

NÚMERO E GÉNERO NOMINAIS NO DESENVOLVIMENTO DO PORTUGUÊS DE TIMOR-LESTE

NOMINAL GENDER AND NUMBER IN THE DEVELOPMENT OF THE TIMORESE VARIETY OF PORTUGUESE

Isabel A. Santos*

imas@fl.uc.pt

Cristina Martins*

crismar@fl.uc.pt

Isabel Pereira*

mipp@fl.uc.pt

O presente estudo contribui para a descrição do português de Timor-Leste (PTL), analisando padrões variáveis de concordância nominal em número e em género registados nessa variedade.

A relevância desta investigação decorre do facto de o PTL constituir uma variedade nacional não nativa (VNN) do português, ainda escassamente descrita. O facto de se tratar de uma VNN emergente torna o seu estudo relevante para a compreensão do processo de formação histórica das restantes VNN do português.

As VNN resultam de um processo de aquisição, como língua não materna (LNM), de um idioma que, num determinado território, assume o estatuto de língua oficial, isto é, de língua segunda (LS). O confronto entre as produções deste tipo de falantes e as de aprendentes de língua estrangeira (LE) permite realçar padrões de comportamento comuns e/ou específicos.

Procede-se, aqui, ao confronto de produções escritas de falantes de PTL e de aprendentes de PLE. Os resultados evidenciaram tendências similares, mas também uma maior expressão da ausência de concordância nominal plena no grupo dos falantes do PTL. Na sua globalidade, os dados sugerem que, como noutras VNN, a concordância nominal variável poderá vir a constituir-se como uma característica do PTL, que o diferenciará do português europeu (PE).

Palavras-chave: Concordância nominal. Género gramatical. Português de Timor-Leste (PTL). Variedade não nativa (VNN). Português língua estrangeira (PLE).

* Centros de Estudos de Linguística Geral e Aplicada (CELGA-ILTEC) - Faculdade de Letras da Universidade de Coimbra, Portugal.

This study contributes to the description of East-Timorese Portuguese (ETP), focusing on the variable patterns of nominal agreement in number and gender operating in this variety.

The relevance of the research hinges on the fact that ETP is an understudied non-native variety (NNV) of Portuguese. Given its emergent state, the study of this particular variety can furthermore shed light on the historical process that led to the formation of other NNV.

NNV are a product of the non-native acquisition of a language that, in a given territory, takes on official status, this is to say, is a second language (SL). Comparing production data by NNV speakers and by foreign language (FL) learners can elucidate both common and specific patterns of behavior. In this study, texts written by ETP speakers and by PFL learners were compared. Results revealed similar trends in both samples, but also a greater preference of ETP speakers for not complying to full nominal agreement. In general, data suggest that variable patterns of nominal agreement are likely to emerge as a defining property of ETP, as is currently the case in other NNV of Portuguese, thus diverging from European Portuguese (EP).

Keywords: Nominal agreement. Grammatical gender. East-Timorese Portuguese (ETP). non-native variety (NNV). Portuguese as a foreign language (PFL).



1. Introdução

Presente em território timorense há pelo menos quatro séculos, o português é, contudo, falado, ainda hoje, por uma minoria dos habitantes de Timor-Leste. De entre os poucos timorenses que fazem, atualmente, um uso produtivo do português¹, a esmagadora maioria adquiriu-o como língua não materna (LNM) e sobretudo em contexto educacional.²

1 Segundo Albuquerque (2011, p. 68, n.3), por exemplo, “apenas 37% da população leste-timorense é fluente em português”. Contudo, o mesmo autor (2011, p. 68, n.3) esclarece que há dados contraditórios sobre a dimensão do uso do português em Timor-Leste: “Os dados citados foram extraídos de *Timor-Leste Census of Population and Housing (2006)* (...) [, mas] o relatório de Desenvolvimento Humano de Timor-Leste (2001) afirma que somente cerca de 5% da população timorense é falante da língua portuguesa”. Já os dados de *Timor-Leste Population and Housing Census 2015* registam apenas 1384 falantes nativos de português, num universo de 1179654 indivíduos (valor equivalente a 0,11% da população) (cf. *Timor-Leste Population and Housing Census: Population by mother tongue and age, Timor-Leste, 2015*, disponível em <http://www.statistics.gov.tl/pt/category/publications/census-publications/>).

2 “[P]ode-se afirmar que a língua portuguesa em Timor-Leste está limitada somente a poucas situações formais e é utilizada apenas por uma minoria sociolinguística: a população mais

Assumindo, a par do tétum, desde 2002, o estatuto de língua cooficial em Timor-Leste (ocupando, por isso, algum espaço nos domínios públicos de uso linguístico, especialmente na forma escrita), e tendo sido, igualmente, a língua dos movimentos de resistência à anexação indonésia a partir de 1974, o português tende a ser positivamente considerado pelos timorenses num plano simbólico (Batoréo & Casadinho 2012). Ainda assim, tal valoração positiva da língua não tem impelido os timorenses para um uso regular do português no quotidiano. As situações de ininteligibilidade mútua decorrentes do plurilinguismo que caracteriza a sociedade timorense são, portanto, ultrapassadas pelo recurso generalizado ao tétum, língua franca do país, e não ao português, que não adquiriu, aqui, e contrariamente ao que aconteceu noutros contextos ex-coloniais, essa função comunicativa.

