup>a</sup> posição, 2 marcas formais precedentes: SS	os meus <b>PAI</b> [OSVALH1]	(14/26)	54
<b>S</b>	SSM	uns <b>riscos</b> assim AMARELO [CARMOM1]	(0/1)	0
<b>T</b>	S0	esses <b>bicho</b> PEQUENO [OSVALH1]	0/30	0
<b>4</b>	S00	esses <b>peixe</b> gordo GRANDE [MANOH3]	(0/5)	0
<b>6</b>	S000	esses peixe gordo grande COMPRIDO [MANOH3]	(0/4)	0
<b>W</b>	S0M	essas <b>galinha</b> memo NATURAL [CLOTIM2]	(1/4)	25
<b>C</b>	SM	<b>jogadores</b> mais INTERESSADO [CASTEH1]	(2/5)	40
<b>M</b>	SMS	tantos e tantas <b>INFORMAÇÕES</b> [CASTEH1]	(1/1)	100
<b>2</b>	SN	desses dois <b>NORMAL</b> [OSVALH1]	(7/22)	32
<b>7</b>	SN0	umas onze <b>jovem</b> FORMADO [CASTEH1]	(0/1)	0
<b>Cadeias sintagmáticas iniciadas por item sem marcação formal de plural</b>				
<b>B</b>	2 <sup>a</sup> posição, zero na 1 <sup>a</sup> posição: 0	a <b>RAZÕES</b> [CASTEH1]	39/46	85
<b>3</b>	00	o seu <b>DONOS</b> [ABILH2]	(3/3)	100
<b>0</b>	0S	<b>rocha</b> todas ESSAS [OSVALH1]	(4/7)	57
<b>i</b>	0SS	uma otras minhas <b>PREOCUPAÇÕES</b> [CASTEH1]	(1/1)	100
<b>b</b>	0N	todo cinco <b>LATA</b> [ZECAH1]	(0/13)	0
<b>9</b>	0NN(N)0	esse vinte e cinco <b>metro</b> QUADRADO [ZECAH1]	(1/2)	50
<b>Cadeias sintagmáticas iniciadas por item com marcação semântica de plural</b>				
<b>N</b>	2 <sup>a</sup> posição, numeral sem -s na 1 <sup>a</sup> : N	vinte <b>ANO</b> [CASTEH1]	121/412	29
<b>R</b>	2 <sup>a</sup> posição, numeral em -s na 1 <sup>a</sup> : Ns	duas <b>MULHERES</b> [CASTEH1]	79/278	28
<b>\$</b>	NS	cinquenta <b>metros</b> QUADRADO [ABILH2]	(0/1)	0
<b>1</b>	NsS	[-]	(0)	-
<b>5</b>	N0	cem <b>metro</b> QUADRADOS [ABILH2]	(1/2)	50
<b>@</b>	Ns0	Dois <b>filho</b> MULHER [FRANCM2]	(0/5)	0
<b>g</b>	NM	noventa tal <b>CONTOS</b> [ZECAH1]	(2/5)	40
<b>8</b>	NsM	[-]	(0)	-
<b>a</b>	NMM	[-]	(0)	-
<b>e</b>	NsMM	duzentos e tal <b>CONTO</b> [SALVH3]	(0/1)	0
<b>#</b>	NN(N)	quarenta e cinco mil <b>DOBRAS</b> [OSVALH1]	14/49	29
<b>%</b>	NNs(N)	trinta e três <b>ANO</b> [CLOTIM2]	(2/9)	22
<b>Totais:</b>			514/1.573	33

**Obs.:**

- (i) Símbolos aplicados aos factores (coluna Factor): S = item com marcação formal de plural; 0 = item sem marcação formal de plural; N = numeral não terminado em -s; Ns = numeral terminado em -s; M = item que não possui flexão de plural;
- (ii) Símbolo / (Coluna Cód.): situação em que o critério não se aplica;
- (iii) Símbolo [-] (Coluna Exemplo): factor constituído, mas que se constatou não apresentar qualquer ocorrência no nosso *corpus*.

A maioria dos factores regista ocorrências mínimas e alguns não apresentam quaisquer realizações. No capítulo 4, dedicado à análise dos grupos de factores do tipo estrutural, voltaremos a debruçar-nos sobre os trabalhos que estudam esta variável (ponto 4.1.2.4). Iremos tratar os nossos dados de diversas formas, seguindo metodologias aplicadas por outros autores (Scherre, 1988; Lopes, 2001; P. Andrade, 2003; Baxter, 2009), a fim compararmos os resultados destes trabalhos com os nossos. Ao constituirmos esta variável independente e seus factores buscamos confirmar a seguinte hipótese:

**Hipótese 3** – *Marcas precedentes*: a presença de marca PL em elemento anterior ao item analisado inibe a marcação neste. Portanto, os itens tenderão a exibir inibição da marcação caso sejam antecidos de quantificadores e determinantes que apresentem marca formal de PL, bem como de numeral, que já possui informação semântica de número intrínseca.

### **3.6.3. Variável independente *contexto fonológico posterior***

É já dado adquirido em trabalhos de linguística que as variáveis fonológicas podem ser influenciadas pelos contextos que se lhes seguem. Segundo Scherre (1988:241), é sustentável inferir três predições no que concerne a este tipo de influência:

- (i) a busca do padrão silábico CV provoca uma influência directa na ausência de concordância PL, uma vez que esta se realiza, na grande maioria dos casos, com recurso à inserção do morfema *-s*, isto é, com adição de um item consonantal em final de palavra;
- (ii) face ao pressuposto anterior, também é lógico predizer que uma vogal em contexto seguinte favorece a inserção da marca formal de PL, porque estarão criadas as condições para ocorrer o fenómeno de elisão. Este, facilita a ligação entre a consoante final *-s* e a vogal inicial da primeira sílaba do item posterior, de forma a originar a sílaba padrão CV;
- (iii) se a vogal favorece a inserção PL, é crível também que o oposto possa suceder, isto é, que o contexto posterior representado por consoante iniba a marcação, uma vez que o fenómeno da elisão deixa de se poder realizar (Guy, 1981a:27-29).

Contudo, diversos trabalhos têm evidenciado que os efeitos produzidos pela realização de consoante ou vogal em contexto posterior não se têm revelado significativamente diferentes para a inserção ou não do *-s* marca de PL no elemento antecedente. Scherre (1978:7-8) havia constatado que algumas consoantes em contexto seguinte revelavam tendência para favorecerem a ocorrência de marca PL em nomes terminados em *-s* ou *-r*, ou seja, substantivos que realizam flexão total ou parcial em número, no primeiro caso com a inserção final de *-es* (*meses* > *mulheres*) e no segundo caso com a inserção final de *-e* (*mese* > *mulhere*). Estes aspectos levaram Scherre (1988:243) a analisar os elementos em contexto posterior em termos de traços, principalmente as consoantes, efectuando pesquisas na linha de Braga (1977) e Guy (1981a). Braga (1977:71-72) havia concluído que os itens palatais condicionam fortemente a marcação, em virtude da sua aproximação fónica com os segmentos de PL. Guy (1981a:146), por seu lado, debruçou-se sobre o efeito da sonoridade e do ponto de articulação das consoantes na inserção do morfema *-s* em itens regulares, constatando que as velares favorecem mais esta do que as labiais e as alveolares (e/ou dentais). Ao mesmo tempo, verificou que a tendência para a marcação decresce com a ocorrência de traço sonoro e aumenta com a existência de traço surdo.

Nos seus trabalhos, Scherre (1978, 1988) agrupou velares e labiais, inviabilizando, assim, comparações com o trabalho de Braga (1977). Paralelamente, como no português é relevante o fenómeno da neutralização, Scherre (1978, 1988) absteve-se de analisar a influência das palatais na marcação PL. Contudo, e tal como Guy (1981a), observou que esta é inibida pelos segmentos consonantais sonoros e favorecida pelos surdos, tendo considerado normal que tal aconteça, uma vez que os segundos ocorrem com mais frequência nas línguas.<sup>144</sup> Simultaneamente, Scherre (1988:255) constatou ausência de efeito na oposição consoante/vogal, mas ocorrendo efeito positivo da consoante e negativo da vogal nos itens terminados em *-s* (*mês* > *meses*).

O efeito da pausa tem sido também analisado em estudos sobre inserção da marca formal de PL, com resultados diversificados, como adiante se verá. Face a esta divergência de conclusões, Scherre (1988:245) incluiu também o factor pausa nos seus estudos sobre a influência do contexto fonético/fonológico posterior na pluralização. Nas suas observações, testou várias hipóteses de agrupamentos de factores: (i) constituindo

apenas quatro factores, que implicam as oposições maiores (vogal/consoante e pausa interna/pausa final)<sup>145</sup>; (ii) separando os itens lexicais em três grupos (regulares, terminados em *-s* e finalizados em *-r*); (iii) analisando detalhadamente o conjunto dos itens regulares em termos de segmentos e traços das consoantes, de acordo com o ponto de articulação, a sonoridade e as caixas de ressonância. Os resultados levaram Scherre (1988:255) a concluir que não só não existe busca do padrão silábico universal CV como também ocorre o fenómeno da haplologia sintáctica<sup>146</sup>. Paralelamente, regista-se uma outra oposição de acção na marcação: a pausa,<sup>147</sup> peguilhando o preconizado por Poplack (1980a:61) e confirmando o defendido por Guy (1981a:166), exerce influência na inserção da marca formal PL nos itens regulares (menos acentuada) e nos elementos terminados em *-s* (mais acentuada); por seu lado, os traços dos segmentos (ponto de articulação, sonoridade e caixas de ressonância), tal como já haviam advogado Braga (1977:40) e Guy (1981a:166), exercem fraca influência na marcação, apesar de as surdas propiciarem um pouco mais a inserção das marcas PL do que as sonoras.<sup>148</sup>

Lopes (2001:225) observou igualmente o *contexto fonológico seguinte* a cada elemento do SN e a sua interferência na presença de marca PL na NURC. Inicialmente, trabalhando também com os quatro factores que envolvem as oposições maiores (vogal/consoante e pausa interna/pausa final), a autora optou por manter sempre dois factores para variável pausa, a fim de não ferir os resultados finais. Estes revelaram que a pausa final é o factor que mais favorece a marcação, que a pausa interna e a vogal exercem um peso neutro na mesma, e que a ocorrência de consoante a desfavorece ligeiramente. Contudo a proximidade de valores entre as variáveis pausa interna, vogal e consoante poderia mesmo sugerir a constituição de um grupo de factores binário: pausa final e pausa interna/vogal/consoante.

Após obtenção destes valores, Lopes (2001:228-239) avaliou também a influência na marcação em termos do traço de sonoridade consonantal, constituindo, para o efeito, duas variáveis: itens regulares e itens terminados em *-s*. A exemplo da metodologia levada a cabo para outros grupos de factores, a autora observou depois o efeito das variáveis na marcação PL das seguintes variáveis sociais: *português popular*, *português universitário*, *escolaridade (Primária, Média e Universitária)*, *sobrenomes religiosos e sobrenomes não-religiosos*. Os resultados não se revelaram uniformes para todas as

variáveis independentes, mas, ainda assim, foi possível concluir que, por exemplo, o fenómeno da haploglia sintáctica (Scherre, 1988:244) actua também nos dados de Lopes (2001:239), com o contexto posterior vocálico desfavorecendo a marcação nos itens em *-s*. Mais elementos sobre as conclusões deste estudo serão apresentados aquando da comparação dos nossos resultados com os de outros trabalhos (capítulo 5).

No PT, Baxter (2004:113-115) estabeleceu cinco factores para analisar a influência da variável independente *contexto fonológico posterior* na marcação PL do SN: consoantes vozeadas, consoantes surdas, vogal, pausa interna e pausa final. Os achados revelam que a pausa final é o elemento que mais influencia a marcação PL na fala da comunidade, observando-se esta tendência em todas as gerações. Seguem-se-lhe, com pesos quase iguais, a pausa interna e as consoantes surdas, embora estes factores revelem tendência para condicionarem mais a marcação nas gerações menos idosas. As vogais e as consoantes sonoras, por seu lado, apresentam resultados inibidores da marcação, apesar de as primeiras se estabelecerem num patamar praticamente neutro. Baxter (2004:111) atribui a responsabilidade dos vários constrangimentos fonológicos verificados no PT às tendências de aquisição e influências de transferência que ocorrem no PtgL2, via DLP's fornecidos à geração mais idosa pelos falantes adultos de L1 africanas.

Ainda relativamente ao grupo de factores *contexto fonológico posterior*, não poderemos levar a cabo comparações quer com as variedades africanas do Maputo e de Cabo Verde quer com o dialecto de HEL-Ba. De facto, esta variável independente foi excluída das análises de Jon-And (2008; 2009) por, na fala dos maputenses e dos cabo-verdianos se revelar exiguamente influente na marcação PL do SN. Por seu lado, P. Andrade (2003) não constituiu esta variável para as suas análises sobre o dialecto brasileiro em questão.

Quanto ao nosso *corpus*, se o compararmos com os de Scherre (1988) ou Lopes (2001), revela-se quantitativamente mais modesto. Assim sendo, este aspecto inviabiliza de imediato a possibilidade de se constituir um elevado número de factores em termos de segmentos e traços para a variável independente *contexto fonológico posterior*, sob pena de dispersarmos muito os nossos dados e vermos alguns dos factores rejeitados por insuficiência de ocorrências. Ainda assim, constituímos um grupo de factores inicial que

contemplasse todas as possibilidades de ocorrências no *contexto fonológico seguinte* ao do item sobre o qual incide a marcação PL.

**Tabela 12.4.** Factores constituídos para a variável independente *contexto fonológico posterior*.

<b>5) Variável independente contexto fonológico posterior</b>				
<b>Cód.</b>	<b>Factor</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
F	Pausa final	esses <b>HOMENS</b> . [CASTEH1]	134/294	46
'	Pausa interna	umas nove ou onze jovem <b>FORMADO</b> [CASTEH1]	106/349	30
V	Vogal	umas <b>VOLTA</b> e recebe [CASTEH1]	234/409	57
P	P	OS <b>pais</b> [CASTEH1]	121/197	61
X	Ch	Tem UNS <b>cheros</b> [CARMOM1]	(2/4)	50
K	K	dois <b>DIAS</b> conosco [CASTEH1]	216/312	69
R	R	A <b>razões</b> [CASTEH1]	(13/17)	76
B	B	BOAS <b>barbatana</b> [OSVALH1]	29/34	85
Z	Z	ESSAS <b>zona?</b> [OSVALH1]	(8/8)	100
D	D	um outro <b>TRABALHOS</b> de televisão [CASTEH1]	79/186	42
-	F	umas nove ou onze <b>JOVEM</b> formado [CASTEH1]	34/48	71
T	T	um <b>OUTRO trabalhos</b> [CASTEH1]	51/101	51
J	J	nos <b>PRIMEIROS jogos</b> [CASTEH1]	16/38	42
L	L	ESSAS <b>lomba</b> [OSVALH1]	19/38	50
G	G	ESSES <b>garoupa</b> assim [ZECAH1]	(6/16)	38
M	M	uma <b>OTRAS</b> minhas <b>preocupações</b> [CASTEH1]	51/86	59
N	N	UMAS nove ou onze <b>jovem</b> formado [CASTEH1]	23/66	35
S	S	<b>BRAÇOS</b> suficiente [CASTEH1]	46/102	45
=	V	UMAS <b>volta</b> [CASTEH1]	14/35	40
			<b>Totais:</b> 1.202/2.340	51

**Obs.:** De acordo com a metodologia de Lopes (2001), considerámos pausa interna a distância da realização fónica que, durante a produção oral, não admite interferências entre o som do constituinte em análise e o som do item seguinte. Por seu lado, a pausa final marca o fim de uma frase ou período.

Convém ainda referir que não optámos pela metodologia de Lopes (2001:226), já que esta entendeu não considerar os “*contextos posteriores /s/, /z/, /x/, /j/*, cujos dados foram desprezados, diante da dificuldade de detectar a real presença ou ausência da

*marca de plural no elemento anterior*”. De facto, entendemos que, nos nossos dados, é possível detectar tais distinções, como o demonstram os exemplos que coligimos e apresentamos na tabela anterior, a qual exhibe ainda os factores por nós constituídos e exemplos das suas realizações, extraídas do nosso *corpus*.

Para a variável *contexto fonológico posterior* consideramos a seguinte hipótese:

**Hipótese 4 – Contexto fonológico posterior:** sendo o PA uma língua de padrão silábico CV, serão mais marcados os itens cujos contextos subsequentes configurem segmentos vocálicos ou pausa final. Por outro lado, caso o contexto subsequente seja consonantal, então o favorecimento da inserção de marcas será mais facilitado pelas consoantes surdas do que pelas consoantes sonoras.

#### **3.6.4. Variável independente *classe gramatical do item analisado***

Os trabalhos sobre a influência da variável independente *classe gramatical* na variação de marcas de PL no SN registam já mais de duas décadas e tiveram o seu pioneirismo em estudiosos como Cedergren (1973:43), quando esta se debruçou sobre o espanhol do Panamá e concluiu que os substantivos e adjectivos retêm a marcação, enquanto os determinantes a desfavorecem. Poplack (1980a:60), estudando o PL do EPR, chegaria, no entanto, a conclusões distintas, confirmando que, nos seus dados, os substantivos revelavam um índice de marcação ligeiramente acima do dos adjectivos. Por seu lado, Lefebvre (1981:80) analisa igualmente o efeito de determinadas classes gramaticais (nomes, pronomes e interrogativos) na marcação PL do quechua de Cuzco, um dialecto oficial peruano, chegando à conclusão que o grupo de factores *classe gramatical* não exerce influência sobre a mesma. Não obstante, os estudos de Lefebvre (1981) não nos colocam face ao fenómeno da concordância em número, que não existe no quechua de Cuzco, mas simplesmente em presença de marcas de PL em nomes e pronomes.

Posteriormente, Scherre (1988:148-151) faria um estudo minucioso sobre a influência da *classe gramatical* dos itens do SN na marcação PL deste. Categorizando-os detalhadamente e efectuando, pela primeira vez, uma análise que considerava esta variável de modo separado, primeiro, e inter-relacionada com as variáveis independentes



*posição linear e marcas precedentes*, depois, Scherre (1988:151-167) comprovou que estes três grupos de factores não dão conta dos fenómenos linguísticos na sua totalidade, caso sejam observados isoladamente. Como tal, apenas a convergência entre as diferentes variáveis conseguiria fornecer explicações viáveis acerca do funcionamento das mesmas na configuração da pluralização no SN. E, ao efectuar tal, Scherre (1988) viria a não concordar com o pressuposto avançado por Guy (1981b:179), quando este postulava que os efeitos da *classe gramatical e posição linear* eram idênticos em termos de influência na marcação PL. Assim, e se Guy (1981b:179) preconizava que a equivalência na marcação se regista a nível dos determinantes e primeira posição, dos substantivos e segunda posição e dos adjectivos e terceira posição, Scherre (1988:154) buscou evidenciar que, nos seus dados, os nomes ocupam a terceira posição numa frequência superior à dos adjectivos, que estes podem igualmente aparecer na segunda posição (na qual são mais marcados) e que os possessivos e restantes elementos não-nucleares surgem, amiúde, na segunda posição. Paralelamente, a autora constatou também que a primeira posição é sempre a mais marcada, independentemente da categoria gramatical do item que a está a ocupar.

Nas análises levadas a efeito para a NURC, Lopes (2001:172-175) buscou confirmar se ocorria também uma relação directa entre *posição linear e classe gramatical* nos moldes de Guy (1981b:179), tendo concluído que este tipo de associação não é acolhida no dialecto de Salvador, apesar de os artigos serem bastante marcados, por ocorrerem, normalmente, na primeira posição. Em adição, Lopes (2001) apresenta uma escala de marcação/inibição das categorias gramaticais que abordaremos em etapa posterior do nosso trabalho, quando compararmos os resultados do PA com os da NURC (capítulo 4).

P. Andrade (2003), apesar de ver a variável independente *classe gramatical* rejeitada nas suas rodadas VARBRUL, optou atender ao princípio sugerido por Scherre (1988) de que este grupo de factores não fornece dados conclusivos, caso seja observado individualmente. Assim sendo, P. Andrade (2003:95-98) cruzou a variável *classe gramatical* com a variável independente *posição linear*, confirmando que o efeito idêntico entre os determinantes e primeira posição se ajustava também aos seus dados, com esta categoria gramatical a evidenciar um número de marcas PL superior ao dos substantivos ou adjectivos na mesma posição. Ainda assim, a alta incidência de marcas

na primeira posição confirmava, em pleno, o pressuposto de Guy (1981a:301-302) acerca da pluralização do PB ocorrer na primeira oportunidade, levantando a possibilidade da sua antecedência crioula. P. Andrade (2003:98) considerou ainda a possibilidade de este tipo de ocorrência ancorar nos princípios psicolinguísticos que advogam uma ordenação na aquisição dos morfemas que estruturam os enunciados, isto é, no Modelo dos Quatro Morfemas (Myers-Scotton & Jake, 2000a, 2000b; Myers-Scotton, 2002). Como os morfemas de PL encaixam no grupo dos sistemas morfêmicos que configuram a classe dos *early system morphemes*, são passados, nas línguas em contacto, via transmissão linguística irregular, o que não acontece com a concordância, um processo que exige a intervenção tardia de outro tipo de morfemas sistêmicos, os *bridge system morphemes*.

Baxter (2004), para o PT, e Jon-And (2008), para o PMP, não apresentaram resultados para a influência das classes gramaticais na marcação PL do SN, tendo-se limitado a analisar a relação entre *posição linear* e *posição em relação ao núcleo*. A segunda justificou o facto de não ter analisado a variável *classe gramatical* em virtude de a mesma apresentar interferências com outros grupos de factores nos seus dados. Não obstante, Jon-And (2009:4) estudou o comportamento desta variável independente no PCV, observando que os definidos apresentam um peso favorecedor da marcação superior aos dos adjectivos e nomes. Ainda assim, os valores aproximam-se todos do patamar da neutralidade, levando Jon-And (2009:7) a concluir que a *classe gramatical* exerce efeito forte na marcação PL, possivelmente devido a transferências provenientes do CCV.

Refira-se ainda que os considerando que tecemos em torno dos resultados de Baxter (2004) e Jon-And (2008, 2009) serão fornecidos no capítulo 4, quando observarmos os resultados da variável independente *posição linear*, nomeadamente na parte em que se aborda a relação que a esta estabelece com outros grupos de factores.

#### **3.6.4.1. Constituição e definição dos factores para a variável *classe gramatical***

Para a nossa análise constituímos os factores constantes na tabela do grupo de factores *classe gramatical* (Tabela 12.5), tendo como base a linha de critérios da gramática tradicional, igualmente adoptada por Scherre (1988:150-151) e por Lopes

(2001:167). Assim, optámos também por classificar os adjectivos em duas classes, uma vez que determinados itens, tipo *mesmo* ou *próprio*, podem ser vistos como adjectivos ou indefinidos, enquanto outros desempenham unicamente a função de adjectivos. Os segundos integram então o factor denominado *adjectivo 1*, enquanto os primeiros foram codificados como pertencentes ao factor *adjectivo 2*. Relativamente aos graus dos adjectivos e substantivos, no nosso *corpus* foram detectadas 8 ocorrências dos primeiros, todas no grau diminutivo, e nenhuma dos segundos.