Nestas circunstâncias, e tomando por referência o modelo de círculos/anéis concêntricos proposto por Kachru (1985) para representar os diferentes resultados da expansão internacional do inglês³, poder-se-á situar, analogamente, o português de Timor-Leste (PTL) num limite do círculo externo do português internacional (ou pluricêntrico, *cf.* Baxter 1992), raiando com o círculo em expansão desta língua. Vejamos: diferentemente do círculo interno, integrado pelas comunidades linguísticas nas quais o idioma goza de estatuto oficial e é predominantemente adquirido como língua materna (LM) (concentradas, no caso do português, em Portugal e no Brasil), o círculo externo (ou estendido) de uma língua com dimensão internacional é constituído pelas comunidades nas quais ela também assume estatuto oficial, em função de processos históricos de institucionalização (como a colonização)⁴, mas em que é adquirida, maioritariamente, como língua não materna (e, mais espe-

idosa, pelo fato de terem nascido antes da invasão indonésia em 1974 e, conseqüentemente, terem frequentado o sistema educacional português. Ao mesmo tempo, a língua portuguesa é também utilizada pelos exilados políticos, que usaram a língua portuguesa como instrumento de resistência à invasão e por uma elite que tem, ou teve acesso à educação portuguesa.” (Albuquerque 2011, pp. 68-69).

3 “The spread of English may be viewed in terms of three concentric circles representing the types of spread, the patterns of acquisition and the functional domains in which English is used across cultures and languages. I have tentatively labelled these: the *inner* circle, the *outer* circle (or *extending* circle), and the *expanding* circle.” (Kachru 1985, p. 12).

4 “The outer (or extended) circle needs a historical explanation: it involves the earlier phases of the spread of English and its institutionalization in non-native contexts. (...) The political histories of the regions where institutionalized varieties are used have many shared characteristics: these regions have gone through extended periods of colonization, essentially by the users of the inner circle varieties. The linguistic and cultural effects of such colonization are now a part

cificamente, como língua segunda - LS⁵). Já o último círculo contemplado neste modelo, o círculo em expansão, recobre as situações em que, não assumindo estatuto oficial na comunidade em que se encontram os aprendentes, a língua é, ainda assim, e fruto do seu prestígio e/ou da percepção do seu valor funcional, aprendida como língua estrangeira (LE).

Apesar das críticas de que tem sido alvo⁶, o modelo de Kachru (1985) tem a vantagem de considerar, na classificação das distintas situações em que se encontram as línguas de dimensão internacional, não apenas variáveis históricas e sociolinguísticas, mas também as condições em que são tipicamente adquiridas pelos falantes em cada caso, explicando, deste modo multifatorial, a heterogeneidade funcional e a variação linguística resultante.

No presente estudo, interessa-nos atender à relação entre as situações de uso e aquisição/aprendizagem da língua portuguesa albergáveis no círculo externo e no círculo em expansão deste idioma, uma vez que, em ambos os casos, o português não é, prototipicamente, adquirido/aprendido como LM. Decorre daí que, quer num quer noutro, o processo aquisitivo do aprendente individual resulta em conhecimento de natureza interlinguística, caracteristicamente instável e altamente variável (Selinker 1972), e que exhibe, nos distintos estádios de desenvolvimento, graus de convergência muito variáveis com as gramáticas dos modelos linguísticos emanados do círculo interno.⁷

of their histories, and these effects, both good and bad, cannot be wished away.” (Kachru 1985, p. 12).

5 A expressão “língua segunda” (LS) é aqui usada na aceção apresentada por Leiria (2004, p. 1), *i.e.*, “uma língua não-nativa [aprendida e usada] dentro de fronteiras territoriais em que ela tem uma função reconhecida (...) [,] indispensável para a participação na vida política e económica do Estado, (...) [sendo] a língua, ou uma das línguas, da escola. Por ser língua do país, disponibiliza geralmente bastante *input* e, por isso, pode ser aprendida sem recurso à escola”. Neste sentido, a LS distingue-se da língua estrangeira (LE), um termo que aqui se reserva “para classificar a aprendizagem e o uso [da língua não materna, ou não nativa] em espaços onde essa língua não tem qualquer estatuto sociopolítico”. Embora criticável sob vários pontos de vista (igualmente discutidos por Leiria 2004), a distinção entre LS e LE, nestes termos, é útil e operacional no quadro dos objetivos do presente trabalho.