O factor *categoria substantivada* engloba todo o tipo de classes gramaticais (adjectivos, possessivos, indefinidos ou numerais), conquanto substituam o nome na sua função de núcleo do SN. A este propósito, convém mencionar que, para a codificação dos nossos factores, também não foi estabelecido um critério diferencial entre determinantes e pronomes. Assim, pronomes substantivos ou absolutos, como o pronome indefinido do exemplo [249], foram tratados como categoria substantivada, visto desempenharem a função de núcleo do SN:

[249] PA:  
*Eu ouvi os outros a contar.* [OSVALH1]

Nos casos como o do exemplo [250], em que estamos perante pronomes adjectivos ou adjuntos que acompanham o substantivo (os apelidados de determinantes, visto determinarem a extensão do significado do nome), optámos por designá-los simplesmente como possessivos, demonstrativos ou indefinidos.<sup>149</sup>

[250] PA:  
*Como é vejo outros amigo a fazê, ê també faço.* [OSVALH1]  
“como eu vejo os outros amigos a fazerem, eu também faço”

No que diz respeito aos artigos, Scherre (1988:150) optou por considerá-los conjuntamente com os demonstrativos, mas Lopes (2001:83-84) separou as categorias, tratando os artigos definidos e indefinidos num único factor.<sup>150</sup> Nos nossos dados seguimos a metodologia de P. Andrade (2003:96), que tratou os artigos definidos e indefinidos separadamente, a fim de observar qual das categorias exercia mais peso na regra de concordância PL. Contudo, é importante salientar que, no nosso *corpus*, bastantes vezes nos vimos confrontados com dúvidas na forma de tratar os artigos indefinidos. Se por um lado adoptámos o critério de Lopes (2001:183) no que concerne à

distinção dos pronomes e artigos indefinidos “um” e “uns”, por outro lado interessava destringir também se estávamos na presença do numeral ou do artigo. O exemplo seguinte pode fornecer uma ideia sobre a dificuldade que determinados SN’s colocaram para efeitos de codificação desta categoria, uma vez que nem o próprio contexto permitia distinguir com exactidão o que o informante pretendia transmitir:

[251] PA:  
*tinha um dois home trazia pão aqui.* [MMDEUSM3]  
“havia uns dois homens que traziam o pão aqui”  
“havia um ou dois homens que traziam o pão aqui”

Como se verifica, caso o item “um” seja tratado como artigo indefinido, o SN incorpora dois elementos à codificação, ambos sem marca formal de PL (artigo e nome núcleo do SN). Contrariamente, se o mesmo elemento for tratado como numeral, o SN apenas apresenta um item sujeito à codificação, também sem marca formal de PL (o nome núcleo do SN). Esclarece-se apenas que, no caso presente (e outros semelhantes), decidimos tratar o SN de acordo com o número e ordem original de constituintes na cadeia sintagmática, o que pressupõe, neste caso, a primeira possibilidade, isto é, estarmos perante um indefinido. Este aspecto é tanto mais compreensível, se levarmos também em linha de conta as questões debatidas no ponto 1.3.2.2, acerca do determinante indefinido *ũa*, do santomense (exemplos [36] e [37]), e que transporta as especificações de número e definitude (Alexandre & Hagemeyer, 2007:40). De facto, e ao que tudo parece indicar, a realização do exemplo [251] poderá indiciar uma interferência do substrato, representando o indefinido “alguns”, oposto de “um”.

Relativamente ao numeral, o mesmo não foi tratado por Scherre (1988) nem por P. Andrade (2003). Apenas Lopes (2001:84) se debruçou sobre esta categoria, incluindo ordinais, multiplicativos e fraccionários num único factor. Por razões óbvias, que se prendem com a não flexão dos cardinais em número, a autora não considerou estes para efeitos de análise. Para o nosso estudo, seguimos exactamente a metodologia adoptada por Lopes (2001), embora em alguns casos, como aquele que é reproduzido no exemplo [239], fôssemos obrigados a considerar a existência de flexão em número em determinados cardinais, quando as realizações assim o exigiam.

Quanto aos possessivos e demonstrativos, foram tratados de acordo com os princípios advogados pela gramática normativa. Nos primeiros, foram incluídas também as formas de pronome pessoal complemento directo contraídas com a preposição *de* (*dele, dela, deles, delas*), visto serem categorizadas como determinantes possessivos pós-verbais.<sup>151</sup> Relativamente aos segundos, as formas variáveis são substituídas, por vezes, pela forma invariável *isso*, que pode surgir flexionada em número PL, como acontece no exemplo [219].

Ainda sobre este item gramatical, Baxter (2004:108-109) entende que o mesmo poderá estar a desempenhar a função do artigo definido do português. De facto, observa-se, tanto no PT como no PA, a existência de formas inovadoras que actuam no lugar do referido artigo, como acontece não só com o exemplo [219] mas também quando o artigo surge contraído com preposições, situação a que fizemos já referência no ponto 1.3.2.3.1 do presente trabalho:

[252] PA:  
*fica com dinheiro três ou quatro dia ni mão* [LUISH2]  
“fica com o dinheiro três ou quatro dias na mão”

No que respeita aos indefinidos, foi nosso critério subdividi-los em dois factores: (i) o dos indefinidos propriamente ditos, englobando os itens que se aplicam à terceira pessoa gramatical, quando esta é considerada de um modo vago e indeterminado (factor H, Tabela 12.5); (ii) o dos quantificadores, que incluem os elementos indicativos de proporções numéricas indeterminadas (factor Q, Tabela 12.5).<sup>152</sup> Perante a dificuldade levantada por alguns indefinidos, detentores de características que os podem englobar em qualquer das subdivisões, como acontece com o item *vários/várias*, ao qual fizemos referência a propósito do exemplo [247], tentámos seguir um critério uniformizado, optando por englobá-los num único factor. Assim, no que concerne ao item em questão, decidimos classificá-lo como indefinido e não como quantificador, porque, seja qual for a sua situação de uso, possui sempre o traço que estabelece a distinção, a diferença, em detrimento do traço de quantidade, que pode ser apagado.

A propósito do quantificador invariável *tudo*, o mesmo regista realizações que reflectem uma ancoragem nas formas crioulas, nas quais é usado em vez do quantificador variável *todo*, com excepção para as ocorrências fossilizadas que

concorrem conjuntamente com numerais. Assim, e a exemplo do que sucedeu com o demonstrativo invariável *isso*, o quantificador *tudo* recebeu, por vezes, o tratamento de quantificador variável, quando o contexto assim o determinava:

[253] PA:  
*tudos os dia eu assistia televisão* [OSVALH1]  
 “todos os dias eu assistia à televisão”

A Tabela 12.5 exhibe os factores constituídos para a variável independente *classe gramatical*, seus códigos, exemplos, número de realizações e respectivos percentuais de marcação PL:

**Tabela 12.5.** Factores constituídos para a variável independente *classe gramatical do item analisado*.

6) Variável independente <i>classe gramatical do item analisado</i>				
Cód.	Classe gramatical	Exemplo	Nr. de Ocorrências	%
N	Substantivo grau normal	<b>TRABALHOS</b> privado [CASTEH1]	480/1.428	34
Y	Substantivo grau diminutivo	[-]	-	-
W	Substantivo grau aumentativo	[-]	-	-
C	Categoria substantivada	os <b>OTROS</b> [OSVALH1]	17/54	32
A	Adjectivo 1	<b>trabalhos</b> PRIVADO [CASTEH1]	19/65	29
D	Adjectivo 2	esses <b>jogadores</b> MESMO [ANTOM1]	(0/8)	0
Z	Adjectivo grau diminutivo	Esses <b>peixe</b> PIQUININO [ZECAH1]	(0/8)	0
“	Adjectivo grau aumentativo		-	-
G	Artigo definido	A <b>razões</b> [CASTEH1]	260/288	90
H	Artigo indefinido	UMAS <b>volta</b> [CASTEH1]	63/80	79
L	Numeral	nos PRIMEIROS <b>jogos</b> [CASTEH1]	(3/6)	50
O	Possessivo	MINHAS <b>preocupações</b> [CASTEH1]	31/37	84
V	Demonstrativo	ESSES <b>homens</b> [CASTEH1]	174/186	94
I	Indefinido	OTRAS minhas <b>preocupações</b> [CASTEH1]	62/64	97
Q	Quantificador	TODOS <b>jogadores</b> [CASTEH1]	93/116	80
R	Relativo	[-]	-	-
T	Pronome pessoal recto	[-]	-	-
<b>Totais:</b>			1.202/2.340	51

Refira-se que foram ainda constituídos dois factores para abarcarem os seguintes casos: pronomes pessoais rectos e relativos variáveis (p.e. *qual/quais*). Contudo, estes tipo de relativos não ocorre no PA (ver Tabela 12.5), confirmando-se a tendência para o padrão das relativas confluir para uma forma comum (o uso de “que”) nas variedades crioulezantes, um fenómeno que vem também afectando o PE. De facto, neste são já tidas como confirmadas quer a supressão do relativo “cujo” quer a estratégia cortadora que vem determinando o apagamento das preposições que ocorrem em paralelo com os relativos (Mateus & Cardeira, 2007:71), fenómenos que há muito se registam no Brasil e que determinam que, neste aspecto, o padrão de realização do PE esteja cada vez mais próximo dos do PA e PB.

Com base nas conclusões dos trabalhos a que nos referimos a propósito da variável independente *classe gramatical*, pretendemos confirmar a seguinte hipótese:

**Hipótese 5 – *Classe gramatical*:** os artigos, os quantificadores, os indefinidos, os demonstrativos e os possessivos configuram-se como os elementos gramaticais que tendem a receber mais marcas de PL, visto desempenharem funções de determinantes, isto é, ocorrerem, por norma, antes do núcleo do SN. As outras classes gramaticais propiciam-se menos à marcação, por configurarem a posição de núcleo do SN, ou se localizarem, preferencialmente, depois deste.

### **3.6.5. Variável independente *posição do item analisado em relação ao núcleo do SN***

Algumas variáveis do tipo estrutural, casos da *posição linear*, *posição em relação ao núcleo do SN*, *marcas precedentes* ou *classe gramatical*, têm-se revelado bastante importantes para se entenderem as motivações que subjazem ao fenómeno da CPL-var no SN. Aliás, se atentarmos à hierarquia dos grupos de factores do PA seleccionados como influentes na marcação PL (Tabela 14.1), constatamos que o topo da pirâmide é ocupado por este tipo de variáveis independentes, em detrimento das variáveis do tipo semântico ou fonológico, por exemplo. Paralelamente, apercebemo-nos também que, no vértice superior da pirâmide do PA, se encontra a variável independente *posição do item analisado em relação ao núcleo do SN*, considerada pela ferramenta VARBRUL como a mais importante para a inserção/inibição do número nos elementos do SN.

Refira-se que Scherre (1988) foi a primeira a tratar os grupos de factores *posição linear*, *marcas precedentes* e *classe gramatical* quer de forma individualizada quer cruzados entre si, já que, até então, os diferentes trabalhos efectuados sobre concordância variável apenas haviam privilegiado, de modo distinto, o efeito dos dois primeiros grupos (*posição linear* e *marcas precedentes*), sem analisarem a influência exercida pelo terceiro grupo (*classe gramatical*). Posteriormente, Lopes (2001:175-198) observou também o efeito exercido na marcação PL pela variável que se estuda neste ponto do nosso trabalho. Quanto a nós, optámos por incluir também uma outra variável de características estruturais nas nossas observações, a *ordem do item na cadeia dos constituintes flexionáveis*, cujo comportamento pensamos não ter sido ainda examinado em estudos de concordância em número. Pronunciar-nos-emos detalhadamente sobre esta variável no ponto 3.6.8.

Guy (1981b:179), analisando a pressão que os constituintes morfológicos, em determinada posição na cadeia sintagmática, exercem na marcação PL das produções dos falantes semiescolarizados cariocas, concluiu existir equivalência entre classe e sua localização na cadeia sintagmática. Scherre (1988:154) viria a discordar deste pressuposto, alegando, por exemplo, que todos os itens gramaticais surgem mais marcados na primeira posição no SN. Contudo, nenhum dos autores se debruçou sobre a variável independente *posição em relação ao núcleo do SN*, apesar de Scherre (1988:219) estabelecer uma oposição entre elementos não-nucleares à esquerda do núcleo nas primeira e segunda posições. A variável, como referimos, viria então a ser tratada por Lopes (2001:171-172), que, tal como Scherre (1988:154), defende existir uma frequência diferente de concordância para os itens da mesma classe, consoante a posição que ocupem na cadeia do SN. Estranhando o alto índice de marcação revelado pelos possessivos em segundo lugar na cadeia do SN, Lopes (2001:174-175) analisou a sua posição face ao núcleo do SN, concluindo que tal lugar na cadeia do SN poderia ser anterior ou posterior ao núcleo. Como tal, a marcação incidia mais sobre a posição pré-nuclear, em detrimento da posição pós-nuclear, mostrando que a variável independente *posição em relação ao núcleo do SN* é um elemento com forte influência na concordância PL. Face a esta evidência, Lopes (2001:175-177) decidiu então observar se os outros elementos não-nucleares revelavam o mesmo tipo de



comportamento que os possessivos, acabando por verificar que os itens possuem um índice muito maior de marcação quando estão localizados à esquerda núcleo. Paralelamente, Lopes (2001:177-178) constatou também que as posições lineares podem desfavorecer a marcação, a qual aumenta a partir da segunda posição. Para a autora, a explicação assenta, em parte, no modelo psicolinguístico dos Quatro Morfemas (Lopes, 2001:183-185), cuja ordem de emanção é análoga na aquisição de L1 ou ASL.

Baxter (2004:117, 2009:276), por seu lado, observando quer o PT falado por descendentes de africanos bantóides quer o dialecto de HEL-Ba, constatou também a tendência para a marcação PL favorecer mais o elemento não-nuclear imediatamente à esquerda do núcleo, enquanto os itens pós-nucleares produzidos pelas gerações mais idosas apresentavam uma quase total falta de marcação. Reanalizando o primeiro dialecto, Baxter (2009:278) apontou para ele um perfil de variação e funcionamento dos factores condicionantes estruturais semelhantes aos do PVB. Portanto, corrobora Guy (1981a:301-302) quanto aos mecanismos responsáveis pela variação da concordância PL no SN, quando o português é adquirido por falantes de determinadas L1's africanas. Nesta perspectiva, a mudança verificada no PT seria condicionada por factores que ancoram na aquisição do PtgL2 e PtgL1 adquiridos em situação que envolve a transmissão de componentes do PtgL2 influenciados pelo umbundu, a língua ancestral dos tongas. Como tal, verificava-se que a componente sociohistórica terá também desempenhado um papel importante, via DLP's do PtgL2, na aquisição do PtgL1 nativizado, leia-se PT. E, assim sendo, a hipótese avançada por Guy (1981a) de que os substratos africanos poderão constituir-se como a principal fonte da CPL-var do PVB, escorando-a a partir da posição imediatamente pré-nuclear, ganhava sustentação com os achados de Baxter (2004).

Porém, convém também adiantar que P. Andrade (2003:98), observando igualmente o dialecto de HEL-Ba, corroborara parcialmente Guy (1981b:179) no que concerne à equivalência entre classe e posição. De facto, a autora avançou que os determinantes revelam um elevado índice de marcação quando ocorrem na primeira posição e respaldou a explicação quer no *Modelo dos Quatro Morfemas*, já antes adiantada por Lopes (2001), quer na possibilidade de as línguas transplantadas de África para o Brasil poderem estar por trás do fenómeno que leva o PB a indicar o PL na primeira

oportunidade. Quanto às segunda e terceira posições serem preferencialmente ocupadas por substantivos e adjetivos, respectivamente, como advogara Guy (1981b:179), o facto é também corroborado por P. Andrade (2003:98), que avança ainda que estes lugares apresentam índices de marcação inferiores aos da primeira posição. Paralelamente, P. Andrade (2003:97) constataria ainda que o determinante *todo* é o único elemento gramatical que ocorre indistintamente em todas as posições, com o maior número de realizações a acontecer na terceira posição, mas a maior taxa de marcação a recair igualmente na primeira posição.

Jon-And (2008:5), por seu lado, apresenta o factor *elemento imediatamente anterior ao núcleo* como o único que marca positivamente o número, com todas as outras posições a revelarem-se inibidoras da marcação. A autora verifica ainda que esta variável sintáctica tem efeitos mais fortes no PMp do que no PT e no PVB, concluindo que a causa do facto deverá ancorar na influência que as L1's bantu exercem nos dados dos seus informantes, todos eles falantes do xironga ou xichangana, em contexto doméstico, e do português, em situações de actividade diária (Jon-And, 2008:7).

Para observar o comportamento da variável independente *posição em relação ao núcleo do SN* no PCV, Jon-And (2009:6) cruzou-a com o grupo de factores *posição linear*. Os resultados revelam também que os elementos em posição imediatamente anterior ao núcleo do SN favorecem bastante a marcação, enquanto a sua localização nas outras posições inibe a pluralização. Deste modo, Jon-And (2009:7) entende que a variável conjunta *posição em relação ao núcleo do SN/posição linear* revela efeitos semelhantes no PCV e outras variedades africanas e brasileiras de português.

Trabalhando com os dados do nosso *corpus*, Figueiredo (2008:28-29) havia também cruzado as variáveis independentes *posição linear* e *posição em relação ao núcleo do SN*, confirmando a existência de um número bastante diminuto de itens em posições pós-nucleares não imediatas. Este aspecto levou-o a criar um único factor que reúne todas as posições posteriores ao núcleo do SN, a fim de o mesmo ser seleccionado pela ferramenta VARBRUL e poder fornecer valores consistentes acerca do desenvolvimento geracional da CPL-var no SN do PA:

“Refira-se que, inicialmente, esta lista estava dividida em onze factores, mas, nas duas primeiras gerações, o número de itens com marca PL nas quarta e quinta posições no SN, sejam eles Núcleo sejam item plural à direita deste, é bastante diminuto. Este aspecto levou-nos a eliminar a 5ª. posição das nossas codificações, reduzindo a lista inicial para apenas sete padrões configuracionais com a criação de uma nova variável dependente: a que reúne todas as Posições Posteriores ao Núcleo” (Figueiredo, 2008:29).

Deste modo, também na presente pesquisa se optou pela constituição de um único factor para análise de todos os itens que ocorrem em posição ulterior ao núcleo do SN, integrando-o no grupo de factores constante na Tabela 12.6:

**Tabela 12.6.** Factores constituídos para a variável independente *posição do item analisado em relação ao núcleo do SN*.

<b>7) Variável independente <i>posição do item analisado em relação ao núcleo do SN</i></b>				
<b>Cód.</b>	<b>Posição</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
<b>0</b>	Núcleo	‘SOAS morta [DULCEM1]	497/1.482	34
<b>A</b>	Anterior ao núcleo, não imediatamente	UNS quatro <b>ano</b> [CELESH3]	57/87	66
<b>I</b>	Imediatamente anterior ao núcleo	ESSES <b>lugar</b> tudo [LUIH2]	631/681	93
<b>U</b>	Posterior ao núcleo	<b>homens</b> CAPAZES [CASTEH1]	17/90	19
<b>Totais:</b>			<b>1.202/2.340</b>	<b>51</b>

Face ao que foi exposto anteriormente, colocamos a seguinte hipótese para a variável *posição do item analisado em relação ao núcleo do SN*:

**Hipótese 6 – Posição do item analisado em relação ao núcleo do SN:** Os elementos antepostos ao núcleo do SN são mais marcados do que aqueles surgem pospostos a este. Entre os primeiros, a marcação é mais acentuada nos itens adjacentes ao núcleo.

### **3.6.6. Variável independente *posição do item na cadeia do SN (Posição linear)***

O grupo de factores *posição linear* permite analisar a frequência com que certo elemento, em determinada posição na cadeia do SN, recebe a marca PL, independentemente da sua categoria gramatical. Este grupo de factores, quando cruzado com outros grupos (caso dos grupos *classe gramatical* ou *posição do item em relação ao núcleo do SN*), deixa também perceber, claramente, a relação de equivalência e

influência na marcação de número que existe entre determinados elementos gramaticais e a posição por eles ocupada no SN. Deste modo, a constituição do grupo *posição do item na cadeia do SN*, em concomitância com outros grupos de factores, permitirá confirmar se o PA cumpre também o pressuposto, já avançado em anteriores estudos sobre a CPL-var no SN de variedades de português, de que os itens em posição pré-nuclear revelam maior pendor para favorecerem a marcação, por oposição aos itens localizados pós-nuclearmente, que tendem a inibi-la (Scherre, 1988; Lopes, 2001; P. Andrade, 2003; Baxter, 2004, 2009; Figueiredo, 2008; Jon-And, 2008, 2009). Por outro lado, os estudos têm também indicado que a primeira posição da cadeia sintagmática tende ser a mais marcada, por oposição às outras localizações, cuja propensão é de favorecerem menos a pluralização. Importa ainda referir que, nos nossos dados, de todas as variáveis independentes que observam a questão da posição estrutural, o grupo de factores que aqui se analisa é aquele que é considerado pela ferramenta VARBRUL como o menos relevante para a inserção de marcas de PL na cadeia sintagmática do PA.

Os trabalhos pioneiros sobre o efeito desta variável independente (Braga & Scherre, 1976) concluem já sobre a oposição de marcação entre primeira posição e outras localizações, com a primeira favorecendo a inserção de marcas e as outras desfavorecendo-a. Posteriormente, Braga (1977:38) analisa o comportamento geracional do mesmo grupo de factores e conclui que os falantes das classes baixa e média aplicam mais a regra da marcação na primeira posição, e que a mesma vai perdendo gradualmente a sua força nas posições seguintes.

Seguidamente, Scherre (1978:109) observa o comportamento da variável nos estratos escolares, concluindo que o índice de favorecimento evidencia duas direcções opostas, já que, em determinadas posições, a probabilidade de aplicação da regra sobe no grupo escolarizado e desce no grupo semi-escolarizado. Ponte (1979:162), no seu trabalho sobre o PPA, constata também que a primeira posição é o factor que mais favorece a marcação PL, decrescendo esta, sucessivamente, nas posições seguintes. Igual padrão de favorecimento e desfavorecimento é também detectado por Nina (1980) na micro-região de Bragantina, e por Guy (1981b:179), nos falantes semi-escolarizados cariocas. Este, contudo, afirma que os efeitos da *posição linear* e *classe gramatical* são idênticos, já que parece existir uma equivalência entre os determinantes e a primeira posição, os

nomes e a segunda posição e os adjectivos e a terceira posição. Assim sendo, será nessas posições que os citados elementos gramaticais apresentarão mais marcas.

Até aqui, todos os trabalhos referidos apontam para uma evidência que parece inquestionável: a variável independente *posição linear* é aquela que exerce uma influência mais uniforme sobre a regra da concordância de número nos itens do SN das diferentes variedades de português no Brasil. Por outra parte, parece confirmar-se também o Princípio Funcionalista de a tendência para a informação semântica relevante ser retida na estrutura de superfície (Kiparsky, 1972:195), levando a que a marcação ocorra preferencialmente na primeira posição sintagmática, tornando a pluralização desnecessária nos elementos situados noutras localizações, visto revelar-se redundante.

Scherre (1988:183), reanalisando o fenómeno da concordância PL no MRJ, foi a primeira a tratar os grupos de factores *posição linear*, *marcas precedentes* e *classe gramatical* em conjunto, em virtude de entender que o estudo isolado das mesmas não dá conta do fenómeno da concordância PL na sua totalidade. Simultaneamente, defendeu também que o português e o espanhol são duas línguas que marcam o PL da mesma forma, isto é, sem revelarem condicionamentos ao princípio funcionalista do tipo “kiparskyano”, de acordo com o detectado por Poplack (1980a) em estudos sobre o portorriquenho. Nesta perspectiva, o que motiva a pluralização do português são fenómenos que ancoram no Princípio do Processamento Paralelo (Scherre, 1988:208), ou seja, no pressuposto de que são as marcas que levam a marcas, por oposição a zeros, que geram mais zeros. Ainda assim, Scherre (1988:213) conclui que, em termos probabilísticos, é a primeira posição que surge sempre mais marcada. E assim sendo, não se confirma o pressuposto de Guy (1981b:179), quando estabeleceu uma relação directa entre posição linear e classe gramatical. Daí que este pressuposto tenha sido minuciosamente observado por Scherre (1988), que considerou o efeito apontado por Guy (1981b:179) como virtual, já que, nos dados do MRJ, a relação entre *classe gramatical* e *posição na cadeia sintagmática* apresentou valores que não confirmam a hipótese avançada por este autor.

Lopes (2001:165-167) observou o comportamento do grupo de factores *posição linear* recorrendo a cinco factores, idênticos aos que Scherre (1988) havia utilizado nas suas análises. Os resultados confirmaram o Princípio Funcionalista de que a primeira

posição é a mais favorecedora para a inserção de marcas de PL (Kiparsky, 1972:195), visto ser esta que atribui a informação nova de número ao SN. No entanto, os dados de Lopes (2001:180-181) não acompanham os resultados de estudos anteriores que revelaram uma diminuição progressiva da marcação à medida que os itens se afastavam da primeira posição, já que a quinta posição dos SN's da NURC revelou um ligeiro aumento do favorecimento relativamente às segunda, terceira e quarta posições. Sobre este aumento nos pronunciaremos mais adiante (capítulo 4), já que discordamos da metodologia adoptada por Lopes (2001) para a codificação dos seus dados.

Por seu lado, P. Andrade (2003:94-95), não estudou esta variável de modo isolado, tendo-a cruzado com o grupo de factores *posição em relação ao núcleo do SN*. Os resultados finais indicam que a fala de HEL-Ba não constitui excepção aos achados anteriores, já que os seus falantes pluralizam mais os elementos pré-nucleares e praticamente não inserem marcas nos elementos situados depois do núcleo do SN. P. Andrade (2003:96) testou também a marcação PL em cada classe gramatical considerando a sua posição no SN, concluindo que o efeito apontado por Guy (1981b:179) só pode ser atribuído à primeira posição, a mais marcada, e que apresenta, na maioria dos casos, um determinante a ocupá-la. Este aspecto permite, contudo, que se corrobore Guy (1981a:301-302) quando este afirma que a indicação do PL, no PB, tem tendência a ocorrer na primeira oportunidade, um padrão que não tem precedentes na história do PE.

Baxter (2004:115-117, 2009:278) adoptou também a metodologia que pratica o cruzamento entre *posição linear* e *posição em relação ao núcleo do SN*. Constituindo onze factores configuracionais,<sup>153</sup> o linguista estudou a influência dos mesmos na marcação PL do dialecto dos tongas (Baxter, 2004) e na marcação de número do PT e de HEL-Ba (Baxter, 2009), numa perspectiva geracional. Considerandos sobre os resultados de Baxter (2004, 2009) e comparação entre os mesmos e os valores do PA serão levados a cabo quando nos debruçarmos pormenorizadamente sobre os nossos resultados, já que testaremos também o efeito dos nossos grupos de factores na marcação PL de acordo com a mesma metodologia utilizada para os tongas e HEL-Ba (capítulo 4).

Jon-And (2008:5), para o PMp, seguiu a mesma proposta de constituição de factores de Baxter (2004), tendo elaborado nove factores<sup>154</sup> e apresentando resultados apenas para a comunidade. O factor elemento imediatamente anterior ao núcleo é o único que favorece a marcação PL nos dados de Jon-And (2008:7), revelando efeitos mais fortes nas produções do Maputo do que nas realizações dos tongas e do Brasil. A autora atribui o facto a uma possível influência das L1's africanas dos informantes maputenses.

Relativamente ao PCV, Jon-And (2009:6) optou por cruzar apenas os grupos de factores *posição em relação ao núcleo do SN* e *posição linear*. Os resultados apresentam a primeira posição como sendo aquela que mais favorece a marcação, mesmo quando esta é núcleo do SN. Como tal, confirma-se também, no PCV, o Princípio Funcionalista (Kiparsky, 1972) de caber a esta posição a atribuição da informação de número no SN.

Para a constituição dos factores que compõem o grupo de factores *posição linear*, adoptámos a metodologia seguida por Scherre (1988) e Lopes (2001), pelo que codificámos cinco posições, conforme Tabela 12.7.

**Tabela 12.7.** Factores constituídos para a variável independente *posição do item na cadeia do SN*.

<b>8) Variável independente <i>posição do item na cadeia do SN</i></b>				
<b>Cód.</b>	<b>Posição</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
<b>1</b>	Primeira posição na cadeia do SN	<b>TRABALHOS</b> privado [CASTEH1]	678/754	90
<b>2</b>	Segunda posição na cadeia do SN	Umas <b>DORES</b> [DULCEM1]	441/1.314	34
<b>3</b>	Terceira posição na cadeia do SN	Os meus <b>FILHOS</b> [MANOH3]	48/168	29
<b>4</b>	Quarta posição na cadeia do SN	sete ou oito <b>MESES</b> [CASTEH1]	29/83	35
<b>5</b>	Quinta ou outra posição na cadeia do SN	umas nove ou onze jovem <b>FORMADO</b> [CASTEH1]	6/21	29
<b>Totais:</b>			<b>1.202/2.340</b>	<b>51</b>

Através da influência exercida por esta variável independente na marcação PL do SN, pretendemos confirmar, ou não, a seguinte hipótese:

**Hipótese 7** – *Posição do item na cadeia do SN*: a primeira posição do SN é a que mais favorece a inserção de marcas. As outras posições tendem a marcar menos o PL, pelo facto de o traço de PL (a informação semanticamente relevante) já ter sido atribuído.

### **3.6.7. Variável independente *marcação de género***

A constituição da variável independente *marcação de género* tem como objectivo verificar se existe uma correlação entre a marcação de género e a marcação PL em cada constituinte do SN. Efectivamente, e dado que o morfema de género surge mais directamente conectado à estrutura mórfica do item analisado, em virtude de estar estreitamente ligado ao radical deste, é de esperar que a marcação PL, por ser posterior à do género, possa ser condicionada pela primeira quando ela tem que ocorrer.

Esta variável independente não foi observada nos trabalhos sobre CPL-var de Guy (1981a), Scherre (1988), Lopes (2001), P. Andrade (2003), Baxter (2004, 2009) e Jon-And (2008, 2009). Contudo, Lucchesi (2000a:264), em estudo sobre a concordância variável de género de HEL-Ba, advogou o Princípio da Coesão Estrutural (Haiman, 1983:782), constatando que, de facto, a marcação de género favorece categoricamente a marcação PL no constituinte analisado, enquanto a não-marcação de género se revela como um forte obstáculo à inserção do número. Na presente análise, pretendemos observar se, de facto, a marcação de género funciona como inibidora da marcação de número. Numa perspectiva generativista, a inserção de marca de género, por ocorrer primeiro (Müller, 1994a; Fransceschina, 2002, 2005; Corrêa, Name & Ferrari-Netto, 2004; Montrul, 2004; White *et alii*, 2004), serve como gatilho para despoletar também a marcação de número. Nesta conformidade, e conforme observado no ponto 2.12 do presente trabalho, o que se espera é que, caso ocorra marcação de número, aconteça também inserção de marca de género.

Em trabalho sobre a aquisição da concordância de PL no SN por aprendentes chineses de PtgL2 em contexto formal de ensino, Godinho (2005:345) observou o efeito desta variável na inserção da marcação PL e confirmou o Princípio de Coesão Estrutural (Haiman, 1983:782), visto que a não-expansão da cadeia da concordância nominal no SN inclui ambas as concordâncias, de género e de número, cujos traços nucleares pertencem à posição SNUM, uma categoria intermédia entre o SDET e o SN (Baxter,



Lucchesi & Guimarães, 1997:2). Deste modo, e do ponto de vista aquisicional, sendo a identificação dos traços de género e número conseguida através da concordância no SDET, a não-interpretação de um dos traços pode ter influência na não-interpretação do outro, o que implica que a categoria intermédia (SNUM) não esteja totalmente adquirida. E no caso dos aprendentes chineses, tal como nos falantes de L1's africanas do grupo níger-congo atlântico, a explicação poderá assentar no facto de a concordância nominal no português não configurar um parâmetro das suas L1's e, conforme preconizado pela FFFH, ocorrer falha no restabelecimento dos parâmetros (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), uma vez que as categorias funcionais deixam de estar disponibilizadas depois do período crítico do desenvolvimento. Por seu lado, também não deve ser descartada a possibilidade de, tal como advogam os defensores da FTFA, os falantes do PtgL2 deterem uma compreensão e um julgamento metalinguístico do género gramatical distinto daquele que acontece na LA, uma vez que determinadas línguas do grupo níger-congo atlântico, pelo facto de não possuírem género gramatical, fazem com que os seus falantes em processo de aquisição de uma L2 detenham já um conhecimento cognitivo-linguístico prévio que vai impedi-los de adquirirem completamente as propriedades relevantes da L2 (cf. Gonçalves, 2004:235; Franceschina, 2005:191).

A fim de elaborarmos os factores para o grupo de factores *marcação de género*, optámos por seguir três critérios distintos, de acordo com os seguintes parâmetros de flexão em género:

1. O item leva correcta marcação de género:

- a) itens biformes regulares, com correcta inserção de marcação (p.e. *trabalhos PRIVADO* [CASTEH1]);
- b) itens biformes irregulares, com correcta inserção de marcação (p.e. *BONS pescadores* [CELESH3]);
- c) itens uniformes com marca padrão de género e correcta inserção de marcação (p.e. *vários TRABALHOS* [CASTEH1]; *nas LOJAS* [CARMOM1]).

2. O item não leva correcta marcação de género:
- itens biformes regulares, com incorrecta inserção de marcação (p.e. *trabalhos CASEIRA* [CARMOM1]);
  - itens biformes irregulares, com incorrecta inserção de marcação (p.e. *ESSES fruta* [ANTOM1]);
  - itens uniformes com marca padrão de género e incorrecta inserção de marcação (p.e. *uns cem CONTAS*<sup>155</sup> [CELESH3]).
3. Itens uniformes sem marca padrão de género:
- itens uniformes com marcação única de género fornecida pelo determinante (p.e. *os MATERIAL* [ABILH2]; *essas POSSIBILIDADE* [CASTEH1]);
  - itens uniformes com aparente marca padrão de género, isto é, não correspondência entre género do determinante e terminação do item (p.e. *os DIAS* [CASTEH1]; *outras minhas PREOCUPAÇÕES* [CASTEH1]);
  - itens uniformes com dupla marcação de género fornecida pelo determinante (p.e. *todos JOVEM* [MMDEUSM3]).

A Tabela 12.8 apresenta os factores constituídos, suas codificações, número de ocorrências e percentuais de marcação:

**Tabela 12.8.** Factores constituídos para a variável independente *marcação de género*.

9) Variável independente <i>marcação de género</i>				
Cód.	Factor	Exemplos	Nr. de Ocorrências	%
c	O item leva correcta marcação de género	<b>trabalhos</b> PRIVADO [CASTEH1] BONS <b>pescadores</b> [CELESH3] vários <b>TRABALHOS</b> [CASTEH1] nas <b>LOJAS</b> [CARMOM1]	1.060/2.030	52
n	O item não leva correcta marcação de género	<b>trabalhos</b> CASEIRA [CARMOM1] ESSES <b>fruta</b> [ANTOM1] uns cem <b>CONTAS</b> [CELESH3]	38/59	64
/	Itens uniformes sem marca padrão de género	os <b>MATERIAL</b> [ABILH2] essas <b>POSSIBILIDADE</b> [CASTEH1] os <b>DIAS</b> [CASTEH1] outras minhas <b>PREOCUPAÇÕES</b> [CASTEH1] todos <b>JOVEM</b> [MMDEUSM3]	104/251	41
<b>Totais:</b>			1.202/2.340	51

**Obs.:** O factor marcação de género aplica-se a todos os elementos do SN, de acordo com o traço semântico do núcleo deste.

Para confirmar, ou não, os pressupostos do Princípio de Coesão Estrutural, levantámos a seguinte hipótese:

**Hipótese 8 – Marcação de género:** as probabilidades de não-marcação de número são maiores caso não ocorra marcação de género. Em contrapartida, a aquisição de marcação de género beneficia a marcação de número, já que os traços de género e número se encontram dentro do mesmo nó funcional SNUM.

### **3.6.8. Variável independente *ordem do item na cadeia dos constituintes flexionáveis em número***

Se o grupo de factores *posição linear* permite verificar a frequência com que certo item, independentemente da sua classe gramatical, recebe a marca de PL em determinada localização da cadeia sintagmática, já a variável independente *ordem do item na cadeia dos constituintes flexionáveis* surge, neste trabalho, com o intuito de observar o comportamento dos elementos flexionáveis nesse mesmo SN. Note-se que, o facto de um elemento ser o primeiro item flexionável do SN não implica, necessariamente, que esteja a ocupar a primeira posição linear da cadeia sintagmática, conforme se pode constatar no exemplo usado para o factor “P” da Tabela 12.9. Assim, e dado que na cadeia sintagmática se inserem itens passíveis e não-passíveis de flexão, poder-se-á observar se a inibição que ocorre em elementos posteriores é consequência da inserção de itens não-flexionáveis entre os elementos flexionáveis da cadeia sintagmática, não se observando o Princípio da Coesão Estrutural (Haiman, 1983:782), em virtude de se cortar o encadeamento da concordância. Relembre-se que Guy (1981b:179) advogou que o efeito da *posição linear* e *classe gramatical* são idênticos, pois parece existir uma equivalência entre ambos os grupos de factores, visto os determinantes ocorrerem preferencialmente na primeira posição, os nomes se situarem usualmente na segunda posição e os adjectivos ocorrerem, por norma, na terceira posição. Discordando do postulado de Guy (1981b:179), Scherre (1988:167) não só opinou que a marcação PL ocorre preferencialmente na primeira localização, independentemente da categoria gramatical que aí se posicione, mas também que a pluralização do SN é determinada pelo processamento paralelo (Scherre, 1988:208), isto

é, pelo princípio de que marcas conduzem a mais marcas e zeros implicam a ocorrência de mais zeros (Scherre, 1988:235).

Assim sendo, a variável independente *ordem do item na cadeia dos constituintes flexionáveis*, que não temos conhecimento de ter sido observada em outros trabalhos sobre concordância variável, analisada *per se* ou em relação com outros grupos de factores, como as variáveis linguísticas *posição linear* ou *classe gramatical* e a variável social *idade*, poderá ajudar na clarificação, entre outras, de questões como as que se seguem: (i) Marcas conduzem a marcas no item imediatamente posterior? (ii) Os itens não-passíveis de marcação afectam a cadeia de concordância quando inseridos entre elementos passíveis de marcação? (iii) Qual a posição preferencial dos diferentes itens gramaticais flexionáveis na cadeia do SN?

Para elaboração do grupo de factores *ordem do item na cadeia dos constituintes flexionáveis* constituíram-se os factores constantes na tabela seguinte:

**Tabela 12.9.** Factores constituídos para a variável independente *ordem do item na cadeia dos constituintes flexionáveis em número*.

<b>11) Variável independente <i>ordem do item na cadeia dos constituintes flexionáveis em número</i></b>				
<b>Cód.</b>	<b>Posição</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
<b>P</b>	Primeiro elemento	duzentos e tal <b>CONTO?</b> [SALVH3]	886/1.487	60
<b>S</b>	Segundo elemento	todos OS <b>dias</b> [ANTOM1]	286/754	38
<b>T</b>	Terceiro elemento	todos os <b>DIAS</b> [ANTOM1]	28/85	33
<b>Q</b>	Quarto ou outro elemento	uma outras minhas <b>PREOCUPAÇÕES</b> [CASTEH1]	(1/10)	10
<b>Totais:</b>			1.201/2.336	51

Esta variável independente permitirá avaliar a seguinte hipótese:

**Hipótese 9 – *Ordem do item na cadeia dos constituintes flexionáveis em número*:** a inserção de itens não flexionáveis em número entre os elementos flexionáveis que vêm manifestando concordância confirma a aplicação do Princípio do Processamento Paralelo (Scherre, 1988:208), já que actua como elemento condicionador da marcação nos itens posteriores, cortando a cadeia de concordância e confirmando o pressuposto de que ausência de marcas leva à não inserção de mais marcas.

### 3.6.9. Variável independente grau de concordância de número no SN

A constituição desta variável independente visou permitir que se observasse, numa perspectiva essencialmente geracional, quais os tipos de SN's que são produzidos pelas diferentes faixas etárias de Almojarife. De facto, atendendo ao número de elementos flexionáveis inseridos nas cadeias sintagmáticas, é possível não só tirar conclusões acerca da extensão e grau de complexidade das mesmas mas também visionar quais são as gerações que as produzem. No entanto, devemos referir que este grupo de factores se adequa mais a um estudo não-atomístico, conforme realizado por P. Andrade (2003:106-107), pelo que nos limitaremos a fornecer o número de ocorrências e fazer uma breve análise sobre as mesmas.

Para a elaboração do grupo de factores grau de concordância de número no SN constituíram-se os seguintes factores:

**Tabela 12.10.** Factores constituídos para a variável independente grau de concordância de número no SN.

12) Variável independente grau de concordância de número no SN				
Cód.	Factor	Exemplo	Nr. de Ocorrências	%
1	SN's com 1 elemento flexionável no SN	duas <b>PISSOAS</b> [CASTEH1]	204/729	28
2	SN's com 2 elementos flexionáveis no SN	<b>TRABALHOS PRIVADO</b> [CASTEH1]	865/1.352	64
3	SN's com 3 elementos flexionáveis no SN	UMAS nove ou onze <b>JOVEM FORMADO</b> [CASTEH1]	125/231	54
4	SN's com mais de 3 elementos flexionáveis no SN	UMA OTRAS MINHAS <b>PREOCUPAÇÕES</b> [CASTEH1]	(8/28)	29
<b>Totais:</b>			1.202/2.340	51

Partindo do pressuposto que as diferentes faixas etárias de Almojarife exibem gramáticas distintas, a variável independente grau de concordância de número no SN permitirá avaliar a seguinte hipótese:

**Hipótese 10 – Grau de concordância de número no SN:** a faixa etária mais idosa de Almojarife produz SN's de estrutura mais reduzida e mais baixo grau de complexificação estrutural. Assim sendo, os SN's vão-se tornando mais extensos e

complexos com as novas gerações, confirmando-se uma aquisição geracional da regra de concordância.

### **3.6.10. Variável independente traço semântico do SN (*Animacidade*)**

No estudo dos seus dados, Scherre (1988:266) observou que a incidência de marcação PL sobre o substantivo “pessoa” (82%) era notoriamente superior à que recaía sobre o nome “coisa” (36%). Diversos trabalhos linguísticos (Ferraz, 1979; Lucchesi, 2000a; Alexandre & Hagemer, 2007; Figueiredo, 2009a) têm atribuído ao traço [+humano], por ser discursivamente mais saliente (Mollica, 1977:37), um papel relevante em vários tipos de fenómenos que ocorrem em variedades de português (p.e. a conservação da flexão de género em nomes adjectivados do tipo *macho/fêmea* ou *homem/mulher*). Assim sendo, Scherre (1988) levantou a hipótese de os itens nominais que possuem este tipo de traço apresentarem maiores índices de concordância.

Note-se que, anteriormente aos achados de Scherre (1988), já Mühlhäusler (1981), ao analisar dados do *pidgin* tok pisin, havia constatado que as propriedades dos nomes com traço semântico [+animado] poderiam ter influência na marcação de número. Ao que tudo indica, estas propriedades semânticas interagem com a sua função sintáctica, levando a que os nomes em posição de sujeito revelem maior tendência a serem marcados, uma vez que nessa posição surgem, maioritariamente, agentes do tipo animado (Young, 1991:111). Posteriormente, Sabino (1983) verificou também a relação entre nomes de traço semântico [+animado] e a função sintáctica de sujeito no crioulo de base inglesa das Ilhas Virgens, concluindo que este traço favorece a marcação de número, enquanto o traço [-animado] a inibe.

Godinho (2005) procurou também verificar se os nomes de traço [+animado] favoreciam a marcação PL nos SN's produzidos por aprendentes chineses de PtgL2 em contexto formal de ensino, mas a ferramenta VARBRUL não considerou esta variável significativa para a observação da CPL-var dos mesmos.

Por seu lado, Jon-And (2009:7) constituiu cinco factores para analisar a influência da variável independente *animacidade* na marcação PL do PCV, uma variedade que tem como substrato quer o crioulo de São Vicente (as gerações mais novas e parte dos falantes mais idosos) quer o crioulo de Santo Antão (os restantes falantes mais velhos).

Assim, além dos nomes de traço semântico [+humano], [+animado] [-humano] e [-animado], a linguista criou um factor para substantivos referentes a partes do corpo e outro para plantas. Os resultados apresentados por Jon-And (2009) permitem concluir que o factor [+humano] favorece a marcação, enquanto o factor [-animado] a inibe. Quanto aos outros três factores, exibem um reduzido número de ocorrências, pelo que os seus pesos não foram considerados pelo programa VARBRUL.

Pesquisas sobre concordância dentro do SN de línguas parcialmente reestruturadas têm também trazido à evidência que existe um paralelo entre os padrões de variação das marcações mórnicas de PL e de género. Em ambas as situações, os constituintes localizados antes do núcleo do SN revelam maior tendência a serem marcados do que os posicionados pós-nuclearmente. Apesar de os fenómenos se assemelharem, têm sido justificados distintivamente. Assim, no caso da CPL-var, tem vindo a ganhar força a possibilidade de as L1's africanas exercerem influência, conjuntamente com as dos crioulos, na configuração do SN do PtgL2 ou dos CP's. Posteriormente, caso estes dois tipos de línguas entrem em contacto, a partilha de traços comuns a ambas pode determinar a incorporação dos mesmos na nova língua que emerge (Clements, 2009:19). Assim, no caso da CPL-var registada no PA, a mesma pode estar sujeita a condicionantes que ancoram na aquisição do PtgL2 e PtgL1 em situação de contacto que envolveu estruturas de PtgL2 e CP (o santomense) influenciadas pelas línguas ancestrais africanas (cf. Guy, 1981a; P. Andrade, 2003; Baxter, 2004, 2009; Figueiredo, 2008, 2009a, 2009b; Jon-And, 2008). Idêntica situação ocorre com a CGEN-var, cujas condicionantes ancoram também nas línguas africanas, tendo estas determinado que o núcleo do SN seja o item responsável pela fixação da morfologia flexional de género e irradiação da concordância de género (Lucchesi, 2000a:189; Figueiredo, 2009a:50), indo esta marcar mais fortemente os elementos adjacentes ao núcleo, especialmente os situados imediatamente antes deste. Como tal, a inibição à marcação do género cresce à medida que os itens se encontram posicionados mais longe do núcleo na cadeia sintagmática ou à direita do mesmo (ponto 2.6.4.2 do presente trabalho).

A premissa que determina o núcleo do SN como pólo irradiador da marcação mórnic de género estabelece ainda aproximações com algumas realizações que ocorrem nos CP's da Alta Guiné e do Golfo da Guiné (Lucchesi, 2000a:189), dado que, nestes, é

também a partir do núcleo do SN que se recompõe a morfologia do género. Nesta conformidade, é visível um paralelo no plano da morfossintaxe de género entre estes CP's, o PA e o dialecto de HEL-Ba, já que, em todos eles, se constata a existência de ocorrências que revelam não-aquisição de matéria gramatical da LA, com a erosão da morfologia flexional do nome a afectar mais o mecanismo morfossintáctico da concordância do que a flexão mórfica do género dos núcleos nominais (cf. Chomsky, 1996 [1995], 2001a; Carstens, 2000) que dizem respeito a seres animados (Lucchesi, 2000a:337-338). Deste modo, com o apagamento do mecanismo morfossintáctico de função meramente gramatical para indicação do género, conservou-se a flexão deste nos nomes adjectivados do tipo *macho/fêmea* ou *homem/mulher*. No caso concreto do santomense (exemplos [53a] e [53b]) e do CCV (exemplo [54]), o valor referencial de sexo ocorre, por norma, com os referidos substantivos adjectivados sendo colocados pós-nuclearmente ao nome, que se mantém invariável na forma masculina.

Dado que este tipo de ocorrências é também recorrente no PA (exemplos [55a] e [55b]), e como em termos generativistas a causa do acontecimento pode ser atribuída ao apagamento dos traços não interpretáveis dos elementos funcionais ou não-parametrizados, adquiridos em fase pós-puberdade (Hawkins, 1993, 1998, 2001; Yip, Rutherford & Clashen, 1995; Hawkins & Chan, 1997; Franceschina, 2002, 2003), decidimos verificar se o fenómeno da animacidade afecta também o mecanismo morfossintáctico da concordância PL.