6 *Cf.*, para uma revisão e discussão dessas críticas, Schmitz (2014).

7 Central, no modelo de Kachru, é a distinção entre “norm-providing varieties”, correspondendo às que se afirmam no círculo interno, “norm-developing varieties”, identificáveis, em estádios de desenvolvimento distintos, no círculo externo, e, por fim, “norm-dependent varieties” (Kachru 1985, pp. 16-17), respeitantes às interlínguas dos aprendentes da língua como LE que, tomando as variedades do círculo interno como referência, se situam no círculo em expansão.

Para além das similitudes, interessa-nos igualmente destacar as especificidades das circunstâncias em que se processa a aquisição da LNM num e noutro tipo de cenário. As mais relevantes, na ótica dos objetivos do presente trabalho, são (i) as oportunidades de exposição ao *input* e de uso produtivo da LNM (associadas ao grau de concentração dos aprendentes numa dada comunidade), (ii) a probabilidade de transmissão geracional das características interlinguísticas não nativas presentes no *output* que esses aprendentes produzem e (iii) a qualidade desse mesmo *input* (podendo este ser predominantemente marcado por propriedades não nativas ou representar, pelo contrário, modelos nativos).

Importa realçar, antes de mais, que, enquanto o conhecimento interlinguístico dos aprendentes de LNM com estatuto de LS (concentrados em comunidades integradas no círculo externo) está sujeito a efeitos comparáveis aos que resultam da transmissão linguística irregular⁸, o conhecimento interlinguístico dos que aprendem a LNM com o estatuto de LE (em situação exógena, no círculo em expansão) não se encontra nas mesmas condições. Tal sucede, em primeiro lugar, em função da presença, nem que apenas nos domínios públicos, da LNM na comunidade em que se inserem os aprendentes do círculo externo, o que lhes proporciona maiores oportunidades de exposição ao *input* dessa língua, criando, simultaneamente, mais oportunidades para o seu uso produtivo. A concentração, em comunidades com uma base territorial, de aprendentes da LNM nestas condições reforça, assim, a exposição de todos ao *input* não nativo. Nesta linha, e em segundo lugar, nestas comunidades do círculo externo, a LNM, tendo o estatuto de LS, pode vir a assumir um prestígio tal que passe também a marcar presença nos domínios de uso privados, nomeadamente nas interações familiares e domésticas, o que cria condições para a sua transmissão geracional e, logo assim, para a progressiva nativização de características não convergentes com as da gramática dos modelos emanados do círculo interno. Assim se geram variedades não nativas (VNN), reconhecidamente típicas de contextos ex-coloniais (Gonçalves 2010a).

8 Segundo Lucchesi e Baxter (2009, p. 101), a transmissão linguística irregular corresponde à nativização, por crianças em fase aquisitiva, de “versões de segunda língua desenvolvidas entre os falantes adultos (...), que apresentam lacunas e reanálises em relação aos seus mecanismos gramaticais”. Este processo é característico de contextos ex-coloniais, nos quais uma língua europeia (língua de superstrato) “se impõe, de modo que os falantes [nativos] das outras línguas, em sua maioria adultos, são forçados a adquiri-la em condições bastante adversas de aprendizado, em função de sua sujeição e marginalização”.

Por contraste, e como já se afirmou, os aprendentes de uma LNM pertencentes ao círculo em expansão não a transmitem, prototipicamente, aos seus descendentes. Na ausência de um estatuto sociopolítico reconhecido nas comunidades em que se inserem estes aprendentes, as funcionalidades comunicativas da LNM, dado o seu estatuto de LE, são muito limitadas, circunscrevendo-se, normalmente, às interações com estrangeiros (falantes dessa mesma língua) e ao domínio académico. Nestas circunstâncias, para o aprendente da LE em situação exógena, as oportunidades de *output* são, por regra, exíguas, assim o sendo, igualmente, as oportunidades de *input*. Ainda assim, são os modelos linguísticos emanados do círculo interno aqueles que, no contexto instrucional, no qual tipicamente ocorre o processo de aquisição/aprendizagem da LE, servem de referência aos aprendentes da LNM do círculo em expansão.⁹

Algumas das condições que acabámos de descrever sofrem, contudo, alterações relevantes, de natureza quantitativa e qualitativa, quando os aprendentes da LE experienciam uma situação de imersão linguística e, mais especificamente, quando essa experiência ocorre numa comunidade do círculo interno. Nesta, a língua alvo de aquisição/aprendizagem (LA) não é apenas dominante, como é, igualmente, a LM da generalidade dos seus membros. Numa situação de imersão com estas características, alargam-se, portanto, os domínios de uso da LA e aumentam, também assim, as oportunidades de *input* e de *output* para os aprendentes da LNM. Para além disso, e muito pertinentemente, atendendo aos objetivos da presente investigação, os aprendentes da LE em situação de imersão passam a estar expostos, de modo sistemático e intenso, a *input* nativo, *i.e.*, aos modelos linguísticos emanados do círculo interno.