Todavia, e independentemente da questão do paralelismo entre os padrões de variação das marcações mórficas PL e género e da possibilidade de incidência de marcação PL sobre o substantivo “pessoa” ser superior à que recai sobre o nome “coisa” (Scherre, 1988:266), o estudo da variável independente *traço semântico do SN* levantou a hipótese de o desenho da CPL-var no SN do PA ser mais condicionado por questões do tipo estrutural do que de carácter semântico, como se verá na secção dedicada à análise dos resultados desta variável (ponto 5.1.1). A fim de se ajudar a trazer mais clarificações a todas estas questões, codificaram-se os substantivos em termos da variável da animacidade do seu traço semântico, constituindo-se o grupo de factores *traço semântico do SN (animacidade)*, ao qual se incorporaram os três factores constantes na Tabela 12.11. Ainda a propósito da codificação destes itens, refira-se que elementos



gramaticais como o colectivo “equipas” foram considerados como portadores do traço semântico [+humano]. Quanto aos resultados finais, eles poderão também ajudar-nos a confirmar o pressuposto de o fenómeno da CPL-var, ao contrário da CGEN-var, não ter como pólo irradiador o núcleo do SN, mas sim a categoria funcional DET, na qual o PL é inserido através de um morfema do tipo *singleton*, o qual serve de âncora para o controlo da pluralização (Baxter, 2009:292).

Para a variável traço semântico do SN, colocamos a seguinte hipótese:

**Hipótese 11 – Traço semântico do SN:** Os nomes com traço semântico [+animado] favorecem a marcação, por oposição aos substantivos com traço semântico [–animado], que revelam tendência para desfavorecê-la.

**Tabela 12.11.** Factores constituídos para a variável independente *traço semântico do SN*.

13) Variável independente <i>traço semântico do SN</i>				
Cód.	Traço	Exemplo	Nr. de Ocorrências	%
<b>h</b>	[+humano]	os <b>HOMENS</b> [CASTEH1]	747	60
<b>a</b>	[–humano] [+animado]	esses <b>GAROUPA</b> assim [ZECAH1]	75	37
<b>z</b>	[–animado]	<b>TRABALHOS</b> privado [CASTEH1]	1.501	48
<b>Totais:</b>			1.201/2.338	51

### 3.6.11. Variável independente *sexo*

Em trabalhos de sociolinguística quantitativa sobre variação, as variáveis independentes do tipo social, apesar de incidirem sobre dados sincrónicos, surgem também intimamente ligadas à noção de prestígio, a qual, por sua vez, pode determinar comportamentos linguísticos conotados a mudanças em progresso. A discussão em torno do modo como a variável *sexo* influencia as mudanças em curso não tem sido consensual. Os estudos pioneiros com recurso a esta variável social (Chambers & Trudgill, 1980:91-93) apresentaram as mulheres como impulsionadoras das mudanças em direcção às formas de prestígio, enquanto os homens seriam os responsáveis por fenómenos de não-realização, já que apresentavam maior índice de ocorrências que se afastam dos registos padronizados. Contudo, Labov (1978:82) discordaria deste pressuposto, uma vez que atribuiu às mulheres a responsabilidade por todos os processos

de mudança. Quanto a Scherre (1988:429), relembra que o papel da variável *sexo* na questão da mudança em curso não é muito claro na literatura em linguística. Ainda assim, os falantes urbanos do MRJ, que produziram as ocorrências que Scherre (1988) observou, pareciam confirmar o pressuposto levantado por Chambers & Trudgill (1980), dado que são as mulheres quem favorece a marcação de número, enquanto os homens a inibem. A possível tendência para o favorecimento da marcação em número revelada pelo estrato masculino de comunidades urbanas seria ainda constatada por Lopes (2001:109), quando se debruçou sobre a fala culta de São Salvador.

Em contrapartida, Baxter, Lucchesi & Guimarães (1997), ao estudarem a CGEN-var das produções dos falantes rurais de HEL-Ba, apontaram uma tendência contrária para este fenómeno, com os homens a propiciarem mais a marcação do que as mulheres. Por seu lado, observando as tendências de mudança nas normas cultas e popular do PB, Lucchesi (1998a) chama a atenção para a falácia que generalizações como a de Chambers & Trudgill (1980) podem encerrar, quando afirmam que as mulheres tendem a dirigir as mudanças em direcção às formas de prestígio. Confirmando a tendência de mudança registada por Baxter, Lucchesi & Guimarães (1997) para HEL-Ba, Lucchesi (2000a:281) reafirma que não se pode generalizar qualquer tendência com vista a atribuir uma maior ou menor influência na marcação a um dos sexos, porque *“não apenas não é muito claro, como também constitui um grande equívoco considerar essas variáveis sociais como parâmetros isolados, sem se considerar um contexto mais abrangente, no qual esses elementos podem ser integrados numa interpretação totalizadora, capaz de dar conta da especificidade de cada processo histórico particular de mudança lingüística.”*. Nesta perspectiva, o estudo compartimentado das variáveis independentes, sejam elas linguísticas sejam elas sociais, pouco contributo dá para a observação dos processos de mudança registados numa determinada comunidade, enquanto todo caracterizado por determinado padrão linguístico. Portanto, só é possível concluir que certos fenómenos se encontram em situação de variação ou processo de mudança caso, a partir das evidências empíricas concedidas por cada uma das variáveis, o pesquisador consiga perceber e justificar, com base quer em fundamentações teórico-científicas já trazidas à evidência quer em inferências coerentemente argumentadas, a forma como tais evidências se articulam entre si na construção do todo coerente que é o

registro linguístico do grupo em questão. Significa isto dizer que a observação da variação e mudança da língua só é possível caso se tenha em conta que a língua é algo heterogéneo, determinado tanto por condicionantes linguísticas inerentes à sua própria estrutura como por fenómenos sociais externos, tal como o preconizam os preceitos da sociolinguística variacionista. E, neste aspecto, também os resultados de P. Andrade (2003:129-130) acabariam por evidenciar a fragilidade das generalizações avançadas por Chambers & Trudgill (1980), já que a autora confirmaria, em definitivo, que, na comunidade de HEL-BA, não são as mulheres, mas sim os homens a liderar a mudança em direcção à norma padrão, tendo, assim, a responsabilidade maior na fixação das regras de concordância. Resta ainda referir que, trabalhando igualmente com dados de informantes rurais, Baxter (2009:285) verificaria igual tendência para o PT, apresentando valores em tudo semelhantes aos de P. Andrade (2003:130).

No caso do PA, a Tabela 12.12 permite observar que as ocorrências passíveis de análise produzidas por homens (1.323) e mulheres (1.017) apresentam valores bastante aproximados. No nosso estudo, os falantes encontram-se equitativamente distribuídos por escalões etários, não se registando também grandes variações de escolarização a nível dos mesmos estratos de idade. No entanto, os falantes do sexo masculino terão tido um contacto mais estreito com o PE em determinada fase de vida da comunidade, já que eram os homens que migravam para os centros urbanos em busca de oportunidades de emprego, por um lado, e eram também integrados nos quadros do exército colonial, por outro lado.

**Tabela 12.12.** Factores constituídos para a variável independente *sexo*.

<b>14) Variável independente <i>sexo</i></b>				
<b>Cód.</b>	<b>Sexo</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
<b>H</b>	Masculino	faço <b>TRABALHOS</b> PRIVADO [CASTEH1]	709/1.323	54
<b>M</b>	Feminino	vinho agora stá <b>DEZ CONTO</b> [MMDEUSM3]	493/1.017	49
<b>Totais:</b>			1.202/2.340	51

Tendo em conta que os aspectos mencionados, formulámos a seguinte hipótese para a variável *sexo*:

**Hipótese 12** – **Sexo**: Ambos os sexos devem apresentar um nível semelhante de favorecimento da pluralização no SN. Contudo, não é de excluir a possibilidade de os falantes do sexo masculino, por condicionantes sociolinguísticas, apresentarem ligeira tendência para realizarem mais a concordância em número do que os falantes do sexo feminino.

### **3.6.12. Variável independente *idade***

A análise da variável *idade* permite detectar possíveis mudanças linguísticas em curso na comunidade, verificáveis através do chamado *tempo aparente* (Labov, 1972a). É o que sucede, por exemplo, quando esta variável apresenta uma distribuição curvilínea, a qual revela que se está perante uma variação estável, isto é, um estágio de evidente mudança consumada. Em contrapartida, se a variável *idade* apresentar uma distribuição inclinada, com o ápice da aquisição da regra de concordância na faixa etária mais nova e o maior índice de variação na faixa etária mais velha, significa que a comunidade ainda se encontra em fase de mudança em progresso (Chambers & Trudgil, 1980; Baxter, Lucchesi & Guimarães, 1997). Contudo, importa referir que a variável *idade*, por si só, não é suficiente para indicar a mudança em progresso, pelo que é necessário recorrer a outros padrões sociais associados a ela, como o sexo dos falantes, os seus anos de escolarização e até o estrato social em que estão inseridos.

De acordo com a metodologia que tem sido aplicada em outros trabalhos sobre análise da concordância variável manifestada por diferenciados grupos linguísticos (Lucchesi, 2000a; P. Andrade, 2003; Baxter, 2004; 2009), distribuímos equitativamente os informantes do PA por faixas etárias distintas: (i) dos 20 aos 40 anos (faixa etária FE-1); (ii) dos 41 aos 60 anos (faixa etária FE-2); (iii) mais de 60 anos (faixa etária FE-3). Ao considerarmos estas diferentes gerações de informantes, é-nos possível estudar a covariação sistemática entre variáveis linguísticas e sociais, verificando em qual delas ocorre a maior frequência de formas inovadoras. Paralelamente, ao contemplarmos um período de pelo menos 20 anos para cada uma das faixas etárias, garante-se, em tempo real, a possibilidade de uma projecção para um período de tempo aparente que é, segundo os especialistas em estudos variacionistas,<sup>156</sup> suficiente para determinar a cristalização da variação das três gerações e confirmar a mudança no sentido da fixação

da regra de concordância de número. Esta, por seu lado, vai contribuir para o desenho da gramática da comunidade de Almojarife.

Dado que alguns trabalhos sobre CPL-var apresentam os informantes estratificados por escalões etários distintos dos nossos (Scherre, 1988; Lopes, 2001), remetemos para o capítulo 5 (análise dos dados das variáveis extralinguísticas) os considerandos apresentados em pesquisas anteriores acerca da variável independente *idade*, nomeadamente quando compararmos os nossos resultados com os de Scherre (1988), Lopes (2001), P. Andrade (2003) e Baxter (2004, 2009). Os factores que constituímos para a variável independente *idade* estão patentes na tabela seguinte:

**Tabela 12.13.** Factores constituídos para a variável independente *idade*.

<b>15) Variável independente <i>idade</i></b>				
<b>Cód.</b>	<b>Escalão etário</b>	<b>Exemplo</b>	<b>Nr. De Ocorrências</b>	<b>%</b>
1	FE-1 (de 20 a 40 anos)	faço TRABALHOS PRIVADO [CASTEH1]	623/984	63
2	FE-2 (de 41 a 60 anos)	são ESSES TONGA de lá. [ABILH2]	373/786	34
3	FE-3 (+60 anos)	vinho agora stá DEZ CONTO. [MMDEUSM3]	206/570	24
<b>Totais:</b>			1.202/2.340	51

Para a variável independente *idade* estabelecemos a seguinte hipótese:

**Hipótese 13 – *Idade*:** A idade e a aquisição da regra de concordância estabelecem uma relação directamente proporcional, com as gerações mais novas a favorecerem cada vez mais a marcação.

### 3.6.13. Variável independente *estadia* (*Permanência fora da comunidade*)

A integração da variável *estadia* no nosso estudo tem a ver com o facto de muitas mudanças ocorrerem em virtude da “elasticidade” da própria língua, que aceita incorporações de outros dialectos. Estas poderão ser trazidas por falantes de dialectos distintos que visitem regularmente uma determinada comunidade, ou passem a viver integrados nela. Em complementação, as incorporações poderão ser também trazidas por membros da própria comunidade que se desloquem para fora desta por prolongados períodos de tempo e regressem com formas novas integradas no seu idiolecto, as quais são depois transmitidas aos outros elementos da sua comunidade.

O contacto com variedades urbanas, normalmente detentoras de um padrão de concordância mais rico, contribui para aumentar a inserção de marcas. A exemplo do que vem sendo adoptado em outros trabalhos sobre concordância variável (Lucchesi, 2000a; P. Andrade, 2003), dividimos os informantes do PA em duas categorias: (i) falantes que se ausentaram da comunidade durante pelo menos um período de seis meses; (ii) falantes que nunca se ausentaram da comunidade ou, caso o tenham feito, não tenham permanecido fora desta por períodos superiores a seis meses. As ausências dos membros de Almojarife são normalmente determinadas por questões laborais, já que alguns dos falantes emprestaram, sazonalmente, mão-de-obra às roças ou ter-se-ão deslocado a povoações mais expressivas ou mesmo a São Tomé, aí exercendo actividades distintas (p.e. construção civil, alfaiataria, etc.) ou servindo no exército colonial. Casos há em que os falantes estiveram em países vizinhos da região do Golfo da Guiné, como o Gabão, aonde se deslocaram para mercadejar, tendo aí prolongado a sua estadia para exercer diversas actividades. Uma das informantes regista também viagens e estadias longas em Angola, país ao qual se deslocava para visitar um filho. Como se verifica, os casos de ausência prolongada não implicaram um contacto mais próximo com a LA (cf. Tabela 8.1), pelo que as mesmas não terão sido um factor favorecedor da aquisição da regra de concordância.

A Tabela 12.14 mostra-nos os factores estabelecidos para a variável independente *estadia*:

**Tabela 12.14.** Factores constituídos para a variável independente *estadia*.

<b>16) Variável independente <i>estadia</i></b>				
<b>Cód.</b>	<b>Ausência</b>	<b>Exemplo</b>	<b>Nr. de Ocorrências</b>	<b>%</b>
<b>S</b>	Sim	tem DOIS GÉMIA de três ano [LUISH2]	358/719	50
<b>N</b>	Não	faço TRABALHOS PRIVADO [CASTEH1]	844/1.621	52
<b>Totais:</b>			1.202/2.340	51

Com a constituição da variável independente *estadia*, procuramos confirmar a hipótese seguinte:

**Hipótese 14 – *Estadia*:** Na comunidade de Almojarife, o factor ausência não contribuiu para que falantes produzam padrões da concordância PL mais próximos dos da LA.

### 3.6.14. Variável independente *escolaridade*

A propósito da variável *idade* (ponto 3.6.12), referimos que as variáveis extralinguísticas, observadas por si só, não são suficientes para indicar a mudança em progresso, sendo necessário recorrer a outros padrões sociais associados a ela, como os anos de escolarização dos falantes ou o estrato social em que estes estão inseridos. Socialmente, a classe alta, sendo por norma escolarizada, utiliza mais as formas padronizadas, que são também sinónimo de prestígio. Por seu lado, as classes média-baixa, normalmente analfabetas ou com baixo índice de escolaridade, apresentam um registo marcado por traços linguísticos populares ou vernáculos. Assim sendo, ocorre uma tensão entre os registos das duas categorias de classes, com a primeira reflectindo uma variação estável e a segunda liderando os processos de mudança (Labov, 1982:77-78).

No ponto 3.6.11, a propósito da discussão para levantamento da hipótese a observar através da variável independente *sexo*, tecemos comentários sobre a forma como o estudo compartimentado das variáveis independentes pouco contributo dá para a observação dos processos de mudança, o que pode também ser confirmado pela variável social *escolaridade*. A constituição deste grupo de factores tem como condição primeira observar se os falantes menos escolarizados usam com menor incidência as formas-padrão do que aqueles que possuem um grau de escolarização mais elevado. De facto, é tido por aceite que a difusão dos modelos normativos concorre para o processo de mudança linguística, uma vez que não só expõe mais os falantes à norma culta da língua falada mas também os submete ao contacto com a língua padrão escrita. No entanto, convém ressaltar aqui que tanto a exposição como o contacto referido não se dão só em ambiente escolar. Com efeito, contingências como o tipo de actividade profissional exercida, a incorporação ou não de elementos do sexo masculino nos quadros do exército colonial ou a deslocação do falante para fora da comunidade podem, também, determinar que o informante, apesar de se enquadrar em determinado factor do grupo de factores *escolaridade*, tenha partilhado experiências sociolinguísticas que determinaram a expansão do seu repertório linguístico. E se tal acontecer, o informante possuirá um idiolecto mais próximo das formas padrão da LA, contrariando, assim, a tendência generalizada do factor em que se encontra incorporado. Como tal, para

contemplar a análise dos aspectos expostos, o grupo de factores *escolaridade* tem que ser observado em concomitância com todas as outras variáveis sociais, isto é, à luz da própria sociohistória do país, em geral, e da comunidade em observação, em particular.

Durante o levantamento dos dados do nosso *corpus* não deixou também de se ter em conta a questão da “representatividade”, isto é, da quantidade de informantes a incluir em cada factor social. Assim, quanto mais informantes se possuir para cada factor (e consequentemente para cada célula VARBRUL), mais se diluem as idiosincrasias, reduzindo-se, logicamente, a possibilidade de adulteração dos resultados finais. Entre os linguistas, a questão da “representatividade” tem sido alvo de algum debate no que concerne ao número mínimo de informantes tido como representativo para indicar a direcção das tendências linguísticas de uma determinada comunidade. O consenso tem apontado para um mínimo de quatro informantes a incluir em cada célula, embora alguns pesquisadores defendam que esse mínimo deverá ser de cinco informantes. A este propósito teremos que referir que a célula *escolarização alta (frequência do ensino pós-primário)*, da variável social *escolaridade*, apenas inclui três informantes, por razões que se prendem com a própria sociohistória da comunidade de Almocharife, sendo os seus membros na quase totalidade indivíduos de estrato social humilde, logo com manifesta dificuldade em acederem ao ensino. Contudo, todos os outros grupos de factores incluídos nas diferentes variáveis sociais possuem mais de cinco informantes, estando, assim, salvaguardada a possibilidade de os nossos resultados saírem viciados por questões idiosincráticas conectadas aos falantes.

Diga-se, por fim, que as variáveis sociais, apesar de serem independentes, partilham semelhanças com as variáveis dependentes, uma vez que são também entidades nominais, revelando-se descontínuas ao não aceitarem valores intermediários. Contudo, a propósito da descontinuidade de algumas destas variáveis, é necessário um esclarecimento. Se é um facto que a variável *sexo*, por exemplo, implica apenas duas alternativas possíveis (masculino *versus* feminino), já outras variáveis, como a *idade* e a *escolaridade*, pressupõem um *continuum* que, à primeira vista, as desenquadrará do modelo nominal. De facto, tanto uma como a outra implicam uma escala de continuidade no domínio social, representada por traços distintivos (jovem/meia idade/terceira idade...; não escolarizado/menos escolarizado/mais escolarizado...), que



poderão ser seleccionados em função de critérios operacionais, com vista a enquadrá-las em factores que abranjam um, e apenas um, grupo social específico. Assim, um determinado falante que se enquadre na faixa etária FE1 (20-40 anos), não pode ser enquadrado também, por exemplo, na FE3 (+60 anos).<sup>157</sup> Da mesma forma, um falante que tenha frequentado o ciclo preparatório não pode, simultaneamente, ser analfabeto.

A inclusão da variável *escolaridade* na nossa análise ficou a dever-se ao facto de a mesma permitir observar até que ponto a difusão de modelos normativos em São Tomé terão contribuído para a aquisição da regra da concordância de PL evidenciada pelo PA. Como se sabe, a frequência do ensino leva os falantes a prestarem mais atenção às regras de concordância, pelo que passam a aplicar mais as marcas de PL. Nesta conformidade, a constituição dos factores para a variável *escolaridade* surge condicionada não só pelos contextos sociais específicos de cada comunidade mas também pela forma como o ensino se encontra estruturado no país a que a comunidade pertence. E, neste aspecto, São Tomé e Príncipe constitui caso peculiar, como se verá aquando da análise dos dados respeitantes a esta variável social (capítulo 5).

Em trabalhos sobre CPL-var no SN, apenas Scherre (1988:445), Lopes (2001:107) e Jon-And (2009:4) apresentam resultados sobre a variável escolaridade. Contudo, como as duas primeiras se debruçaram sobre variedades urbanas do PB, a tendência será para os falantes destas serem mais escolarizados que os de Almojarife, uma comunidade rural. Paralelamente, o sistema organizativo do ensino no Brasil é distinto do de São Tomé e Príncipe ou Cabo Verde, tanto na época em que estes arquipélagos eram colónias portuguesas como na actualidade. Como tal, os critérios que subjazem à constituição dos nossos factores para a variável independente *escolaridade* divergem daqueles que foram levados em conta para observação da CPL-var das variedades brasileiras de português, mas assemelham-se aos elaborados por Jon-And (2009:4) para estudo do PCV. Não obstante, estes aspectos não negam a viabilidade operacional desta variável, dado que qualquer dos ensinos em questão contribuiu para uma maior exposição dos falantes à norma padrão do português. Considerações sobre os estudos de Scherre (1988), Lopes (2001) e Jon-And (2009), bem como comparações dos resultados dos mesmos com os nossos serão apresentados no capítulo 5, dedicado à análise dos

resultados das variáveis sociais. A tabela 12.15 apresenta os factores que constituímos para a variável escolaridade, para a qual estabelecemos a hipótese seguinte:

**Hipótese 15 – Escolaridade:** Quanto mais alto é o nível de escolaridade, mais atenção é prestada pelos falantes ao mecanismo da concordância. Como tal, os falantes mais escolarizados flexionam mais os itens pluralizáveis do SN.

**Tabela 12.15.** Factores constituídos para a variável independente *escolaridade*.

17) Variável independente <i>escolaridade</i>				
Cód.	Escolaridade	Exemplo	Nr. de Ocorrências	%
a	Analfabetos	tem DOIS <b>GÉMIA</b> de três ano [LUISH2]	187/596	31
s	Escolarização baixa (Frequência do ensino primário)	leva pa OUTROS <b>LUGAR</b> também [CARMOM1]	420/695	60
e	Escolarização média (Ensino primário concluído)	pissoa mandô AS <b>MOBÍLIA</b> [OSVALH1]	347/697	50
p	Escolarização alta (Frequência do ensino pós-primário)	faço <b>TRABALHOS PRIVADO</b> [CASTEH1]	248/352	71
<b>Totais:</b>			1.202/2.340	51

### 3.7. Codificação dos itens de acordo com os códigos atribuídos aos factores

A quantificação dos dados pela ferramenta VARBRUL só é possível caso os mesmos tenham sido codificados. A codificação, feita de acordo com os factores estabelecidos, visa confirmar, ou não, as hipóteses levantadas para estes. Não nos delongaremos sobre a forma de proceder às codificações, já que este aspecto se encontra bastante contemplado na leitura da especialidade (p.e. Young & Bayley, 1996; Guy, 1988; Pintzuk, 1988; Paolillo, 2002; Scherre & Naro, 2003; Guy & Zilles, 2007). No nosso caso, será suficiente referir que os dados foram codificados de acordo com as categorias e subcategorias seleccionadas como variáveis-alvo, constituindo cadeias de códigos como as que se apresentam seguidamente (SN constituído por dois elementos):

**Tabela 13.1.** Exemplo de cadeia de códigos resultante da codificação dos elementos do SN.

	<b>Cadeia de códigos:</b>	<b>Exemplo:</b>
1.	(0DPP'AU2cS2hH1Np	<b>jovens</b> HABILIDOSO [CASTEH1]
2.	(SYP/VC01/P2hH1Np	<b>JOVENS</b> habilidoso [CASTEH1]

- Obs.:** (i) O item analisado surge grafado em caixa alta (maiúsculas), estando o núcleo do SN marcado a negrito;  
(ii) O primeiro elemento da cadeia de códigos, depois do parêntesis, diz respeito a um dos factores do grupo de factores *variável dependente*, enquanto cada um dos restantes quinze itens se refere a um factor de uma das quinze variáveis independentes constituídas.

Note-se que o processo de codificação dos dados não está concluído quando se finaliza a atribuição dos códigos a todos os itens passíveis de marcação PL, já que, como veremos seguidamente, após processamento dos dados com recurso ao suporte computacional, alguns dos factores constituídos podem apresentar-se como irrelevantes para a observação da variação. É o que sucede, por exemplo, com os factores identificados pela ferramenta VARBRUL como *knockouts* (prefiguram 0% ou 100% de marcações) ou com os grupos de factores apontados como *singletons* (variáveis independentes que apresentam um único factor), que terão que ser retirados da análise.