Face ao que ficou dito, o português de Timor-Leste pode ser considerado uma VNN do português emergente, em fase incipiente de consolidação, já que, para além do baixo número de falantes reais, não evidencia, atualmente, um processo de nativização em curso. Neste aspeto, a situação do PTL difere da que caracteriza outros contextos de génese de variedades não nativas do português, como o moçambicano, onde não apenas se observa uma difusão significativa do uso efetivo do português (sobretudo entre a população dos meios urbanos), como já se registam contingentes de falantes do português como LM, promoto-

9 Cf. nota 7.

res da nativização das propriedades linguísticas não convergentes com a gramática do português europeu (Gonçalves 2013).

Igualmente diferenciador do que acontece com algumas outras VNN do português é o estado claramente incipiente e lacunar em que se encontra a descrição do PTL. Contribuir para o conhecimento do seu estado atual é, pois, o primeiro objetivo do presente estudo. A este associam-se outros, nomeadamente o de avaliar as conexões, largamente inexploradas na literatura, entre padrões de aquisição de estruturas específicas do português em situações de exposição ao *input* muito distintas, confrontando comportamentos de aprendentes de PLE, em contexto de instrução formal e em imersão em Portugal, com os comportamentos dos aprendentes de PLS no contexto timorense.

As estruturas escolhidas para avaliação neste estudo são as de concordância nominal em número e em género, reconhecidamente sujeitas a variação no português global e pluricêntrico, não apenas nos círculos externo e em expansão, mas também no círculo interno. Na verdade, está amplamente documentada a presença de regras variáveis (Labov 2003) de concordância nominal em número no português do Brasil (PB) (Brandão & Vieira 2012; Vieira & Brandão 2014). Já no português europeu (PE) a concordância nominal plena em número se manifesta como regra semicatórica, com valores de ocorrência, em dados relativos a três comunidades (Vieira & Brandão 2014), muito próximos dos 100% que caracterizam uma regra categórica.¹⁰ Dito de outro modo, a concordância nominal plena em número (e também em género) é uma característica muito robusta do PE, que distingue esta variedade nativa quer do PB, quer das variedades nacionais com o estatuto de VNN.

A concordância sintática, quer verbal, quer nominal, tem-se revelado um mecanismo gramatical desafiante para os aprendentes em diferentes cenários de aquisição/aprendizagem do português como LNM.

No caso específico das categorias sujeitas à concordância selecionadas para análise na presente investigação, *i.e.*, género e número nominais, importa, desde já, destacar algumas propriedades diferenciadoras. Assim, em português:

10 Labov (2003) prevê a existência, nas comunidades de uso, de regras linguísticas categóricas (que se manifestam com uma frequência de 100%), semicatóricas (cuja frequência é de 95% a 99%) e variáveis (com uma frequência que oscila entre os 5% e os 95%). Se as regras linguísticas categóricas nunca são violadas e as semicatóricas apenas raramente o são nos usos dos falantes, as regras variáveis, por definição, não são violáveis, antes representando a existência de normas concorrentes e conviventes na comunidade em questão.

(i) os valores de número (singular; plural) não são apenas de natureza gramatical, carreando, concomitantemente, valores semânticos (representando informação de natureza quantificacional ou marcando a genericidade/especificidade das expressões nominais (Peres 2013)); já os valores de género nominal (masculino; feminino) são, apenas e tipicamente, de natureza gramatical, sendo minoritários, em português, os nomes cujo valor de género se correlaciona com um valor semântico, correspondendo ao sexo do referente nominal;

(ii) na morfologia do nome, os valores de número expressam-se flexionalmente, com sistematicidade e representam casos de elevada transparência forma-função (*i.e.*, o morfema flexional –Ø representa consistentemente o valor de singular e o morfema –s o de plural); já o género poderá ser entendido como uma propriedade lexicossintática¹¹ ou, então, como uma categoria morfossintática não realizada por flexão (Villalva 2000; 2003), cujo valor não é transparentemente representado na forma morfológica do nome.¹²

A investigação revista na Secção 2 do presente trabalho revela, consistentemente, que existe assimetria na aquisição/aprendizagem das categorias de número e de género nominais do português como LNM, muito embora os estudos não convirjam na identificação da categoria que é mais desafiante para os sujeitos observados. Para esta circunstância deverão concorrer vários fatores (os perfis dos sujeitos estudados, a natureza dos dados empíricos considerados e os procedimentos de análise adotados), cuja valorização resulta, por um lado, de opções metodológicas distintas e, por outro, de perspetivas teóricas não coincidentes. Fornecendo aos investigadores ferramentas conceptuais nem sempre compatíveis entre si, os diferentes posicionamentos condicionam, como se esperará, a interpretação dos dados.