Refira-se ainda que a recodificação dos dados pode também ser necessária, caso determinados factores apresentem pesos relativos próximos ou certas variáveis independentes actuem conjuntamente sobre a regra. Nestes casos, a situação resolve-se com a amalgamação de grupos de factores para se constituírem grupos de factores mais robustos ou, em alternativa, com a alteração do arquivo de condições, que fornece ao VARBRUL as informações sobre o modo como ele deve conduzir a análise dos dados, isto é, quais os grupos de factores que devem ser processados em cada uma das rodadas (Paolillo, 2002:58). Para tanto, combinam-se os grupos de factores com recurso à linguagem LISP (*List Programming*) (Young & Bayley, 1996:275), especialmente programada para uso informático, e submetem-se os dados a nova análise VARBRUL.

### **3.8. Quantificação dos dados para selecção das variáveis independentes e dos factores a serem incorporados na análise final**

Depois de se constituírem as variáveis independentes e seus factores, e se codificarem também, em função destes, todos os itens em que se supõe ocorrer marcação de número PL, introduziram-se os dados no suporte computacional e procederam-se às primeiras

rodadas VARBRUL. Estas não apresentaram resultados sólidos, revelando, desde logo, haver interferências quer entre as variáveis independentes quer entre factores dos grupos de factores, como aliás se esperaria, face ao elevado número de grupos de factores e factores constituídos.<sup>158</sup> Como tal, situações houve em que alguns grupos de factores foram simultaneamente seleccionados e rejeitados, enquanto outros não apresentaram qualquer tipo de resultados. Nestes casos, a metodologia propõe que se proceda a uma análise cuidada de todas as variáveis independentes e seus factores, quer aplicando testes de significância estatística dentro dos grupos de factores quer observando a “convergência” de cada rodada, a fim de fazer valer, uma vez mais, o princípio da Lei da Parcimónia e se eliminarem os factores ou mesmo grupos de factores que não apresentem valores estatisticamente significativos. A eliminação de determinados factores poderá então passar pela retirada dos mesmos dos dados a analisar ou, em alternativa, pela sua junção a outros factores, criando-se um novo factor amalgamado.

Tido como um método bastante eficaz para testar a significância de análises alternativas dentro dos grupos de factores, o cálculo de qui-quadrado permite observar se a amalgamação de factores dentro de um mesmo grupo de factores compõe um novo factor que possa ser incorporado à análise. Reduzindo-se os factores, um a um, vão-se efectuando rodadas VARBRUL, testando-se os sucessivos efeitos das alterações no *log-likelihood*, até se obter a melhor análise, isto é, aquela que recorre ao menor número de factores para explicar os dados. No entanto, é também possível amalgamar vários factores num único, conquanto tal se justifique estatística ou teoricamente. Nos capítulos 4 e 5, ao detalharmos os passos dados no sentido de tratar cada um dos grupos de factores constituídos, referiremos também se estes foram objecto de qualquer dos tratamentos anteriores, e porquê.

Quando a análise efectuada engloba mais de dez grupos de factores – como acontece no nosso caso –, a ferramenta VARBRUL gera arquivos com muitas células, o que implica uma sobrecarga de cálculos para a obtenção do melhor modelo, isto é, para conseguir a convergência nos valores mais adequados que modelam o conjunto de dados inseridos na rodada. Para efectuar os referidos cálculos, o VARBRUL recorre a um algoritmo que, em sequências cíclicas, modifica os valores dos pesos relativos para

tentar alcançar o modelo matemático que mais fiavelmente se aproxime da distribuição constatada nos dados inseridos.

Após cada sequência cíclica, o algoritmo estabelece comparações entre o último e o penúltimo ciclo para efectuar ajustes nos valores do modelo e melhorar este. Quando a diferença entre um ciclo e outro é menor do que o valor limite estabelecido pelo algoritmo, a ferramenta VARBRUL pára as rodadas porque não consegue melhorar mais o modelo. Neste caso, o programa alcançou a convergência do algoritmo, ou seja, alcançou o modelo que mais e melhor se aproxima daquele que é verificado na fala da comunidade. Atingida a convergência, a ferramenta VARBRUL permite não só imprimir os valores dos pesos e do *input* do último ciclo mas também verificar a quantidade de ciclos (“iterações”) a que o algoritmo recorreu até alcançar a convergência.

Quanto mais elevado é o número de variáveis independentes ou de factores inseridos no mesmo grupo de factores, mais sequências cíclicas necessita o algoritmo de efectuar para obter o melhor modelo, ou seja, alcançar a convergência. Para evitar que o algoritmo produza sequências infinitas na busca dos ajustes, o programa VARBRUL permite que aquele force ciclos até à vigésima iteração, mas pára na seguinte, independentemente dos valores alcançados entre o último e o penúltimo ciclo. Nestes casos, a convergência não é alcançada e compete ao pesquisador verificar se tal foi consequência do elevado número de factores ou grupos de factores, ou, pelo contrário, resultou da falta de ortogonalidade entre duas variáveis independentes que compartilham factores, isto é, factores que definem exactamente os mesmos dados.<sup>159</sup>

As situações de falta de convergência podem ser resolvidas reduzindo-se o número de variáveis independentes a serem incluídas nas rodadas VARBRUL, isto é, omitindo os grupos de factores não significativos, ou, em alternativa, efectuando cruzamentos entre os vários grupos (*cross-tabulations*) para verificar quais os factores envasados em grupos de factores diferentes e que poderão estar a compartilhar traços idênticos. No ponto seguinte do presente trabalho (5.2.4) damos conta da metodologia aplicada nas nossas rodadas como vista a eliminar iterações e permitir a convergência dos nossos dados. Ressalve-se ainda que, apesar de termos sempre procurado identificar e resolver a causa da falta de convergência, não foi possível, em certas situações, conseguir uma

resolução definitiva. Contudo, ainda que os resultados obtidos não se revelassem os melhores possíveis nestas situações, o problema não se apresenta como fatal, já que eles se aproximam bastante dos hipotéticos melhores valores, concedendo assim uma indicação fiável dos efeitos das variáveis (cf. Guy & Zilles, 2007).

### **3.8.1. Iteração entre grupos de factores**

A redução de factores dentro de um mesmo grupo de factores pode torná-lo mais robusto para a análise que se pretende efectuar. Contudo, os próprios grupos de factores podem revelar iterações entre si, levando a ferramenta VARBRUL a produzir resultados sem significância e a eliminar da análise variáveis independentes que poderão ser significativas para o estudo em causa, ou, em contrapartida, seleccionar grupos de factores com resultados viciados. Não obstante, cruzando-se grupos que compartilhem traços ou características, podemos constituir novos grupos de factores que se revelem ajustados e significativos, permitindo descobertas importantes.

Paralelamente, poder-se-á também dar o caso de alguns grupos de factores não serem ortogonais, ocorrendo incompatibilidade entre eles. Estas situações não são resolvidas pelo cruzamento de variáveis independentes, e a metodologia aconselha que se efectuem rodadas VARBRUL retirando das mesmas, ora um, ora outro grupo, a fim de a análise alcançar resultados fiáveis e que a tornem válida. Olhando para os grupos de factores constituídos para se efectuar a análise dos nossos dados, verifica-se que as variáveis independentes *posição do item na cadeia do SN (posição linear)*, *posição em relação ao núcleo do SN* e *ordem do item na cadeia dos constituintes flexionáveis* são constituídas por factores que compartilham características. De facto, todos eles comungam traços em termos de efeito de variação, pois reflectem uma determinada posição do item analisado na cadeia linear do SN. Contudo, a amalgamação de dois ou dos três grupos de factores num único factor iria impedir que se cruzasse, independentemente, qualquer um deles com outros grupos de factores, inviabilizando-se a obtenção de alguns achados que entendemos serem bastante significativos para o estudo da CPL-var no SN do PA.

Em estudos sobre a CPL-var no SN, efectuados até à data presente, não temos conhecimento de ter sido analisado a variável independente *posição do item na cadeia dos constituintes flexionáveis*. Por outro lado, o comportamento da variável

independente *classe gramatical* tem sido observado quer conjuntamente quer em cruzamentos com os grupos de factores *posição do item na cadeia do SN* e/ou *posição em relação ao núcleo do SN*. Ora, é nosso entendimento que a variável *classe gramatical* não deixa também de representar, ela mesma, tanto a posição linear como o posicionamento relativamente ao núcleo do SN. De facto, em SN's de estrutura reduzida, como acontece maioritariamente no PA, quando se fala em determinados itens gramaticais, está a inferir-se, logicamente, os dois citados posicionamentos que o elemento gramatical ocupa preferencialmente na cadeia sintagmática. Assim, as referências aos determinantes implicam, por norma, a indicação de elementos pré-nucleares situados na primeira posição linear, as alusões ao nome ou categorias substantivadas implicam, geralmente, a indicação de itens que ocorrem em posição nuclear e em segundo lugar na cadeia linear, e a menção a adjectivos, salvo poucas excepções, está a referir-se a constituintes posicionados pós-nuclearmente e em terceira colocação no SN. Como tal, a nossa inferência leva-nos a concluir que os quatro grupos de factores referidos terão que apresentar interacções caso sejam incluídos, total ou parcialmente, na mesma rodada. E assim sendo, retira-se robustez aos mesmos, levando-os a apresentarem resultados enviesados, em vez dos não-viciados. Aliás, a idêntica conclusão terá chegado Jon-And (2008:3) nas suas análises sobre a CPL-var do PMP, uma vez que excluiu das suas rodadas a variável *classe gramatical*, “*por causa de interferência com outras variáveis*” que a autora não explicita, mas que acreditamos serem precisamente a *posição linear* e a *posição em relação ao núcleo do SN*.

### **3.8.2. Variáveis independentes a serem incorporadas na análise final**

Face à necessidade de preservarmos as quatro variáveis independentes mencionadas, que, de certa forma, reflectem uma determinada posição do item analisado na cadeia linear do SN, mas conscientes também que as mesmas apresentariam iteração caso qualquer delas fosse incluída em simultâneo na análise com uma das outras, optámos por efectuar diversas rodadas, incluindo apenas uma das variáveis independentes em cada uma destas. Assim, para efeitos de análise geral de todos os grupos mantivemos apenas um destes grupos de factores em cada rodada, a fim de conservarmos como modelo a rodada que permitisse uma visão mais abrangente sobre a totalidade dos grupos de

factores, ou seja, aquela que eliminasse o menor número de variáveis independentes. No caso das rodadas que eliminaram igual número de variáveis independentes, levou-se em consideração, para efeitos de análise geral, a que apresentou o nível de significância mais baixo.

Seguindo esta metodologia, apuraram-se resultados não enviesados, tendo as rodadas eliminado os seguintes grupos de factores:

- (i) Variável eliminada na rodada em que se manteve o grupo de factores *classe gramatical: marcação de género*.
- (ii) Variáveis eliminadas na rodada em que se conservou o grupo de factores *posição do item em relação ao núcleo do SN: animacidade do nome e marcação de género*.
- (iii) Variáveis eliminadas na rodada em que se preservou o grupo de factores *posição do item na cadeia do SN (posição linear): marcação de género*.
- (iv) Variáveis eliminadas na rodada em que se manteve o grupo de factores *ordem do item na cadeia dos constituintes flexionáveis: tonicidade e marcação de género*.

A ferramenta VARBRUL considerou a variável linguística independente *marcação de género* como não-significativa para aplicação da regra da CPL-var no SN do PA em todas as rodadas. Assim sendo, este grupo de factores, pela sua nula influência na CPL-var do SN de Almojarife, não será objecto de análise no nosso trabalho.<sup>160</sup> Relativamente às outras variáveis independentes, é possível constatar que as rodadas (ii) e (iv) eliminaram dois grupos de factores, enquanto as rodadas (i) e (iii) eliminaram apenas a variável independente *marcação de género*. Assim sendo, passamos a considerar para efeitos de análise apenas as duas últimas rodadas. De entre estas, a rodada (i) é aquela que apresenta o nível de significância mais baixo (0,011 – Tabela 11.1). Deste modo, passaremos, doravante, a tê-la como modelo para observar o efeito das diferentes variáveis independentes na marcação PL do SN do PA. As três restantes rodadas, isto é, as que mantiveram, em separado, os grupos de factores *posição do item em relação ao núcleo do SN* (nível de significância = 0,036 – Tabela 16.1), *posição do item na cadeia do SN* (nível de significância = 0,038 – Tabela 21.3) e *ordem do item na cadeia dos constituintes flexionáveis* (nível de significância = 0,017 – Tabela 17.1), serão utilizadas quer para se fazer uma leitura do modo como os grupos de factores que



foram retirados da rodada anterior intervêm na marcação PL do SN do PA quer para ajudar a definir o quadro hierárquico de importância decrescente das diferentes variáveis independentes na mesma marcação. Assim, e após observarmos não só os valores das quatro rodadas em questão mas também das rodadas gerais em que se incluíram todos os grupos de factores, estabeleceu-se a escala decrescente de importância das variáveis independentes no desenho da CPL-var do SN do PA:

**Tabela 14.1.** Escala hierárquica decrescente das variáveis independentes seleccionadas como válidas para análise da CPL-var no SN do PA.

<b>Escala hierárquica</b>	<b>Variável independente</b>
01	Posição do item em relação ao núcleo do sintagma nominal
02	Ordem do item analisado na cadeia dos constituintes flexionáveis em número
03	Classe gramatical do item analisado
04	Escolaridade
05	Marcas precedentes ao item analisado
06	Saliência fónica (Processos morfofonológicos de formação plural)
07	Grau de concordância de número no sintagma nominal
08	Posição do item na cadeia do sintagma nominal (Posição linear)
09	Idade
10	Estadia (Permanência fora da comunidade)
11	Sexo
12	Tonicidade
13	Traço semântico do sintagma nominal (Animacidade do nome)
14	Contexto fonológico posterior

No presente capítulo debruçámo-nos sobre a metodologia aplicada ao estudo da variação, feita numa perspectiva atomística, com recurso a métodos quantitativos processados pela ferramenta VARBRUL. Para tanto, não só descrevemos esta mas também a forma como foram recolhidos, tratados e codificados os dados submetidos a análise. Para levar a cabo a análise, foram constituídas uma variável dependente binária

(presença ou ausência de marca PL nos elementos do SN), onze variáveis independentes linguísticas (três fonológicas, sete morfossintáticas e uma lexical) e quatro variáveis independentes extralinguísticas. Para cada uma das variáveis independentes foram constituídos os respectivos factores e formulada uma hipótese, que pretendemos ver confirmada, ou não, nos próximos capítulos. Foi também descrita a metodologia aplicada na quantificação dos dados para, com recurso às primeiras rodadas VARBRUL, se proceder à selecção e escalonamento hierárquico das variáveis independentes que têm intervenção directa na inserção da marcação PL e configuração da regra da concordância de número PL no SN do PA. Assim, os passos seguintes do nosso estudo darão conta do modo como trabalhámos, com recurso à ferramenta VARBRUL, as diferentes variáveis independentes linguísticas e extralinguísticas tidas como confiáveis, a fim de apresentarmos e comentarmos os resultados que permitem observar o modo como estas actuam nas referidas inserção e configuração.



---

<sup>1</sup> Na época renascentista, por exemplo, Dante debruçou-se sobre 14 dialectos italianos, embora os considerasse línguas corrompidas.

<sup>2</sup> No Brasil, a “*variação na concordância de número é bem significativa, chegando a ocorrer entre as camadas mais cultas, e estando amplamente generalizada entre as variedades populares da zona rural, onde o plural no SN é quase que só marcado no determinante*” (Lucchesi, 2000a:19).

<sup>3</sup> Se bem que os conceitos de pidginização/crioulização e transmissão linguística irregular (*imperfect language shift*) digam respeito a processos sociohistóricos de contacto prolongado entre línguas, nos quais intervêm um idioma Y que é tomado como modelo para os falantes envolvidos no contacto, estamos perante dois processos distintos de aquisição: o primeiro envolve a aquisição/criação de uma L2, enquanto o segundo envolve a aquisição/criação de uma L1. Deste modo, o conceito de pidginização/crioulização é menos abrangente do que o conceito de transmissão linguística irregular, visto envolver processos de natureza linguística e sociohistórica que conduzem ao emergir de uma nova língua X (pidgin ou crioulo), a qual se assume como predominante na situação de contacto. Por conseguinte, a pidginização/crioulização pressupõe uma situação de reduzido acesso à língua-alvo Y, determinante para o desencadeamento de um processo de estruturação linguística, que decorre do material primário disponível adquirido e das necessidades estruturais e sociopragmáticas do que está a ser adquirido. Estes aspectos são determinantes para o aparecimento da nova língua X, bastante distinta das que lhe concederam os modelos iniciais para aquisição, uma vez que incorpora itens morfossintácticos, fonológicos, lexicais e semânticos de uma língua Z, e que não foram adquiridos a partir da língua Y. Quanto à transmissão linguística irregular, por configurar um *continuum* de níveis diferenciados de socialização/nativização de uma L2 (a língua X) adquirida em situação sociohistórica específica, apresenta resultados diferenciados entre si, caracterizados por não-realização de morfologia flexional que pode ir de grau mais acentuado a grau mais leve. As questões em torno da transmissão linguística irregular e tipos de variação por ela determinados encontram-se debatidos no ponto 2.11.3 do presente trabalho.

<sup>4</sup> Os falantes do PA terão ainda tido contacto quer com o PtgL2 influenciado por estruturas bantu, falado por africanos “contratados” em finais do século XIX para trabalharem nas roças de cacau e café de São Tomé, quer com o CCV, L1 de trabalhadores enviados, posteriormente, para as mesmas roças. Deste modo, não será também de desconsiderar a possibilidade de o PA apresentar influências decorrentes do contacto com estas línguas.

<sup>5</sup> Do contacto entre o português e os dialectos ameríndios resultaram, essencialmente, contribuições para o léxico do PB. Os linguistas têm sido unânimes em considerar que foi mais forte a ligação entre o PE e as línguas dos africanos, ainda que não seja de desconsiderar a intencionalidade por parte dos colonos em manter a diversidade linguística, a fim de se evitar a unidade nos grupos escravos, mantendo-os submissos (Rodrigues, 1985). Dado que a diversidade foi uma realidade, alguns estudiosos acreditam que as influências dos substratos ameríndios possam estender-se quer ao sistema vocálico, afectando a pronúncia e a entoação, quer ao sistema morfológico (Houaiss, 1985; Baxter & Lucchesi, 1997; Lucchesi, 1999, 2009a; Lucchesi & Baxter, 2006), originando “*a redução de flexões, fenómeno comum aos falares crioulos*” (Elia, 1979:41). Por outro lado, a CPL-var no SN estabelece-se a nível geral após a emergência de uma situação de transmissão linguística irregular no Brasil, quando os falantes dos substratos bantu e kwa, detentores de um código linguístico marcado por ausência formal de concordância a nível do SN, passaram a contactar simultaneamente com o PE e os dialectos ameríndios. Como consequência desse contacto, o sistema de marcação africano, por ser de reduzida flexão, terá sido transposto quer para o PtgL2 defectivo quer para o PBL1 (Guy, 1981a, 1989, 2005), quer ainda para dialectos ameríndios, como acontece na linguagem dos caipiras (Mendonça, 1936).

<sup>6</sup> A deriva é aqui entendida como sinónimo de mudanças fonéticas que condicionam o comportamento morfossintáctico das línguas no decorrer do tempo, originando novas formas, como dialectos, línguas ou blocos linguísticos (Sapir, 1971 [1921]). Deste modo, a *deriva românica* enquanto fenómeno inerente à própria língua, afectou a flexão gramatical das línguas europeias, em geral, e das línguas romance, em particular, levando as segundas a diferenciarem-se do latim. Trata-se, pois, de um fenómeno histórico que, no caso do português, e como esclarece Lucchesi (2000a), conduziu à simplificação morfológica do idioma luso desde as suas origens mais remotas no indo-europeu, com destaque para a perda da flexão casual dos nomes, na passagem do latim para o romance. Idêntico fenómeno ocorreu, por exemplo, nas línguas germânicas, que sofreram mudanças internas que as levaram a tornarem-se diferentes do germânico.

<sup>7</sup> Chomsky usa o “*termo técnico «língua-I» (de «língua interiorizada»), opondo-se ao termo «língua-E» (de «língua exteriorizada»), que refere o conjunto de frases e expressões determinadas pela língua-I*” (Raposo, 1992:27, nota 3), para se referir à linguagem, isto é, ao sistema de conhecimentos interiorizados na mente do homem, e que permite que este reconheça, interprete e produza enunciados na sua L1. Deste modo, o objecto de estudo dos generativistas é a língua-I, não a língua-E.

---

<sup>8</sup> No presente estudo, o conceito de LA é essencialmente sinónimo de idioma que é objecto de apropriação por parte de uma comunidade (aquisição massiva), que tende a substituir a sua L1 pela LA. Neste caso, a comunidade pode, diacronicamente, vir a conceder-lhe o estatuto de L1, isto é, de língua dos seus afectos.

<sup>9</sup> Disponível em: <http://www.instituto-camoes.pt/revista/descbroutro.htm>.

<sup>10</sup> Apesar de admitir que os casos de discordância são frequentes na língua oral, Bechara (1999 [1928]:544) atribui estes ao facto de o fluxo do pensamento ocorrer mais rapidamente do que a formulação e estruturação da oração. Deste modo, o gramático sugere que o falante aproveite cuidadosamente a liberdade de concordância que a língua portuguesa oferece para não prejudicar a clareza da mensagem e a harmonia do estilo.

<sup>11</sup> No estudo da concordância variável de género levado a cabo por Lucchesi (2000a) para o dialecto de HEL-Ba, o linguista deu também conta do fenómeno da analogia, avançando que o mesmo é inerente às situações de *transmissão linguística irregular* e resulta da generalização da interpretação de que as palavras terminadas em *-a* serão femininas quanto ao género.

<sup>12</sup> Guy (1981a) estabeleceu as seguintes variáveis sociais para as suas observações: *sexo, idade (adolescentes, adultos jovens e adultos velhos) e estilo de fala (informal e cuidadoso)*.

<sup>13</sup> No estudo atomístico, cada elemento do SN é visto como um dado de análise, definido pela presença ou ausência de marca de PL. Lucchesi (2000a:22) optou por apelar à análise de “mórfica”.

<sup>14</sup> O estudo sintagmático analisa o SN como um todo (uma única unidade de análise), debruçando-se sobre dois ou mais constituintes extraídos dos dados. Como tal, atribui-se um valor positivo aos SN’s que tenham todos os seus constituintes marcados e um valor negativo aos que contenham pelo menos um elemento sem marcação de número, quando esta deveria figurar no mesmo.

<sup>15</sup> O Princípio do Processamento Paralelo é um dos princípios que contribuem para a elaboração do Processamento Serial (algoritmo incrementativo e determinístico), que determina a aplicação de operações mentais sobre cadeias de símbolos, em que o *output* de uma operação serve de *input* para a próxima. De acordo com o Princípio do Processamento Paralelo, o processo de compreensão engloba estágios distintos: acesso lexical, *parsing*, forma fonética, etc. Tais processamentos ocorrem um de cada vez, sem debordamento. Portanto, num modelo de processamento em paralelo, todos esses processos podem ocorrer simultaneamente, isto é, pode especificar-se foneticamente um vocábulo, ao mesmo tempo que se acessa a outro e se constrói a estrutura sintáctica. No caso concreto em que “*marcas levam a marcas e zeros levam a zeros*” (Scherre, 1988:208), é o efeito do paralelismo sintáctico, consequência talvez de um mecanismo mental, que faz o indivíduo ser levado a repetir e a imitar contextos linguísticos.

<sup>16</sup> Para os considerandos elaborados acerca do conceito de Scherre (1988) e Silva & Scherre (1998 [1996]), sobre a intervenção do Princípio do Processamento Paralelo na marcação PL dos itens do SN, P. Andrade (2003), terá partido do pressuposto que SN se encontra licenciado em SDET (Abney, 1987). Nesta conformidade, o DET é uma extensão da projecção da expressão nominal, sendo aí realizado o morfema de PL, em virtude de assumir a posição de núcleo do SDET.