No momento histórico atual, a comunidade científica não dispõe de uma teoria unificada e consensualizada que dê satisfatoriamente conta quer dos muitos resultados não convergentes já apurados e relatados na literatura, quer das múltiplas dimensões necessariamente implicadas no processo de aquisição/aprendizagem de uma LNM.¹³ Sendo certo que

11 “De acordo com esta visão, partilhada por linguistas em relação a várias línguas (*cf.* Eichler *et al.*, 2012), number has been considered a morphological marker that combines with the stem it modifies (Ritter, 1988), whereas grammatical gender has been considered a fixed property of stem (Harris 1991)” (Barber e Carreiras 2005, p. 137 *apud* Martins 2015, n.9).

12 Para uma revisão destes argumentos, *cf.* Martins (2015, pp. 29-31).

13 Uma demonstração ilustrativa disto mesmo é constituída pelo conjunto dos capítulos, representativos de correntes de pensamento diferentes, incluídos em Van Patten e J. Williams (Eds.) (2008).

nenhuma das perspectivas teóricas atualmente concorrentes nega **um** papel ao *input* linguístico neste processo, a verdade é que o peso que o fator *input* assume em cada uma delas difere muito substancialmente. De modo semelhante, poder-se-á afirmar que poucos serão os investigadores que negam a participação, neste processo, do conhecimento linguístico prévio do aprendente de uma LNM (e, em particular, da sua L1/LM), mantendo-se, no entanto, ainda controversa a natureza dessa participação. Finalmente, e se todos concordarão que o aprendente convoca, no processo, mecanismos cognitivos adequados à tarefa complexa de aquisição/aprendizagem, está, contudo, por consensualizar se esses mecanismos são de âmbito geral ou se (com graus igualmente divergentes de pré-especificação inata de categorias linguísticas) estão seletivamente dedicados ao tratamento de representações e dados linguísticos.

No presente trabalho, e na linha do que tem vindo a ser descrito em vários estudos que têm comparado padrões de aquisição linguística por aprendentes que se encontram em distintas situações de contacto de línguas¹⁴ admite-se que não apenas o grau de exposição ao *input*, mas sobretudo a qualidade desse mesmo *input* são fatores com impacto relevante na configuração das interlínguas resultantes de processos de aquisição/aprendizagem de LNM. À luz deste pressuposto, prevê-se, então, que uma exposição prolongada e intensiva a *input* nativo, marcado pela presença de regras categóricas ou semicategóricas, reúne mais condições de redundar numa interlíngua mais convergente com a gramática da línguavalvo (LA), caracterizada por padrões de concordância sintática plena. Do mesmo modo, uma exposição prolongada e intensiva a *input* não nativo, marcado pela presença de regras variáveis, promoverá a aquisição desses mesmos padrões de variação.

Neste trabalho, apresentamos dados empíricos que permitem comparar padrões de concordância nominal em número e em género em grupos de aprendentes de PLNM de distintos perfis e revisitaremos, na secção do estado da arte, os resultados de estudos relativos aos padrões de concordância nominal observados tanto noutras VNN do português, como no português do Brasil. No estudo empírico, compararemos, em concreto, os dados relativos a aprendentes de PLE em contexto de imersão, em Portugal, com experiência instrucional e cujas línguas maternas são o espanhol, o alemão e o chinês (anteriormente apresentados por

14 Para uma revisão de alguns exemplos ilustrativos de estudos deste tipo, cf. Almeida e Flores (2017).

Martins 2015), com os relativos a aprendentes de PTL, com instrução de nível universitário, residentes em Timor-Leste. Procurar-se-á avaliar, assim, até que ponto os dados empíricos recolhidos junto de falantes timorenses de português apresentam especificidades ou, pelo contrário, revelam padrões coincidentes não só com outras VNN do português em mais adiantado estado de desenvolvimento histórico (e, assim também, de “nativização”), mas também com o português aprendido enquanto língua estrangeira (PLE), fundamentalmente por via instrucional, por aprendentes com diferentes níveis de proficiência.

2. Estado da arte

Estudos anteriores têm consistentemente revelado que a concordância nominal em número e em género representa uma área especialmente sujeita à formação de padrões variáveis. Comportamentos não convergentes com a estrutura alvo têm sido observados quer junto de falantes de variedades não nativas do português em contextos em que ele assume o estatuto sociopolítico de LS, quer junto de aprendentes de PLE.

No entanto, poucos destes estudos permitem uma comparação efetiva, na mesma amostra de sujeitos, dos padrões de concordância nominal em número e dos padrões de atribuição de género ao nome e respetiva concordância (Figueiredo 2009; Godinho 2010; Leiria 2006; Martins 2015). Por essa razão, este trabalho tem como objetivo observar o desempenho de um grupo de aprendentes no que diz respeito ao domínio de ambas as estruturas, assumindo que género e número configuram categorias gramaticais do nome com motivações semânticas e comportamentos formais distintos. Dá-se, desta forma, continuidade à investigação empírica de Martins (2015), em que os padrões variáveis de concordância nominal em número e em género são cotejados. Esse estudo envolveu uma amostra de produções escritas de 90 sujeitos de 3 LM distintas (chinês, alemão e espanhol), distribuídos por diferentes níveis de proficiência (A2, B1, B2 e C1), cujo comportamento será agora confrontado com os resultados obtidos junto de aprendentes timorenses de PLS.