<sup>17</sup> Os morfemas de conteúdo atribuem semântica aos enunciados e são os únicos cujos lemas contêm a informação necessária para a realização das estruturas de superfície. Por seu lado, os morfemas de sistema dividem-se em duas categorias: morfemas sistémicos estruturais (*early system morphemes*) e morfemas de nível funcional sistémico (*late system morphemes*) (Myers-Scotton, 2002:74). Destes, os primeiros (*early system morphemes*) são activados a nível conceptual para estabelecerem as ligações das intenções semântico-pragmáticas (*chunks*) do falante às estruturas de superfície (*content morphemes*) (Myers-Scotton, 2002:14). Como os lemas dos *early system morphemes* são activados de modo indirecto pelos lemas que suportam os morfemas de conteúdo, ambos são adquiridos de modo mais fácil. O mesmo não acontece com os *late system morphemes*, uma vez que são apenas activados a nível estrutural, ou seja, quando se procuram expandir as estruturas linguísticas.

<sup>18</sup> Baxter (2004:101-102) aponta cinco motivos que levaram o santomense a não penetrar em Monte Café: (i) o facto de os ancestrais escravos terem abandonado as roças em massa, após a abolição da escravatura, no século XIX; (ii) o facto de grande parte dos antigos capatazes das roças ser constituída por europeus; (iii) o facto de os trabalhadores transplantados continuarem a falar as suas L1’s ancestrais, mantendo-as entre os seus descendentes; (iv) o facto de o sistema de controlo das roças não favorecer os contactos com outras comunidades; (v) o facto de a localização geográfica das próprias roças contribuir, também ela, para o hermetismo das mesmas.

---

<sup>19</sup> Variáveis constituídas por Baxter (2004) para análise da CPL-var do PT: (i) saliência da forma morfofonológica; (ii) contexto fonológico seguinte; (iii) marcas antecedentes; (iv) classe gramatical; (v) posição relativamente ao núcleo do SN; (vi) idade; (vii) sexo; (viii) descendência; (ix) educação.

<sup>20</sup> Em 1930, o governo português adoptou o modelo de “assimilação” francês, que procurava, através da língua colonialista, a única permitida no sistema escolar, assimilar e aculturar a população africana não-indígena aos modelos europeus (Barreto, 1977). Para além da fluência no português, outras condições impostas, para se poder ascender à condição de “assimilado”, eram: (i) ter rendimentos regulares; (ii) possuir nível de educação satisfatório (4<sup>a</sup>. classe); (iii) adoptar a monogamia (Firmino, 1995:260, nota de rodapé 8).

<sup>21</sup> A implementação de uma política de língua e educacional consistente apenas emergiu em 1930, quando o governo português adoptou o modelo de “assimilação” francês (Barreto, 1977). Deste modo, a primeira escola primária só viria a ser criada no final do século XIX, sendo o seu número, em meados do século seguinte (1945), apenas de 92 (Mateus, 1999:212).

<sup>22</sup> Na análise VARBRUL, as probabilidades devem ser observadas em função dos pesos relativos e não dos percentuais, já que estes não são indicadores fiáveis para os efeitos exercidos pelas variáveis linguísticas e extralinguísticas no fenómeno em estudo.

<sup>23</sup> Os informantes que produziram o *corpus* de Jon-And (2008) são todos falantes que têm uma língua bantu como L1 (xironga, confinado a cerca de 650000 indivíduos na região do Maputo, ou xichangana, falado por 11,4% da população moçambicana) (Instituto Nacional de Estatística de Moçambique, 2010).

<sup>24</sup> O povoamento das ilhas de Cabo Verde iniciou-se no Sul, a partir de Santiago, em 1462. Seguiu-se-lhe o do Fogo e, depois, o da Brava. Santo Antão e S. Nicolau, no lado Barlavento, foram povoadas com escravos trazidos de Santiago e do Fogo, no século XVII, ao passo que só aconteceu um verdadeiro povoamento de S. Vicente a partir dos finais do século XVIII (Carreira, 1979; Santos, 2002).

<sup>25</sup> Segundo Ferraz (1975:36), a ausência do artigo definido nos CP's da Alta Guiné e do Golfo da Guiné deve-se à influência dos substratos africanos.

<sup>26</sup> O desenquadramento da tradução em português visa conservar o espírito original da produção no crioulo.

<sup>27</sup> Definido como *Lexicon*, na perspectiva de Chomsky (1996 [1995]:6).

<sup>28</sup> Dentro destas funções encontram-se processos como a formação de palavras, armazenamento de correspondência imprevista na relação som/significado e de formas morfofonológicas irregulares ou a geração das unidades atómicas da análise sintáctica (Silva, 2010:19).

<sup>29</sup> O clássico conceito de crioulo exógeno ou de plantação (*plantation creole*), proposto por (Holm, 1988/1989), isto é, de crioulo que emerge entre falantes de distintas L1's geográfica e culturalmente deslocados por colonos, é questionado à luz da crioulistica actual (p.e. McWhorter, 2010), que só admite a emergência de crioulos endógenos (*fort creoles*). Assim, os crioulos apenas se constituem em regiões multilingues, às quais aportam grupos dominantes vindos de fora para trocas comerciais, fazendo com que os naturais dessas regiões e das áreas circundantes se associem a si cultural e economicamente. Desta forma, os indígenas, ainda que não deixem de falar as suas L1's, necessitam também de criar um código linguístico de emergência para comunicarem não só entre si mas também com os grupos estrangeiros, detentores da LA. Pelo contrário, nas situações de plantação, os escravos transplantados já são detentores de códigos linguísticos de emergência formados endogenamente (pidgins ou proto-crioulos), que se expandem e desenvolvem no seio das comunidades dos falantes de distintas L1's.

<sup>30</sup> Tendo em conta que o edo (Nigéria) é o mais antigo e mais significativo substrato dos CGG's, a etimologia de *inen* radicar-se neste dialecto, apontando Alexandre & Hagemeyer (2007:40, nota 5) três possíveis origens para a mesma: (i) terceira pessoa PL do pronome fraco e forte *iran* “eles”, do qual *inen* é homófono; (ii) determinante definido (*n)ene* “o”; (iii) pronome demonstrativo *ene/enena* “estes/estas”, aparentemente relacionados a (ii) (cf. Melzian, 1937; Agheyisi, 1986, 1990).

<sup>31</sup> Exemplo disponível em: <http://www.abeokuta.org/yoruba.htm> [Consultado em 24 de Maio, 2010].

<sup>32</sup> Exemplos disponíveis em: <http://www.linguakimbundu.com/> [Consultado em 24 de Maio, 2010].

<sup>33</sup> Alguns nomes do santomense, a exemplo do que sucede nas línguas bantu, possuem formas masculina e feminina distintas, como é o caso do substantivo *sogru* > *sogla* (sogro > sogra) (Ferraz, 1979:60). Estes aspectos encontram-se discutidos no ponto 1.3.2.2.1 do presente trabalho.

---

<sup>34</sup> Os exemplos extraídos do *corpus* do PT e que surgem referenciados com o nome do informante foram-nos fornecidos por Alan N. Baxter, que dirigiu o projecto “*Semi-creolization: testing the hypothesis against data from Portuguese-derived languages of São Tomé (Africa)*”. Informações detalhadas sobre o projecto poderão ser consultadas no capítulo 2 do presente trabalho, dedicado à metodologia aplicada no mesmo.

<sup>35</sup> A origem etimológica de *ũa* é atribuída ao indefinido “uma” do português (Alexandre & Hagemeyer, 2007:39, nota 5), que detém o mesmo significado do santomense. Neste, o determinante indefinido e o numeral *ũa* são homónimos.

<sup>36</sup> “Ka” é um morfema aspectual largamente difundido no santomense e, ao que tudo indica, tem origem nas línguas kwa. É utilizado não só para modificar o próprio verbo mas também conjuntamente com outros verbos, já que a conjunção coordenativa copulativa “e” não existe no forro. Deste modo, as orações são ligadas umas às outras pelo método de aposição e, caso o sujeito seja o mesmo para todas elas, a sua retenção é obrigatória apenas na primeira oração.

<sup>37</sup> Exemplos disponíveis em: <http://www.linguakimbundu.com/> [Consultado em 29 de Outubro, 2009].

<sup>38</sup> Exemplo disponível em: <http://vidademacumbeiro.blogspot.com/2008/12/gramtica-yoruba.html> [Consultado em 25 de Agosto, 2010]

<sup>39</sup> Por consoante homorgânica entenda-se aquela que é idêntica a uma outra, em relação ao modo de articulação, à zona de articulação e ao papel das cavidades bucal e nasal, mas diferente quanto à sonoridade. Assim, uma é surda e a outra sonora. Tendo como exemplo a consoante /p/, vemos que ela partilha os seus fechos (traços fónicos) de *oclusividade*, *bilabialidade* e *oralidade* com a consoante /b/. Contudo, /p/ possui o traço de não-sonoridade (é surda), enquanto /b/ é marcada pelo traço de sonoridade (é sonora). O mesmo ocorre com outras consoantes, como /f/ e /v/. Ambas são constritivas-fricativas, labiodentais e orais, mas a primeira é surda e segunda sonora. Na produção de /v/ ocorre vibração das cordas vocais, o que não sucede com a articulação de /f/. Listam-se ainda outros pares homorgânicos: /t/ e /d/; /k/ e /g/; /s/ e /z/; /ʃ/ e /ʒ/.

<sup>40</sup> A vogal morfofonémica [i] é definida por Ferraz (1979:20), em ocorrências no santomense, como “*a nasal central high vowel with lip spreading. Nasality is the only linguistically significant feature of this nucleus. [i] occurs only in two instances in the language: in optional variation with /m/ ~ /n/ [n, ŋ] as a locative morpheme, and as the 1st p.s. subject pronoun. [i] is thus the representation of a morpheme whose main phonological identity is nasality*”.

<sup>41</sup> A dobra é a unidade monetária de S. Tomé, introduzida na ilha em 1977 para substituir o escudo português.

<sup>42</sup> Jornal de São Tomé e Príncipe [em linha]. Disponível em: <http://www.jornal.st/directorio.php?page=173> [Consultado 12 Dez. 2009].

<sup>43</sup> Artigo 13º da lei 53/88.

<sup>44</sup> A ausência de um conceito claramente definido para zona urbana em São Tomé e Príncipe levou o Instituto Nacional de Estatística do país a optar por considerar esta como sendo uma área que possua uma “*aglomeração de casas com algumas infra-estruturas, tais como arruamentos, asfaltos, luz eléctrica nas casas e nas ruas, canalização de água, redes de esgotos, e em que a actividade económica predominante sejam os serviços e / ou a indústria. Também todas as capitais distritais foram consideradas como urbanas*” (Instituto Nacional de Estatística de São Tomé, 2003a:8).

<sup>45</sup> Outro dos factores que contribui para que o número de mulheres seja superior ao dos homens é a maior esperança de vida actual das primeiras (Instituto Nacional de Estatística de São Tomé, 2003a:14).

<sup>46</sup> Matabichar: termo usado nas variedades de português africano, com o significado de “tomar o pequeno-almoço”.

<sup>47</sup> O informante refere-se ao tempo em que o Comité de Libertação de São Tomé e Príncipe, posteriormente denominado Movimento de Libertação de São Tomé e Príncipe (MLSTP), desenvolvia acções de luta contra as autoridades coloniais portuguesas.

<sup>48</sup> Os estudos iniciais sobre ASL partiram do pressuposto que era mais fácil adquirir o que é aproximado à L1 e mais difícil aprender aquilo que se afasta dela (Lado, 1957). Corder (1967; 1971; 1974; 1981b) constataria este pressuposto, afirmando que a interferência pode provir de outra língua semelhante ou intermediária. Corder (1967) atribuiria também a denominação de *idiosincratic dialect* (dialecto idiosincrático) ao sistema linguístico da L2, termo que Selinker (1972:216) modificaria para *interlanguage* (interlíngua).

---

<sup>49</sup> A Hipótese do Período Crítico (*Critical Period Hypothesis – CPH*) para aquisição de uma língua, proposta por Lenneberg (1967), é de carácter marcadamente neurológico, advogando que as mudanças que se operam a nível cerebral impedem os adolescentes e adultos de accionarem na perfeição as capacidades inatas para aquisição de línguas, o que não sucede com as crianças. A abrangência deste conceito e as incertezas em torno da proposta do CPH não deixam que se determine com exactidão os seguintes aspectos: (i) se a habilidade para aquisição declina somente no período crítico; (ii) se a mudança no ponto crítico se dá brusca ou gradualmente; (iii) se o ponto crítico tem influência em todos os aspectos da linguagem ou se apenas em alguns; (iv) se o ponto crítico é meramente biológico ou secundado por outros aspectos que intervenham na capacidade da aprendizagem.

<sup>50</sup> As siglas *FLO* e *FLf* dizem respeito à terminologia adoptada por Raposo para a tradução de Chomsky (1999a). Chomsky (1996 [1995]) recorreu às siglas S0 (estado inicial) e S8 (*estado estável*), respectivamente. Por seu lado, Silva (2010) optou pelas siglas S0 (cf. Chomsky, 1996 [1995]) e Sp, no sentido de Língua Produto, conforme mencionado frequentemente por João Costa durante o seminário “Tópicos em Sintaxe Generativa”, ministrado na Universidade de Macau, entre 25 e 31 de Março, 2008.

<sup>51</sup> A este propósito, o *Princípio do Subconjunto* defende que “se um parâmetro tem dois valores «+» e «-», e o valor «-» determina uma linguagem que é subconjunto da linguagem determinada pelo valor «+», então o valor «-» é o valor não-marcado, seleccionado pela GU na ausência de qualquer evidência linguística” (Chomsky, 1986a:146). É o que acontece no caso da *informação negativa* (incluindo a *informação negativa indirecta*), que não desempenha nenhum papel na aquisição da linguagem se o valor não marcado do *parâmetro* for a posição *não-pro-drop* (realização fonética obrigatória do sujeito), já que o valor inicial do *parâmetro do sujeito nulo* é a posição *não-pro-drop*. Deste modo, os parâmetros possuem um valor inicial não-marcado, que apenas é alterado durante a aquisição caso ocorra evidência positiva em contrário nos DLP’s. Nesta conformidade, o valor que desencadeia os movimentos, a partir dos núcleos funcionais, é o valor «+», uma vez que os movimentos são responsáveis por representações relativamente complexas. Por outro lado, a noção de marcação será um conceito teórico-I, uma vez que tem a ver com o modo como é aplicado o módulo das regências e ligações.

<sup>52</sup> Os termos técnicos “movimento” e “combinação” radicam nos conceitos “*move*” e “*merge*”, avançados por Radford (1997:61) na teorização sobre o minimalismo.

<sup>53</sup> Os sons que as crianças produzem nesta fase não são considerados manifestações de língua, uma vez que carecem de criatividade. De facto, representam apenas respostas involuntárias do recém-nascido a situações de desconforto (como a fome) ou conforto (por exemplo, a necessidade de ser embalado).

<sup>54</sup> As frases holofrásticas contêm uma única palavra, cujo significado se pode alargar para abranger uma classe mais ampla. Na fase holofrástica, as crianças detêm um vocabulário de cerca de 20 palavras, que utilizam para servir três funções principais: exprimir o desejo de actividade (por exemplo, “colo” pode significar o desejo de ser levado ao colo); manifestar emoção (por exemplo, “não” pode significar “*Não gosto!*”); e servir para denominar algo (por exemplo, “cão” pode significar “*Aquilo é um cão.*”). Paralelamente, a criança usa também as palavras isoladas já com fins sociais. De facto, e voltando a pegar na palavra “não”, verificamos que a mesma pode, neste caso, ser utilizada quer com sentido imperativo (“*Não me dês sopa!*”) ou assertivo (“*Não gosto de sopa.*”).

<sup>55</sup> “... the trigger, or PLD (*primary linguistic data*) must consist only of the kinds of things that children routinely experience and includes only simple structures” (Lightfoot, 1999b:51).

<sup>56</sup> Em psicologia, o conceito de *priming* é aplicado às situações em que um estímulo precoce influencia estímulos posteriores. Os testes neste sentido têm evidenciado que, por exemplo, quando um indivíduo lê uma lista de palavras que incluiu um determinado vocábulo (p.e. comida), caso lhe seja pedido que complete uma palavra iniciada com letras pertencentes a esse vocábulo (p.e. com\_), a probabilidade de ele escrever esse mesmo vocábulo é maior do que se não o tivesse lido anteriormente. Outro exemplo de *priming* envolve pessoas a quem é mostrada uma figura inacabada, que eles não são capazes de identificar. Seguidamente, são-lhes apresentados mais elementos da figura, até que eles a consigam reconhecer. Posteriormente, se os elementos da figura lhes forem novamente exibidos, elas identificá-los-ão mais precocemente do que da primeira vez (Kolb & Whishaw, 2003:453-457).

<sup>57</sup> A alternância de códigos pressupõe o uso alternativo de duas ou mais línguas na mesma situação de conversação. Genericamente, o termo diz respeito a diferentes formas de bilinguismo, já que a alternância pode acontecer nos actos de fala de vários indivíduos durante uma conversação, nas elocuições dentro de um único acto de fala e ainda dentro de uma simples elocução (Milroy & Muysken, 1995).



---

<sup>58</sup> Os padrões morfológicos e sintáticos das alternâncias linguísticas são regidos por regras gramaticais específicas e fronteiras especiais nas quais a alternância de códigos pode ocorrer. Assim, a literatura da especialidade tem apontado alguns constrangimentos que impedem a alternância de códigos:

- (i) o “Constrangimento do Morfema Livre” (*Free-morpheme Constraint*), que determina que a alternância de códigos não pode ocorrer entre morfemas presos (Sankoff & Poplack, 1980);
- (ii) o “Constrangimento da Equivalência” (*Equivalence Constraint*), que advoga que a alternância de códigos apenas pode acontecer em posições onde a ordem de dois dos elementos da frase (um antes e outro depois da alternância) não é excluída em nenhuma das gramáticas (Sankoff & Poplack, 1980);
- (iii) o “Constrangimento das Classes Fechadas” (*Closed-class Constraint*), que entende que os itens que constituem classes fechadas não podem ser alternados (Joshi, 1985);
- (iv) o “*Matrix Language Frame Model*”, que faz a distinção entre as regras das línguas detidas pelo falante (Myers-Scotton, 1997);
- (v) “Constrangimento do Núcleo Funcional” (*Functional Head Constraint*), que indica que a alternância de códigos não pode ser aplicada entre um núcleo funcional (um determinante, uma flexão, etc.) e o seu complemento (SN, SV, etc.) (Belazi, Rubin & Toribio, 1994).

No caso do “*Constrangimento da Equivalência*”, Sankoff & Poplack (1980:18) apresentam o seguinte exemplo, para justificarem o seu pressuposto: *I like you porque eres simpatico* (“*I like you because you are likable*”). Como se verifica, ocorre alternância de código porque há uma obediência às regras de formação das orações de ambas as línguas (orações causais no exemplo fornecido para o espanhol e o inglês), que é exactamente o que sucede nas situações apontadas para o PA. Note-se ainda que o princípio do “*Constrangimento das Classes Fechadas*” (Joshi, 1985) também é contemplado no PA, já que não detectámos nenhuma situação de alternância envolvendo, por exemplo, preposições.


<sup>59</sup> O projecto DUFDE (*Deustsh und Französisch – Doppelter Erstspracherwerb*), detalhadamente descrito por Köppe (1994a), foi desenvolvido entre 1986 e 1992, sempre sob coordenação de Jürgen Meisel, e levou a cabo estudos longitudinais sobre o desenvolvimento linguístico de crianças bilingues em alemão e francês, desde o seu início (1;0 – 1;6 anos) até à idade aproximada dos 5 anos. A colecta dos dados era efectuada a cada duas semanas, tendo o projecto baseado a sua metodologia de selecção de famílias no “princípio de uma pessoa/uma língua”, enquanto pressuposto inerente ao desenvolvimento de um bilinguismo equilibrado. Desta forma, todas as crianças eram filhas de pai alemão, que com elas só comunicava em alemão, e mãe francesa, que com elas só comunicava em francês. Inicialmente, a pesquisa incidiu sobre o desenvolvimento linguístico de treze crianças, mas seis destas foram posteriormente descartadas do estudo, já que não apresentavam um bilinguismo equilibrado, como era exigido pela investigação. Todavia, os registos produzidos por estas crianças não foram abandonados, sendo também utilizados para efeitos comparativos. Quanto às restantes sete crianças, o seu desenvolvimento linguístico foi seguido até final do tempo previsto, embora apenas sejam contempladas nas análises dos estudos editados por Meisel (1994b) as produções de cinco dessas crianças.

<sup>60</sup> Alguns autores (Vainnika & Young-Scholten, 1994, 1996a, 1996b, *apud* Godinho, 2005:107) defendem mesmo que apenas a GU e parte da L1 têm papel interventivo no estado inicial da ASL.

<sup>61</sup> O Modelo do Restabelecimento dos Parâmetros proposto por Schwartz & Sprouse (1996) e desenvolvido por Sprouse (2006) assemelha-se ao da relexificação, avançado por Lefebvre (1986, 1998) para defender que a crioulição é um processo mental e sociolinguístico. Contudo, a referida semelhança dá-se unicamente a nível do conteúdo, diferindo na forma. Deste modo, na relexificação, o estágio final do modelo interlíngua (o crioulo) é semelhante ao do estado inicial da aquisição da L1. Segundo Sprouse (2006), no modelo de transferência em bloco da gramática da L1, a génese dos crioulos é uma espécie de aquisição da L2 em reverso. Assim, o crioulo é comparável a uma variedade fossilizada de L2 do superstrato. No entanto, a hipótese parece pouco razoável, porque todos os crioulos de uma determinada base teriam que reflectir, na totalidade, iguais estágios iniciais de ASL.

<sup>62</sup> A Condição de Visibilidade determina que a teoria temática se associe à teoria do caso, advogando o seguinte: *um elemento só é visível para atribuição de um papel temático se estiver numa posição com caso ou se for PRO*.

<sup>63</sup> De acordo com o processamento da sintaxe observado por Pienemann (2000) no inglês, Plag (2008a:123) elabora a seguinte tabela hierárquica de aquisição para este idioma L2:

Desenvolvimento	Estrutura	Exemplo
Estágio inicial	Realizações holofrásticas	<i>Ball.</i>
 LA	Ordem canónica (SVO)	<i>John kick ball.</i>
	Neg + V	<i>He no like coffee.</i>
	Frase iniciada com advérbio	<i>Then Bob kick ball.</i>
	Topicalização	<i>That I did'n like.</i>
	Frase iniciada com “do”	<i>Do you like it? / Do she like it?</i>
	Inversão “yes/no”	<i>Has he seen you?</i>
	Inversão da cópula	<i>Where is John?</i>
	Partículas verbais	<i>Take the hat off</i>
	“Do” / Auxiliar 2ª posição	<i>Why did she sell that car?</i>
		<i>Where has he gone?</i>
	Cancelamento da inversão	<i>I wonder why he sold that car.</i>

Os aprendentes iniciam a aquisição com realizações holofrásticas e vão, posteriormente, adquirindo estruturas mais complexas, mas seguindo sempre um padrão bastante definido. Alguns dos aprendentes conseguem chegar ao estágio em que realizam estruturas complexas e outros alcançam mesmo o patamar que lhes permite inverterem as orações subordinadas interrogativas.

<sup>64</sup> A emersão tardia da flexão inerente pode também ser constatada no desenvolvimento da flexão verbal de tempo e aspecto. Segundo Bardovi-Harlig (2000:25), a aquisição da morfologia verbal na ASL requer três estágios de desenvolvimento, adquirindo os falantes as marcas de flexão apenas na última fase. Antes, os falantes limitam-se a recorrer a estratégias pragmáticas (p.e. ordem canónica), no primeiro estágio de aquisição, passando a usar formas lexicais para marcar as distinções temporais e aspectuais na segunda etapa do desenvolvimento da interlíngua (p.e. advérbios de tempo ou conectores fráscos).