Os resultados obtidos no âmbito de algumas investigações visando línguas diversas levam à conclusão de que são maiores as dificuldades sentidas no processamento da atribuição do género nominal (AGN) e da concordância nominal em género (CGN) do que no domínio da con-

cordância nominal em número (CNN). No que respeita ao português, essa interpretação é sustentada pelos dados apresentados tanto em estudos sobre PLE (Leiria 2006; Martins 2015), como em estudos sobre VNN (Gonçalves 1997; Gonçalves 2010b, sobre o português de Moçambique). No entanto, outros trabalhos apresentam resultados distintos, permitindo antes concluir que é na concordância envolvendo a categoria gramatical de número que se regista maior variação (Figueiredo 2009, sobre dados do português de Almojarife, São Tomé¹⁵; Godinho 2010, sobre dados de PLE).

Por outro lado, há maior convergência nos dados sobre a forma como se manifestam os desvios de concordância nominal entre diferentes grupos de aprendentes de PLNM.

No que respeita ao género, diversas investigações têm revelado uma forte tendência para a adoção de uma estratégia de *masculino por defeito*, quer em PLE (Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Martins 2015), quer em VNN (Cavele 1999; Chavagne 2005, sobre o português de Angola; Figueiredo 2009, sobre o português de Almojarife, São Tomé; Gonçalves 1997; Gonçalves 2010b, sobre o português de Moçambique). Atente-se no que, sobre o português de Moçambique, escreve Gonçalves (1997, p. 60): “os ‘erros’ de concordância em género afectam particularmente a classe dos adjectivos, e a tendência dominante é para o uso de formas masculinas em contextos que requerem a sua flexão no feminino”. Esta tendência é também referida por Mariotto e Lourenço-Gomes (2013, p. 1282) relativamente a variedades interlinguísticas de aprendentes de PLE de línguas maternas inglesa e espanhola: “os aprendizes priorizam uma marcação *default* da forma masculina”.

Os estudos apontam para uma maior ocorrência de desvios de AGN/CGN em nomes cuja estrutura formal e/ou valor semântico não constituem indícios do valor de género (Ferreira 2011; Martins 2015; Pinto 2015), sendo, em alguns deles, os nomes atemáticos os que apresentam maior número de desvios (Ferreira 2011; Martins 2015).

Ainda no que concerne ao género, considerando os constituintes sintáticos que são afetados pelo desvio, verifica-se que, segundo a maioria dos estudos relativos a PLE, são os especificadores (e, em particular, os determinantes) que, predominantemente, surgem com formas não concordantes com a estrutura-alvo (Ferreira 2011; Godinho 2010;

15 Também Lucchesi e Baxter (2009) defendem a mesma posição relativamente ao português de Helvécia, comunidade rural de origem africana, no estado da Baía (Brasil).

Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Martins 2015). No entanto, Lacsán (2015), na sua investigação sobre a aquisição do género por aprendentes húngaros de português, refere que a classe dos adjetivos é aquela em que os informantes apresentam maior número de desvios.¹⁶

Tendo em conta o contexto específico do PLE, os estudos que têm essa variável em consideração mostram que existe uma tendência para um efeito positivo da instrução sobre o domínio de AGN/CGN. À medida que se avança no nível de proficiência, diminuem os desvios, sem, no entanto, se atingirem desempenhos semelhantes aos nativos (Ferreira 2011; Godinho 2010; Lacsán 2015; Mariotto & Lourenço-Gomes 2013; Martins 2015; Pinto 2015).

Já no que diz respeito aos desvios de concordância em número, várias investigações sobre VNN do português têm salientado que há um comportamento claramente distinto entre os elementos à esquerda do nome e os que se encontram à sua direita no sintagma nominal (SN), sendo os especificadores na adjacência esquerda do nome o local privilegiado de marcação do número plural no SN (Brandão & Vieira 2012¹⁷; Figueiredo 2009; Gonçalves 1997; Inverno 2009; Jon-And 2011; Miguel & Mendes 2013; Nascimento *et al.* 2008), dispensando-se frequentemente a marcação do número nos demais constituintes.

Apontam no mesmo sentido dados relativos a várias outras variedades de português, aí incluindo a brasileira (Brandão 2013; Brandão & Vieira 2012; Jon-And 2011; Miguel & Mendes 2013) e a europeia (Brandão & Vieira 2012), embora nesta a ausência de concordância seja residual, configurando uma regra (semi)categorica.

Por outro lado, no que concerne ao PLE, os estudos revelam que os desvios de CNN apresentam valores muito baixos, mesmo em fases iniciais de instrução, o que poderá decorrer do facto de a CNN obedecer a um padrão regular, sistemático e consistente. Além disso, embora de forma menos evidente do que nas VNN, observa-se também a ten-

16 Os dados de Lacsán (2015) coincidem com a descrição do português de Moçambique feita por Gonçalves (1997), que afirma que os erros de concordância de género afetam sobretudo a classe dos adjetivos.

17 Estas autoras comparam três variedades do português: português europeu, variedade do português do Brasil de Nova Iguaçu e Português de São Tomé. A variedade de Nova Iguaçu, apesar de não ser uma VNN, apresenta marcadas semelhanças com a variedade de São Tomé.

dência para marcação do número nos especificadores, prescindindo dos sufixos nos outros constituintes do SN (Martins 2015).¹⁸

Em alguns dos estudos referidos, as manifestações de desvio nas estruturas de AGN/CGN e de CNN são atribuídas, pelo menos parcialmente, a efeitos de transferência das LM dos aprendentes. No caso concreto de VNN africanas e de algumas variedades do português brasileiro, estará em causa a influência de substrato/adstrato de línguas bantu. No que respeita ao número, por exemplo, em diversas línguas bantu, o plural é marcado através de um elemento prefixal, que identifica a classe nominal. Esta característica estrutural das línguas ajudaria, segundo vários autores, a explicar a marcação do plural num constituinte à esquerda do nome, dispensando outras marcas no SN.

No entanto, se é verdade que grande parte dos falantes de PLNM considerados nessas investigações são também falantes de uma LM da família bantu, essa não é a situação de todos os informantes envolvidos, que evidenciam, ainda assim, comportamentos afins. Os dados, nomeadamente de PLE, sugerem que a presença/ausência, na LM, de um sistema de género nominal e respetiva configuração, bem como as formas de expressão e distribuição das marcas de número, não explicarão por si só as tendências gerais observadas (Figueiredo 2009; Lacsán 2015; Leiria 2006; Martins 2015; Pinto 2015).

3. Trabalho empírico

3.1. Metodologia

A amostra de informantes de PLNM que fornece os dados analisados neste trabalho é constituída por 120 sujeitos, autores de textos escritos que, recolhidos de acordo com um protocolo muito similar, integram o *Corpus* de Produções Escritas de Aprendentes de PL2 (PEAPL2) (<http://teitok2.iltec.pt/peapl2/>) e o *Corpus* PEAPL2 – *Subcorpus* Timor (<http://teitok2.iltec.pt/peapl2-timor/index.php?action=files>).¹⁹

18 Os dados de Godinho (2010) não confirmam inteiramente este facto. Os seus informantes tendem a marcar o plural no nome núcleo do SN.

19 Em ambos os casos, as informações relativas ao perfil sociolinguístico foram obtidas por auto-testemunho dos sujeitos. A principal diferença entre os procedimentos adotados na constituição do *Corpus* PEAPL2 (principal) e do *Subcorpus* Timor reside no número de estímulos usados para a recolha de dados, *i.e.*, 9 no primeiro caso e 5 no segundo.

Deste conjunto de informantes, 30 são falantes timorenses adultos com formação universitária que, no momento da recolha de dados, frequentavam o 3.º ano de uma licenciatura na Universidade Nacional de Timor-Lorosae. A média de idades desta subamostra (PTL) é de 24 anos (estando os valores individuais compreendidos entre os 20 e os 30 anos²⁰) e os sujeitos repartem-se de forma equilibrada entre os 2 sexos. Não dispondo de um critério objetivo para os distribuir de acordo com o seu nível de proficiência, e na ausência de indícios habitualmente considerados para o efeito (como a distribuição, em contexto de ensino formal, por níveis diferenciados), optou-se por prescindir dessa variável no estudo do comportamento linguístico destes informantes.²¹

Outro dado relevante, no domínio da caracterização da subamostra PTL, diz respeito à idade de exposição destes falantes ao *input* do português. Considerando as respostas à questão relativa ao ano do início de estudo do idioma e combinando esse dado com a idade do indivíduo no momento da recolha, foi possível segmentar esta sub-amostra em dois grupos: distinguiram-se, então, aqueles que foram expostos ao português até aos 10 (mas depois dos 6) anos e os que só depois desta idade o aprenderam (*cf.* Tabela 1). Observamos, então, que uma expressiva maioria entra em contacto com o português durante a infância, no ensino básico (*i.e.*, em contexto instrucional); o português surge, neste contexto, como a língua de escolarização claramente dominante (*cf.* Gráfico 1).

20 O valor de M0 (Moda) é de 23.

21 Os únicos dados disponíveis resultam de um exercício de autoavaliação que suscitou muitas dúvidas aos informantes, tendo sido frequentemente obtidas respostas múltiplas e contraditórias a uma mesma questão. Para uma descrição mais detalhada desta subamostra, *cf.* Xie Shanna (2017).