<sup>65</sup> Note-se que a noção de protótipo não implica que o mesmo seja categórico. De facto, o protótipo assume grau de variação nos membros que constituem uma certa categoria ou classe. Tal variação fica a dever-se à dinâmica dos sistemas linguísticos, da mesma forma que acontece em determinadas áreas científicas, como a biologia (Clements, 2009:4).

<sup>66</sup> Para uma consulta mais pormenorizada acerca dos tipos de desencadeamento (*bootstrapping*) envolvidos na aquisição de L1 e L2, consulte-se Bates & Goodman (1999).

<sup>67</sup> Segundo Bechara (1999 [1928]:122-123), estamos perante um caso de sincretismo e não de neutralização, pelo que “*não se pode aceitar a doutrina corrente que vê nessas formas um singular que permanece invariável no plural*”. De facto, a pluralidade surge marcada através do adjunto nominal e, em determinados casos, apesar de existir a forma PL, recorre-se normalmente àquela que é invariável (p.e. *simples* possui o PL *simpleces* e *cós* tem o PL *coses*).

<sup>68</sup> O PL deste tipo de substantivos abstractos produz uma alteração semântica, denotando uma repetição ou multiplicidade de actos, com conotação aumentativa (Brunot, 1989).

<sup>69</sup> Informações pormenorizadas sobre os substantivos massivos e seu comportamento podem ser consultadas em Oliveira (2003:151-152) e Duarte & Oliveira (2003:219-221/236-242).

<sup>70</sup> Bechara (1999 [1928]:115-116) chama a atenção para a diferença entre colectivos universais e particulares e os colectivos que representam nomes de grupos, definidos apenas como colectivos pela gramática tradicional. Assim, são colectivos os que representam nomes de grupos, os conjuntos de pessoas (p.e. *quadrilha*; *elenco*), os grupos de animais (p.e. *matilha*; *ninhada*) e os grupos de coisas (p.e. *cacho*; *mobiliário*). Nestes casos, estamos perante um conjunto de elementos contáveis de uma espécie definida (os grupos de animais), ou total ou parcialmente indefinida (os grupos de pessoas e os grupos de coisas) que, ao contrário dos colectivos universais e particulares, “*requerem a determinação explícita da espécie de objectos que compõem o conjunto*”. Deste modo, é necessário esclarecer se estamos perante *um bando de ladrões* ou *um bando de pardais*, mas não é necessário especificar que estamos perante *um casario de casas*.

<sup>71</sup> Remetemos para Bechara (1999 [1928]:127-128) informações detalhadas acerca destes plurais, nomeadamente no que respeita à regra ortográfica oficial.

---

<sup>72</sup> Os nomes próprios de pessoas, quando usados referencialmente, são obrigatoriamente precedidos de um artigo definido no PE:

- a. *O João leu um livro.*
- b. \**João leu um livro.*

São exceção à regra os nomes próprios de personagens históricos ou que fazem parte da memória colectiva portuguesa, especialmente quando referenciados em discurso literário ou histórico relativamente formal. Neste caso dispensam o uso do artigo definido:

- a. *Diogo Cão foi o primeiro navegador português a chegar à foz do Rio Zaire.*
- b. ?*O Diogo Cão foi o primeiro navegador português a chegar à foz do Rio Zaire.*

<sup>73</sup> Exemplos avançados por Godinho, 2005:115:

*Isto não passa de uma **simples/mera** dor de cabeça.*  
\**Isto não passa de uma dor de cabeça **simples/mera**.*

<sup>74</sup> Exemplos retirados de Casteleiro (2001, vol. G-Z:3417).

<sup>75</sup> Para uma visão acerca da literatura produzida sobre concordância, consulte-se:

[http://www.surrey.ac.uk/LIS/SMG/projects/agreement/agreement\\_bib\\_unicode.htm](http://www.surrey.ac.uk/LIS/SMG/projects/agreement/agreement_bib_unicode.htm).

<sup>76</sup> Conforme demonstrado por alguns autores (Rizzi, 1982; Jaeggli, 1984; Burzio, 1986; Jaeggli & Safir, 1989), as línguas de sujeito nulo evidenciam o seguinte feixe de propriedades:

a. Sujeitos sem matriz fonológica:

(1) PTG:  
*(Eles) Telefonaram.*

(2) INGLÊS:  
\**(They) called.*

b. A possibilidade de o sujeito ocorrer em posição pós-verbal (“inversão livre”):

(3) PTG:  
a. *O João telefonou.*  
b. *Telefonou, o João.*

(4) INGLÊS:  
a. *John called.*  
b. \**Called John.*

c. Ausência de *efeitos vestígio-que* (*that-trace effects*). Em caso de extracção do sujeito, esta dá-se a partir da posição pós-verbal e não da posição pré-verbal (Campos, 1997:98):

(5) PTG:  
Quem disseste que comprou um computador?

(6) INGLÊS:  
\**Who did you say that bought a computer?*

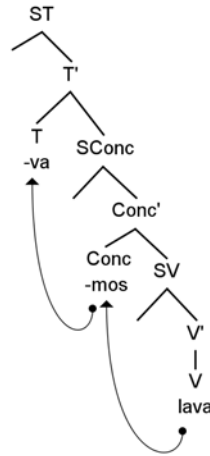
A ausência de *efeitos vestígio-que*, verificada em (5), ocorre em todas as línguas românicas de sujeito nulo (Jaeggli & Safir 1989:37). Por seu lado, Rizzi (1982:174) e Jaeggli (1984:139) entendem que a ausência de *efeitos vestígio-que* se deve ao facto de o sujeito, nas línguas de sujeito nulo, ser extraído directamente da posição pós-verbal e não da posição pré-verbal. Burzio (1986:165) comprova-o em dados do italiano, que indicam que a extracção do sujeito se dá, invariavelmente, a partir da posição pós-verbal e nunca da posição pré-verbal.

<sup>77</sup> As palavras resultantes de incorporação assemelham-se às formadas por composição, mas distinguem-se destas por manterem o seu referencial. Pelo contrário, os elementos compostos são esvaziados do seu conteúdo referencial (p.e. *o guarda-fatos* > *os guarda-fatos* > \**as guarda-fatos*).

<sup>78</sup> Tradução nossa.

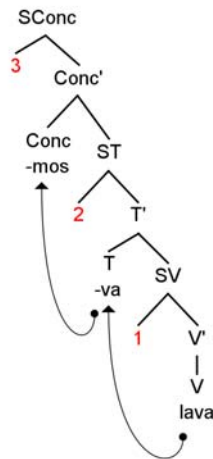
<sup>79</sup> Por *merge* entendam-se as operações de concatenação estrutural. Raposo (1992) atribui ao termo a denominação “*compor*”.

<sup>80</sup> Pollock (1989:386), verificando que o movimento do verbo estabelece correlações com a morfologia de tempo e concordância, mas também que o posicionamento necessita de dois núcleos independentes entre o SC e o SV, sugere que o verbo é formado na sintaxe. Contudo, a hierarquia proposta permite a produção de formas agramaticais, como acontece como a forma portuguesa \**lavámosva*:



- Hierarquia de Pollock (1989) -

Partindo da proposta do PEH (Baker, 1985), Belletti (1990:63) sugere a inversão da hierarquia de Pollock (1989:386), estabelecendo um movimento curto do verbo para T e estando as posições 1, 2 e 3 destinadas ao sujeito:



- Hierarquia de Belletti (1990) -

Apesar de a posição 2 estar destinada ao sujeito, este passa pela mesma sem se deter. Por outro lado, o verbo não vai até à flexão. Assim, pese embora o modelo permitir a realização da forma portuguesa *lavávamos*, não consegue justificar, por exemplo, as situações de inserção da marca de terceira pessoa do PL 's em inglês, já que esta se posiciona em FLEX, isto é, numa posição mais baixa. De qualquer forma, a questão revela aproximações às abordagens estabelecidas na MD (Halle & Marantz, 1993).

<sup>81</sup> De acordo com as assunções de Chomsky (1993), a concordância não tem que ser necessariamente visível, quando ocorre este tipo de configuração. De acordo com o quadro teórico *chomskyano*, a concordância representa uma conjunto de traços(-phi) abstractos, presentes nos núcleos CONC e nos elementos superficiais das posições SPEC, podendo a correlação que ocorre entre tais traços ser visível, ou não.

<sup>82</sup> Note-se que os movimentos não visíveis ou aparentes (*covert movements*) da FL não contam como movimentos reais.

---

<sup>83</sup> As dez categorias de mecanismos gramaticais minimalistas propostas pelo Programa Minimalista são: artigos, formas de TMA, palavras interrogativas, pluralizadores, pronomes de todas as pessoas e números, formas para marcar os casos oblíquos, preposições locativas gerais, um complementador irrealístico, partículas relativizadoras, reflexivos e recíprocos.

<sup>84</sup> Os nomes nucleares, além de próprios ou comuns, podem também ser coordenados (p.e. *Os contos e poemas que estudámos são interessantes.*) e compostos ou lexicais (p.e. *O meu guarda-fatos é de pinho.*).

<sup>85</sup> Os pronomes pessoais podem também constituir todo o SN da frase (p.e. *Tu vais ao teatro amanhã?*), já que não admitem especificadores nem complementos (à excepção de epítetos que se podem considerar um tipo especial de complementos), possuem como referência as pessoas do discurso ou aquela(s) de quem se fala (conhecida(s) dos interlocutores) e são identificadores, tal como sucede com nomes próprios (Brito, 2003a:351-352).

<sup>86</sup> Os estudos de Faria, Freitas & Miguel (1998, 2001) revelaram que a estrutura prosódica das crianças, em fase inicial da aquisição do PtgL1, se comporta de forma idêntica no SN e no SV. No caso do primeiro, “*a presença de informação morfo-sintáctica numa sílaba em final de palavra leva a criança a seleccionar a sílaba não tónica para fazer surgir a fricativa em coda de sílaba em final de palavra, que marca o plural em português, contrariando o padrão do desenvolvimento silábico*” (Godinho, 2005:139). Relativamente ao SV a fricativa morfológica em coda de sílaba é interpretada como morfema de pessoa.

<sup>87</sup> Ferrari-Neto, Corrêa & Augusto (2005) descartam a alternativa proposta por Magalhães (2004:169), que defende que o traço interpretável de número se encontra localizado apenas no DET.

<sup>88</sup> Aronoff (1976) apresenta um modelo voltado para o problema da produtividade lexical, que determina que quanto mais regulares forem os produtos de uma Regra de Formação de Palavras (RFP), maior será a sua produtividade. Como exemplo, compara as formas nominalizadas deadjectivais em inglês formadas com os sufixos *-ness* e *-ity*, apontando que, enquanto o acréscimo de *-ness* não produz qualquer alteração fonológica nas bases, o mesmo não acontece com o acréscimo de *-ity*. Além disso, as formas *X-ness* são semanticamente previsíveis, isto é, têm os significados previstos na RFP, enquanto as formas *X-ity* podem assumir significados idiossincráticos ou perder um ou outro significado previsto na Regra de Formação de Palavras. Assim, para o autor, a regra de formação de substantivos deadjectivais em *-ness* é mais produtiva do que a que forma a mesma classe de palavras em *-ity*. O autor propõe também que, diante da total regularidade das formas *X-ness*, elas não precisariam de estar listadas no léxico, uma vez que podem ser geradas por regra, ao passo que as formas *X-ity*, por não serem inteiramente regulares, têm de estar listadas no léxico. Por seu lado, Spencer (1991) refere que a produtividade também é atestada com o accionamento da regra em situações de emergência comunicativa, as chamadas formações esporádicas. De acordo com as condições morfossemânticas, o falante, perante a presença de itens lexicais envolvidos no processo mais uma situação discursiva do momento, pode criar uma palavra para designar um lugar definido pela situação. É o que acontece no português, por exemplo, nas situações em que emergem formas novas com o afixo *-dromo*, sempre proparoxítonas. Observando a formação e produtividade, Salles & Mello (s.d.) registam que, no estágio actual do PB, se está em fase de produtiva deste afixo (para a RFP *X-dromo*), o que determinou o aparecimento de palavras como sambódromo, ficódromo (lugar onde os namorados ‘ficam’, isto é, onde ficam juntos para namorar), beijódromo (lugar onde casais se beijam), fumódromo (lugar onde se pode fumar), etc. Contrariamente, formas derivadas com os afixos *-este* e *-estre*, como ‘agreste’ e ‘silvestre’, deixaram de ser atestadas diacronicamente em novas derivações. Todavia, note-se que a improdutividade que acontece na RFP *X-este* e *X-estre* constitui improdutividade da regra, e não do afixo. Estes exemplos apontam ainda para a relação entre a diacronia e a produtividade nas línguas, cuja discussão requer o exame das condições de formação dos itens no período diacrónico em que os elementos lexicais são produzidos.

<sup>89</sup> Müller (1994a), no âmbito dos estudos do projecto DUFDE (*Deutsch und Französisch – Doppelter Erstspracherwerb*), discute a emergência dos traços gramaticais de género e número na gramática de duas crianças em fase de aquisição bilingue das línguas alemã e francesa como L1’s, com vista a verificar as consequências sintácticas das opções feitas com base nas especificações lexicais dos núcleos. Segundo a autora, a gramática inicial das crianças não apresenta marcação sistemática de género e número, verificando-se a primeira por volta dos 2;0 anos e a segunda cerca dos 2;4 anos.

<sup>90</sup> As questões em torno da possibilidade de a sequência DET+N ter sido fonologicamente interpretada como um todo indistinto (Koehn, 1994:49; DeGraff, 1999b:482; Faria, Freitas & Miguel, 2001:53) voltarão a ser debatidas no ponto 4.1.2.6.2 do presente trabalho (Relação entre *posição linear* e *posição em relação ao núcleo do SN* na marcação PL dos itens do SN).

<sup>91</sup> Ao contrário do que sucede no PE, que admite a estrutura VSO, no PB e no PA, a alteração de ordem padrão que determina a estrutura SV acontece apenas com verbos inacusativos.

---

<sup>92</sup> Para Selinker (1992:226), recaída é sinónimo de aparecimento repentino de um sistema da interlíngua que há muito se encontrava ausente do desempenho do falante. O fenómeno afecta estruturas que pareciam ter sido erradicadas da gramática da interlíngua, mas é perfeitamente entendido pelo falante, que pode proceder à sua correcção sem necessidade da intervenção de terceiros. Este processo distingue-se da fossilização, uma vez que esta não é uma forma desviada esporádica, antes representa uma variação inalterável, o que impossibilita a sua correcção.

<sup>93</sup> Hale (1998) levantou a questão do “problema de Orwell”, que se debruçou sobre as dificuldades da aquisição de certos tipos de conhecimentos linguísticos, ainda que a exposição à evidência seja constante, ou seja, ainda que o *input* seja bastante frequente. Chomsky (1986a) retomaria o termo, estendendo a sua denominação a “problema de Freud ou Orwell” e aplicando-o ao desenvolvimento do indivíduo no domínio da vida social e das instituições políticas. A discussão acerca das lacunas registadas entre o *input* e *performance* da interlíngua não levou a fundo as considerações teóricas do “problema de Orwell”, até que Franceschina (2005) demonstrou a sua importância enquanto ferramenta heurística para a investigação que se debruça sobre o desenvolvimento da ASL.

<sup>94</sup> Schachter (1974:208) define a esquiva como a estratégia aplicada pelo aprendente para evitar determinadas estruturas que são difíceis para ele, fazendo pressupor que a ausência de certos erros no registo do falante de L2 não reflecte, necessariamente, a sua real competência interlinguística. Kleinman (1978:162) refere que a esquiva pode ter origem em factores como a personalidade, nervosismo, falta de confiança ou inibição, entre outros.

<sup>95</sup> Ellis (1994:351) propõe cinco motivações para ocorrerem fossilizações dos itens, regras e subsistemas linguísticos: transferência linguística (*language transfer*), transferência por treinamento (*transfer of training*), estratégias de aprendizagem da L2 (*strategies of learning*), estratégias de comunicação da L2 (*strategies of second language communication*) e sobregeneralização do material linguístico da LA (*overgeneralization of TL linguistic material*) A conjugação dos cinco processos de fossilização produzirá, então, uma competência interlinguística plenamente fossilizada.

<sup>96</sup> Em trabalho sobre a aquisição do género gramatical em gramáticas fossilizadas, Franceschina (2005), tendo como ponto de sustentação os quadros teóricos dos Princípios e Parâmetros e do Programa Minimalista, debruça-se sobre as causas que originam a divergência no estágio final do desenvolvimento morfossintáctico da L2. Partindo do pressuposto que a GU e o dispositivo computacional da faculdade da linguagem se encontram totalmente disponíveis no processo de ASL (Cook, 1994; White, 1989, 2003; Bruhn de Garavito & White, 2002; Montrul, 2004; Cook & Newson, 2007 [1988]), o conjunto de suposições da autora vai girar em volta tanto do “problema de Platão” (o *input* pobre) como do “problema de Orwell ou Freud” (o *input* rico), abordados por Chomsky (1986a) no estudo sobre o conhecimento da língua. As conclusões do trabalho apontam para a possibilidade de a variação registada por falantes não-nativos não ser tão aleatória como alguma literatura de ASL sugere. Da mesma forma que o “problema de Platão” pode ter influência na questão da variação, também o “problema de Orwell” é determinante para esta, permitindo a articulação entre ambos que se observe a forma como o conhecimento interior interage (ou falha na interacção) com os estímulos externos para determinar o desenvolvimento das gramáticas mentais. Assim, enquanto a aquisição de uma L1 fornece uma perspectiva privilegiada para se observar o primeiro problema, a visão de acordo com o “problema de Orwell” ajuda a analisar, perspicazmente, as questões inerentes à ASL por falantes adultos, dado que se pode entender como uma norma, e não como um desvio ou uma fossilização, a falha em alcançar a competência da LA em determinadas áreas.

<sup>97</sup> Para Selinker (1972:218), a transferência da L1 pode ser positiva ou negativa. A primeira diz respeito às similaridades repartidas por L1 e L2 e que contribuem para facilitar a ASL. Por seu lado, a transferência negativa refere-se às diferenças entre L1 e L2, sendo responsável pelas interferências da primeira na ASL.

<sup>98</sup> Este pressuposto da Teoria da Processabilidade vai ao encontro de postulados avançados previamente em estudos de ASL, que defendem que um determinado traço só poderá ser transferido se houver lugar na estrutura da interlíngua onde o mesmo possa ser encaixado (Andersen, 1983c:182).

<sup>99</sup> Sims (1989:70) atribui às fossilizações temporárias a designação de fossilizações suaves (*soft fossilizations*) ou *jellifications*.

<sup>100</sup> Scovel (1969:248) rotulou a fossilização fonológica de “fenómeno Joseph Conrad” em homenagem ao escritor britânico de origem polaca, cujo nome de baptismo era Józef Teodor Nałęcz Korzeniowski (Berdyczew, 3 de Dezembro de 1857 – Bishopboune, 3 de Agosto de 1924). Educado na Polónia ocupada pela Rússia, Conrad viajou para Marselha, onde, com a idade de 17 anos, iniciou uma carreira de marinheiro. Em 1878 passou a servir num barco britânico, escapando assim ao serviço militar russo e conseguindo a nacionalidade inglesa, em 1884. Aos 21 anos era já fluente em inglês, língua que mais tarde dominaria com excelência, mas mantendo sempre o sotaque polaco.

---

<sup>101</sup> As pesquisas na área da voz e emoção têm demonstrado que alegria, tristeza e raiva constituem o tridente das emoções básicas, podendo ser identificadas perceptualmente em diferentes línguas e culturas, o que aponta para a possibilidade de a expressão emocional e a percepção serem regidas por determinados princípios universais. Contudo, no caso específico do português, os estudos têm evidenciado certas particularidades desta língua, que a distinguem de outras, como o inglês, o alemão ou o finlandês. De facto, no português ocorre uma grande aproximação entre os valores de emoções pertencentes a famílias distintas, no caso a expressão neutra, a tristeza e a alegria, possivelmente por razões de carácter cultural (Nunes, 2009:279). Em resultado de tal, as produções de falantes nativos do português que envolvam estas três emoções poderão não ser devidamente descodificadas por indivíduos em situação de aquisição do PtgL2, principalmente se forem detentores de códigos linguístico-culturais que se afastam bastante dos da comunidade falante da LA. Retoma-se aqui, então, a questão das fossilizações terem também como motivação o grau de aculturação do aprendente em relação ao grupo detentor da LA (*The Acculturation Model*, Schumann, 1978:34).

<sup>102</sup> Para uma discussão detalhada em torno da possibilidade de os crioulos registarem casos de flexão contextual, como a marcação tonal em certas construções com verbos seriais no saramacano (Good, 2003), as formas verbais curtas e longas no crioulo de base francesa maurício para fazer a distinção de verbos transitivos usados com ou sem objecto, a concordância de género em determinadas orações copulativas no crioulo do Haiti (DeGraff, 2001a; Lefebvre, 2003), a distinção entre pronomes sujeito e objecto no crioulo de Tobago (James, 2003) ou a marcação de caso em determinadas estruturas do crioulo caraió de base inglesa, do saramacano (Veenstra, 1994, 1996) e do crioulo do Haiti (Lefebvre, 1998; Sterlin, 1998), consulte-se Plag (2008a:126-128), já que este defende a possibilidade de se estar perante situações meramente aparentes, configurando todas elas casos efectivos de flexão inerente.

<sup>103</sup> A Hipótese da Sequência Não-marcada (*Unmarked Alignment Hypothesis*) postula que, na ASL, os aprendentes organizam inicialmente a sintaxe mapeando os papéis semânticos mais relevantes para o sujeito. Por seu lado, a expressão estrutural do sujeito ocupará a posição linear mais proeminente na estrutura-C, ou seja, a posição inicial (Pienemann, Biase & Kawaguchi, 2005:229).

<sup>104</sup> Para reforçar o postulado de que os crioulos são interlínguas convencionais de um estágio primário de aquisição, Plag (2008b) analisa ainda a questão das interrogativas e a realização de estruturas negativas, concluindo que, em ambos os casos, os padrões encontrados são sustentados em termos da HI. Quanto aos casos que têm sido entendidos como transferências pela literatura convencional, o autor entende que encontram sustento na Hipótese do Desenvolvimento Moderado da Transferência. Para uma leitura mais detalhada acerca destas questões, consulte-se Plag (2008b:320-324).

<sup>105</sup> Adoptando o quadro teórico da Função do Paradigma da Morfologia (*Paradigma Function Morphology*), que advoga que a flexão morfológica se realiza mais do que a lexical (Stump, 2001:14), Kihm (2010) estabelece comparações entre os paradigmas de flexão verbal dos crioulos da Alta Guiné e os crioulos do Golfo da Guiné, e entre os crioulos de base portuguesa da costa ocidental africana e os crioulos indo-portugueses, tendo em conta que a etapa mais evoluída de aquisição alcançada numa interlíngua relativamente estável por parte de aprendentes adultos, ou seja, a Variedade Básica (Klein & Perdue, 1997:303), é a extinção dos paradigmas flexionais e a não-realização correlativa no interior da estrutura das palavras. Kihm (2010:12) conclui que muitos dos traços morfossintácticos não são filtrados porque, simplesmente, categorias como tempo, aspecto ou número são indispensáveis para uma apropriada expressão de eventualidades a que os seres humanos necessitam de recorrer para comunicarem. Assim, alguns paradigmas limitados sobrevivem e os que se extinguem são também substituídos por paradigmas generalizados — e construções perifrásticas —, aspectos estes que contribuem especificamente para a formação dos pidgins e crioulos. Quanto à diversidade de crioulos portugueses que emergiu quer na Índia quer na costa ocidental de África, tal ficar-se-á a dever, segundo o autor, ao contacto entre o português e diferentes substratos e/ou adstratos.

<sup>106</sup> O European Science Foundation Second Language contém um arquivo computadorizado de dados coligidos por grupos de pesquisadores do European Science Foundation Project em cinco países europeus: França, Alemanha, Grã-Bretanha, Holanda e Suécia. O projecto analisa a aquisição espontânea de L2's por parte de quarenta imigrantes adultos que vivem na Europa Ocidental, bem como o código linguístico utilizado na comunicação com os falantes nativos dos seus países de acolhimento.

<sup>107</sup> Mateus & Villalva (2006:27), por exemplo, chamam a atenção para um caso de mudança que poderá estar em curso no PE, afectando a 2ª pessoa do singular do Pretérito Perfeito Simple e concorrendo para a regularização do paradigma da 2ª pessoa do singular. De facto, este é o único tempo verbal cuja 2ª pessoa do singular não termina em /s/. Daí que muitos falantes maternos utilizem, inconscientemente, as formas “*comprastes*”, em vez de “*compraste*”, “*lestes*”, em lugar de “*leste*”, ou “*dormistes*”, substituindo “*dormiste*”.

<sup>108</sup> O fanakalo, um pidgin de base zulu, inglesa e africânder, viu o seu uso extravasar para lá das zonas mineiras da África do Sul, criando pequenas bolsas de falantes na República Democrática do Congo, Namíbia, Zâmbia e Zimbábue. Apesar de usado apenas como L2, o fanakalo é utilizado como língua franca por centenas de milhares de falantes de diversas L1's. O fanakalo é também um dos poucos pidgins que tem como base uma língua não colonizadora e o único pidgin de base zulu. O nome da língua tem como étimo uma truncção de termos de origem ngungi e significa “faz isso assim”, reflectindo o uso funcional deste pidgin. Informações detalhadas sobre o fanakalo podem ser consultadas em Adendorff (2002).

<sup>109</sup> Os movimentos visíveis (*overt movements*), como a deslocação de itens do SN, podem também ocorrer nos crioulos, mesmo que não estejam presentes na sua língua lexificadora:

Exs.: FRANCÊS:

<i>Le</i>	<i>chat</i>	<i>qui</i>	<i>est</i>	<i>assis</i>	<i>sur</i>	<i>ma</i>	<i>table.</i>
DET	gato	REL	está	sentado	LOC	1SG	mesa.

“O gato que está sentado na minha mesa”

CRIOULO DO HAITI:

<i>Chat</i>	<i>ki</i>	<i>chita</i>	<i>nan</i>	<i>tab</i>	<i>mwen</i>	<i>an.</i>	(Kihm, 2008:416)
Gato	REL	senta	LOC	mesa	1SG	DET	

“O gato que está sentado na minha mesa.”

<sup>110</sup> Note-se que Chomsky (1998) conclui que, no fundo, não existe “imperfeição” no deslocamento dos itens, já que esta é a forma de a sintaxe se tornar legível às componentes fonológica e conceptual.

<sup>111</sup> No que concerne à ausência de movimento aparente do verbo para domínios funcionais mais altos nos crioulos, Roberts (1997:12-13) argumenta que, nestes, os verbos nunca se deslocam da sua posição base, enquanto núcleo do SV. Assim, a adjunção do verbo para um módulo mais alto e/ou um núcleo de concordância (“movimento de V-para-I”) encontra-se impossibilitada de acontecer, como sucede nos crioulos de base romance, ainda que o movimento exista na língua lexificadora. Contudo, Kihm (2008:414) entende que o argumento de Roberts não colhe para provar que a posição do verbo serve de pista para demonstrar a não-marcação sintáctica dos crioulos, já que não pode ser aplicado aos crioulos de base lexical francesa, nem mesmo aos CP's, conforme, na linha de Roberts (1997), o haviam postulado DeGraff (1997), para o crioulo do Haiti ou outros crioulos de base francesa e Baptista (2002), para o CCV. Assim sendo, os crioulos também não diferem das outras línguas no que respeita ao pressuposto do “movimento do verbo de V-para-I”.

<sup>112</sup> A argumentação de Roberts (1999) acerca da ausência do “movimento do verbo de V-para-I” nos crioulos de base francesa e sua comparação com idêntica ocorrência na língua inglesa assenta na Teoria dos Traços, que defende que a FLEX contém traços verbais fortes ou fracos, a fim de poder fazer a verificação das propriedades do verbo seleccionado a partir do léxico (Chomsky, 1993:28). Portanto, os traços do infinitivo e dos verbos auxiliares são interpretados na PF, mas como não estão directamente associados às matrizes fonológicas têm de atrair visivelmente o verbo principal para, com ele, produzirem o *spell out*. Contrariamente, os traços fracos, como acontece no inglês, não têm visibilidade fonológica, pelo que não necessitam de atrair visivelmente o verbo auxiliar, a fim de este ser eliminado. Estabelecendo correlação com o poder distintivo dos traços do verbo há ainda a relativa “riqueza” da flexão verbal. Este aspecto determina que os verbos *have* e *be*, em virtude de terem uma flexão mais rica do que os verbos auxiliares, se movimentem abertamente para o nóculo FLEX no inglês moderno. Esta diferença não ocorre no francês e no português, uma vez que os seus verbos possuem flexão rica.

<sup>113</sup> Note-se que o ADV funciona como especificador ou como adjunto, que é um predicado secundário. Desta forma, não é classificado com os traços das quatro categorias lexicais ou temáticas (N, V, A, F), dado que não pertence a uma categoria principal.

<sup>114</sup> A complexidade apresentada pelo posicionamento do ADV na cadeia frásica é também uma realidade no francês. De facto, Kihm (2008:420) considera como parciais as generalizações de Pollock (1989) sobre o francês, já que este entende que o “movimento do verbo de V-para-I” fixa o ADV na posição pós-verbal, daí a agramaticalidade sugerida para a frase “\**Jean souvent embrasse Marie*”. Contudo, Kihm (2008), enquanto falante nativo do francês, não considera agramatical a construção (exemplo [236b]) e outras realizações que possibilitam a colocação pré-verbal do ADV: “*Maria souvent embrassait son père avant de partir pour l'école*” (Kihm, 2008:421).

<sup>115</sup> Kihm (2008:421-422), contrariando o postulado de Roberts (1999), lista também evidências de que o posicionamento do ADV no crioulo de base francesa do Haiti é inconclusivo, principalmente quando ocorre conjuntamente com os marcadores de TMA:



---

Ex.: CRIOULO DO HAITI:

Minis la (deja) ap (deja) l diskou a (deja). (Kihm, 2008:421)  
Ministro DET (já) ASP(já) ler discurso DET (já)  
“O ministro já está a ler o discurso”

<sup>116</sup> *Projecto Vertentes do Português Rural do Estado da Bahia – Fundamentos: O Conceito de Transmissão Lingüística Irregular* [em linha]. Disponível em: <http://www.vertentes.ufba.br/conceito.htm> [Consultado 04 Dez. 2009].

<sup>117</sup> “*This facilitates substantial comparison and, in many respects, avoids ‘reinventing the wheel’*” (Baxter, 2004:105).

<sup>118</sup> O processo de mudança linguística só se completa, por norma, na terceira geração. A primeira geração é caracterizada pelo bilinguismo em estado incipiente, já que o discurso dos falantes revela bastantes traços da L1 e da criação de formas analógicas e originais. Seguidamente, a segunda geração demonstra já indícios de uma mudança parcial e, por fim, a terceira geração evidencia a mudança completa.

<sup>119</sup> Neste caso, norma será também entendida como um dos dialectos da língua, enquanto as variedades regionais e de sistemas linguísticos deterão estatuto idêntico ao da norma-padrão, isto é, não referem formas diferentes ou desprestigiadas de falar a língua, antes reflectem tão só a forma de falar a língua numa determinada região.

<sup>120</sup> *Projecto vertentes do português rural do Estado da Bahia. Chave de transcrição do projeto vertentes.* [em linha]. Disponível em: <http://www.vertentes.ufba.br/chave.htm>.

<sup>121</sup> Para uma completa identificação das convenções ortográficas aplicadas na transcrição dos nossos dados, consulte-se Tagliamonte (2006:56-64).

<sup>122</sup> As comparações com o Português Vernacular de Moçambique têm como base os achados apresentados por Jon-And (2008) no *Congresso Anual da Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBPLE), Universidade de Lisboa, Portugal, 03-05 de Julho, 2008*, e que serão parte integrante da dissertação de doutoramento ainda em processo de elaboração.

<sup>123</sup> As comparações com o Português de Cabo Verde têm como base os achados apresentados por Jon-And (2009) no *Joint Summer Meeting of the Society of Pidgin and Creole Linguistics (SPCL) and the Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBPLE), Universidade de Colónia, Alemanha, 11-15 de Agosto, 2009*, e que serão também parte integrante da dissertação de doutoramento ainda em processo de elaboração.

<sup>124</sup> As faixas etárias registam, entre si, um intervalo temporal de vinte anos, período suficiente para possibilitar um estudo das mudanças ocorridas no tempo aparente, enquanto projecção do tempo real. Os conceitos de tempo real e tempo aparente, bem como a aplicação prática do segundo no estudo dos processos de mudança, encontram-se amplamente descritos em Lucchesi (2000a:277-280) e P. Andrade (2003:80-82).

<sup>125</sup> A *Lex Parsimoniae*, também conhecida como “Navalha de Occam” ou “Navalha de Ockham”, foi primeiramente preconizada pelo frade franciscano inglês William de Ockham, que defendeu que a explicação para qualquer fenómeno científico deve assentar na economia, parcimónia e simplicidade.

<sup>126</sup> “*Linguistic variables which are frequent but have minimal variation are less viable for investigation by this method. Although the structures themselves may be interesting, if the data at your disposal is near categorical (either 100 per cent or 0 per cent), then there is little room for quantitative investigation. (...) In this case, you may rely on the constraint ranking of factors for comparative purposes*” (Tagliamonte, 2006:86).

<sup>127</sup> O valor do *log-likelihood* representa a qualidade da aproximação entre o modelo constituído (os factores que caracterizam os contextos, os pesos que indicam o efeito dos factores sobre o uso da variação investigada, o *input* representativo do nível de uso de determinado valor da variável dependente e o modelo matemático logístico) e os dados submetidos à análise.

---

<sup>128</sup> O substantivo *aldeão*, por exemplo, possui três formas distintas de marcar o PL. Assim, se, por exemplo, tivéssemos que codificar este nome para o grupo de factores *saliência fónica*, seríamos obrigados a optar por três factores distintos, consoante a realização do informante: um para os plurais regulares (*aldeãos*); outro para os plurais –*ão/-ões* (*aldeões*); e outro ainda para os plurais –*ão/-ães* (*aldeães*). Caso ocorrências deste tipo acontecessem em número substancial no nosso *corpus*, interessaria confirmar se as mesmas se comportavam de um modo livre, isto é, se eram aplicadas pelos falantes de modo aleatório, ou, em alternativa, eram utilizadas segundo um padrão ordenado e homogêneo, constituindo, portanto, uma variação instalada na comunidade. Para estudo da segunda situação interessaria, então, tal como preconizado por Labov, 1972a, 1982 e Weinreich, Labov & Herzog (2006 [1968]:100), observar a problemática do encaixamento que estaria subjacente à variação (*embedding problem*), ou seja, determinar quais seriam os factores da estrutura linguística que estariam a motivar a referida variação.

<sup>129</sup> Para uma explicação mais detalhada acerca da determinação da *aplicação* ou *não-aplicação da regra variável* na gramática generativo-transformacional, consulte-se Guy & Zilles (2007:144-145).

<sup>130</sup> No trabalho de estudo sobre o Português do Rio de Janeiro, Scherre (1988) considerou os itens com marcação fónica parcial de PL como elementos com marca de pluralidade (p.e. *mulhere*), enquanto Braga (1977) os havia codificado como singulares.

<sup>131</sup> Nina (1980:103), alegando que todos os casos de PL em –*es* apresentam um comportamento semelhante, funde-os num único factor.

<sup>132</sup> Para o grau de diferenciação fónica na relação singular/plural, Guy (1981a:185) optou por uma escala que estabelece factores distintos para os elementos terminados no singular em –*ão* e –*l*. Relativamente ao item “vez”, como referimos a propósito do critério por nós adoptado para a caracterização dos dados (ponto 3.1.4), o linguista criou também dois factores diferentes: um para a expressão “às vezes”; e outro para as restantes realizações em que intervém o referido item.

<sup>133</sup> À semelhança dos achados de Lucchesi (2000a) no estudo do género, Emmerich (1984), ao analisar o dialecto do Alto Xingu, adquirido por contacto, concluiu que os morfemas flexionais do verbo do português xinguanu tendem a ser incorporados, primeiramente, quando são fonologicamente mais salientes.

<sup>134</sup> Partindo do pressuposto que terá havido uma aquisição distinta do português pelos escravos africanos e, conseqüentemente, pelos seus descendentes, Lopes (2001) acredita que os segundos poderão, de alguma forma, fazer uso de uma linguagem que contenha traços dos seus ancestrais. Como o fenótipo (aparência racial) não se tem revelado suficiente para identificar a ancestralidade do povo brasileiro, marcado por múltiplas misturas, Lopes (2001:155-164), para caracterizar melhor os seus informantes, decidiu constituir uma variável assente nos pressupostos das pesquisas científicas em genética levadas a cabo pela Universidade Federal da Bahia. Esta, para conseguir relacionar a origem étnico-racial com algumas patologias, optou por estudar o significado dos sobrenomes, já que tanto africanos como índios abandonaram os seus sobrenomes de origem para adoptarem os da língua portuguesa. A pesquisa acabou concluindo que os africanos revelavam apetência pela escolha de sobrenomes religiosos, enquanto os índios davam preferência aos sobrenomes portugueses não-religiosos, com conexão aos reinos animal ou vegetal. O número de sobrenomes contabilizados pela UFB provou a coerência do critério e, conseqüentemente, tornou a variável do estudo de Lopes (2001) estatisticamente confiável.

<sup>135</sup> P. Andrade (2003:102) considerou três factores para os elementos que pluralizam com inserção de –*s* final e alteração silábica: itens finalizados em –*l*, itens nasais em –*ão* que pluralizam em –*ões* e itens nasais em –*ão* que pluralizam em –*ães*. Em adição, o factor dos casos regulares foi constituído com elementos que terminam em vogal oral (p.e. *mesa* > *mesas*) ou nasal (p.e. *irmã* > *irmãs*). Estes últimos foram por nós considerados como itens nasais que pluralizam de modo regular (inserção de –*s* final).

<sup>136</sup> O novo grupo de factores *saliência fónica*, após as amalgamações levadas a cabo por P. Andrade (2003:103), passou a incluir os seguintes factores: (i) –*l*, –*r*, –*s*; plurais duplos e nasais sem alteração silábica; (ii) ditongos nasais com alteração; (iii) regulares. A propósito destas amalgamações, note-se que a autora não faz referência a quaisquer testes qui-quadrado, uma vez que só com recurso a estes se poderia confirmar se as amalgamações efectuadas constituíam uma distinção linguística válida para efeitos de análise.

<sup>137</sup> Para análise dos seus dados, Baxter (2004:109) constituiu factores de acordo com as seguintes características: (i) quatro factores considerando a presença ou ausência de tonicidade na sílaba PL, e que acompanham uma mudança no radical da palavra flexionada em número; (ii) um factor referente aos monossílabos átonos; (iii) um factor englobando plurais cujos singulares terminam em –*s*.

---

<sup>138</sup> Nos nossos dados verificam-se realizações com o nome “avós”, um caso de metafonia em que ocorre marca de feminino na realização da pluralização do masculino “avô” (p.e. *mia mãe e mia avó sempre dizia que meus avós...* [CARMOM1]). Este é também o único item metafônico cujo singular é oxítono, já que todos os outros possuem a tonicidade na penúltima sílaba, configurando ocorrências paroxítonas.

<sup>139</sup> Guy (1981a:292-293:) exemplifica os fenómenos de manutenção e supressão no português com recurso a alguns fenómenos deste idioma, como a queda intervocálica da consoante [ʒ] depois de uma vogal átona (p.e. *pejõrem > pior*), ou a sua manutenção depois de uma vogal tónica (p.e. *cũjum > cujo*). Para uma observação detalhada dos fenómenos de apócope, síncope e aférese de sílabas átonas do cabo-verdiano, sugere-se a consulta de Almada (1961).

<sup>140</sup> Realizações configurando o fenómeno apontado por Scherre (1988) e Lopes (2001) são também recorrentes no PA (p.e. *eu vim praqui ê tive Z'ONZE anos* [ANTOM1]).

<sup>141</sup> Braga, (1977:30-31), no seu estudo sobre a concordância PL no SN no Triângulo Mineiro, codifica já esta variável, apresentando um grupo binário de factores: *presença de marca formal plural no segmento precedente ao elemento analisado* e *ausência de marca formal plural no segmento precedente ao elemento analisado*. Braga (1977:39) conclui que o segundo factor favorece a marcação no item seguinte.

<sup>142</sup> Nas suas observações, Poplack (1980a) levou em conta a relação entre *posição linear* e *posição dos marcadores plurais precedentes*, mas avançou que, para um melhor entendimento acerca da CPL-var do EPR, a variável independente *marcas precedentes* deveria ser também analisada em parceria com a *classe gramatical*, o *segmento fonológico seguinte* e a *saliência fónica*.

<sup>143</sup> Lucchesi (2000a), nas pesquisas efectuadas sobre a variação da concordância no género de HEL-Ba, concluiu que a marcação mórfica predomina sobre a economia funcional da língua na aquisição/fixação das regras de concordância, contrariando-se, assim, o princípio de que o processo mental de associação da língua em agrupar formas semelhantes – o *Processo Paralelo* – conduz a uma economia de língua em que “*marcas levam a marcas e zeros levam a zeros*”. Assim, não será a marca no elemento anterior da cadeia linear que leva à marcação, mas sim a marcação mais explícita de género no núcleo que favorece a marcação do género na cadeia linear. Consequentemente, quanto mais explícitas forem as marcas morfológicas presentes na posição de núcleo, mais facilmente se dará a aquisição da regra de concordância nominal no interior do SN. Paralelamente, estes resultados opõem-se às teorias funcionalistas de Martinet (1955), que determinam que os determinantes e modificadores têm maior valor funcional na definição da marcação do género, visto serem os únicos que promovem a informação referencial nos casos, por exemplo, dos nomes comuns de dois géneros. Como tal, Lucchesi (2000a) conclui que a informação se perde durante a conversação, caso não seja recuperada pelos elementos contextuais ou pragmáticos.

<sup>144</sup> Guy (1981a:146) ancora a explicação do fenómeno em estudos sobre percepção da fala, os quais avançam que as fricativas não-labiais surdas favorecem a permanência dos traços não-sonoros, uma vez que são mais distintas e mais perceptíveis do que as sonoras correspondentes.

<sup>145</sup> A constituição de dois factores para a pausa (pausa interna e pausa final), “*teve como objetivo verificar possíveis diferenças neste tipo de condicionamento de comportamento considerado arbitrário*” (Scherre, 1988:247). Assim, a pausa interna é configurada pela distância de realização fónica que trava a interferência entre sons na elocução, isto é, entre o item analisado e o elemento que se encontra em posição posterior a ele. Por seu lado, a pausa final é bastante mais longa e marca o fim de uma sentença ou período.

<sup>146</sup> De acordo com Bechara (1999 [1928]:343), a *haplogogia* ou *braquilogia* é um fenómeno comum que visa “*evitar reduplicações de sílabas que pertencem à palavra primitiva e ao sufixo*”, como em: *caridad(e) + oso > caridoso* (por *caridadoso*) ou *bondad(e) + oso > bondoso* (por *bondadoso*). Este tipo de simplificação pode ainda ocorrer em palavras formadas por composição: *trágico + cómico > tragicómico*.

<sup>147</sup> Os valores finais de Scherre (1988:249) para a pausa dizem respeito a apenas um factor, já que a linguista acabou amalgamando as pausas interna e final num único factor.

<sup>148</sup> A propósito dos seus achados, Scherre (1988:256) chama a atenção para o facto de o morfema marca de PL –s apresentar um comportamento totalmente diferente do –s não-morfémico do português, cujo cancelamento não é afectado pela pausa, ficando antes a dever-se aos itens surdos e às consoantes. Como tal, as regras que condicionam a inserção de ambos os –s terão de ser explicadas com base em pressupostos distintos.

<sup>149</sup> Igual critério foi adoptado em relação aos artigos, em que nos abstivemos de os classificar em determinantes artigos definidos ou determinantes artigos indefinidos.

---

<sup>150</sup> Lopes (2001:83) classificou como artigos indefinidos todas as formas “um” e “uns” em que não era nítida a distinção entre pronome ou artigo indefinido.

<sup>151</sup> Scherre (1988:150) tratou estes itens como pronome pessoal de 3a. pessoa, ao passo que Lopes (2001:85) os considerou como outras formas de indicar posse, incluindo-os, tal como nós, na categoria dos possessivos.

<sup>152</sup> Lopes (2001:85) e P. Andrade (2003:97) apenas consideraram como quantificador os indefinidos invariável *tudo* e variável *todo* (e suas flexões).

<sup>153</sup> Factores constituídos por Baxter (2004:115-117) para análise geracional da influência da variável *posição linear x posição em relação ao núcleo do SN* na marcação PL do PT: (i) item PL à esquerda, adjacente ao núcleo; (ii) item PL à esquerda, não-adjacente ao núcleo; (iii) núcleo em 1ª. posição; (iv) núcleo em 2ª. posição; (v) núcleo em 3ª. posição; (vi) núcleo em 4ª. posição; (vii) núcleo em 5ª. posição; (viii) item PL à direita do núcleo, 2ª. posição no SN; (ix) item PL à direita do núcleo, 3ª. posição no SN; (x) item PL à direita do núcleo, 4ª. posição no SN; (xi) item PL à direita do núcleo, 5ª. posição no SN.

<sup>154</sup> Factores elaborados por Jon-And (2008) para análise da influência da variável *posição linear x posição em relação ao núcleo do SN* na marcação PL do PMp: (i) elemento imediatamente anterior ao núcleo; (ii) elemento anterior ao núcleo, mas não imediatamente; (iii) núcleo em 1ª. posição; (iv) núcleo em 2ª. posição; (v) núcleo em 3ª. posição; (vi) núcleo nas demais posições; (vii) elemento posterior ao núcleo, 2ª. posição no SN; (viii) elemento posterior ao núcleo, 3ª. posição no SN; (ix) elemento posterior ao núcleo, demais posições.

<sup>155</sup> O falante refere-se à antiga unidade monetária portuguesa correspondente a mil escudos, isto é, ao conto de réis.

<sup>156</sup> Labov (1981) considera que o lapso de tempo ideal para se detectar uma mudança linguística em curso será o que corresponde, no mínimo, a metade de uma geração e, no máximo, a duas gerações, isto é, que contempla um intervalo temporal configurado entre 12 e 50 anos.

<sup>157</sup> Face ao exposto, resulta que a variável idade, por exemplo, pode ser tratada sob duas perspectivas: i) a da variável contínua; II) a da variável *continuum* de idades, segmentada em grupos etários distintos (a variável pela qual optámos no presente trabalho).

<sup>158</sup> Lembre-se que o grupo de factores tido como verdadeiramente fiável é o do tipo binário, que apenas permite duas alternativas. Como tal, a constituição de variáveis independentes com elevado número de factores viola este pressuposto e poderá apresentar problemas analíticos sérios devido à provável distribuição desequilibrada de dados, os quais irão ferir a ortogonalidade do grupo de factores em que estão inseridos.

<sup>159</sup> A partilha de factores idênticos por dois ou mais grupos de factores vai determinar que “*tudo o que é codificado como x também é codificado como y. Nesse caso é impossível para o programa testar como os dados do contexto x se comportam quando estão na presença do factor y, e vice-versa*” (Guy & Zilles, 2007:199).

<sup>160</sup> No ponto 3.6.7. fizemos alusão aos critérios que levaram à constituição da variável independente *marcação de género* e seus factores. Quanto ao número de ocorrências produzido para cada factor constituído, bem como os percentuais de marcação PL motivados pelos mesmos, poderão ser consultados na Tabela 12.8. Como se pode constatar, os percentuais de inserção de pluralização rondam todos o patamar da neutralidade, daí o grupo de factores *marcação de género* ter sido considerado irrelevante pela ferramenta VARBRUL, enquanto variável que intervém na configuração da CPL-var do SN do PA. Deste modo, relativamente à hipótese 8 que constituímos, verifica-se que a marcação de género não constitui condição *sine qua non* para conduzir à marcação de número, quando esta é exigida em itens flexionados em género, isto é, que a aplicação de uma primeira flexão, a do género, seja ela regular ou irregular, não implica necessariamente a aplicação de uma flexão posterior, a do número.