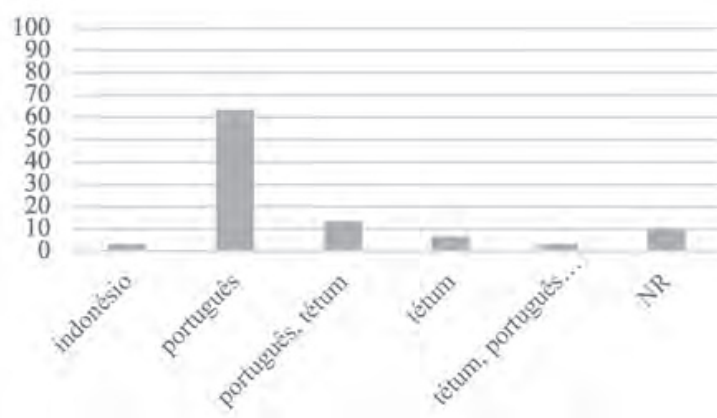


Gráfico 1. Caracterização da subamostra PTL: língua de escolarização (valores percentuais)

Refletindo a complexidade linguística e sociolinguística de Timor-Leste, são muito diversas as LM dos aprendentes (*cf.* Gráfico 2). Não obstante a limitação que esse facto, totalmente expectável, introduz no estudo (tornando complexa a análise do valor da LM no processo de aquisição/aprendizagem do português), é relevante, no âmbito do presente trabalho, o facto de, nas línguas mais representadas (mambai, makasae e tétum, língua franca e oficial do território), não se registar nem sistema flexional de número nem categoria de género, não havendo lugar, portanto, a mecanismos de concordância (Fogaça 2011; Himmelmann 2005; Hull & Eccles 2005).²²

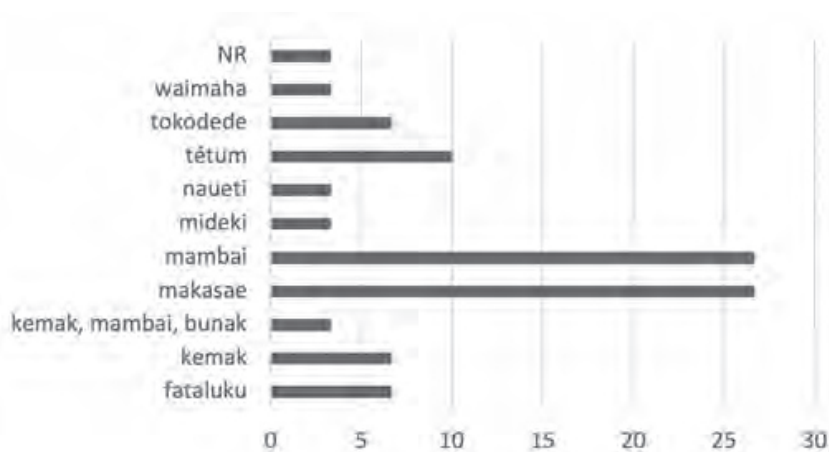


Gráfico 2. Caracterização da subamostra PTL: LM dos informantes (valores percentuais)

22 “From a morphological point of view, nouns in western Austronesian languages are usually unmarked. That is, they are not overtly marked for case, number or gender but rather occur in their lexical base form in most uses” (Himmelmann 2005, p. 128).

Atendendo à dispersão e ao número limitado e muito assimétrico de informantes por LM da subamostra PTL, bem como à escassez de dados atualmente disponíveis para a descrição desta variedade do português (facto que não permite constituir uma amostra de sujeitos mais equilibrada quanto a esta variável) não discutiremos, neste trabalho, eventuais efeitos de transferência da LM no desempenho dos aprendentes. Tratando-se, no entanto, de uma variável com relevância para a compreensão dos processos de aquisição/aprendizagem de LNM, essa abordagem deverá ser objeto de futuras investigações, baseadas em dados empíricos de outra natureza.

Finalmente, a subamostra PTL foi organizada considerando os domínios de uso nos quais, segundo os próprios, os informantes recorrem atualmente ao português. Esses dados permitiram, então, distinguir dois segmentos: (i) o dos indivíduos que usam a língua portuguesa tanto em contexto familiar como em domínios extrafamiliares; (ii) o dos informantes que usam o português unicamente em contextos extrafamiliares. Nenhum dos 30 falantes timorenses afirmou recorrer ao português unicamente em contexto familiar, o que confirma a natureza tendencialmente não nativa do português em Timor-Leste (*cf.* dados referidos na nota 1). Na Tabela 1 mostra-se, então, a distribuição dos falantes timorenses considerando as duas variáveis referidas, i. e., a idade de exposição ao *input* do português e os domínios de uso desta língua. Destaca-se, aí, o predomínio de sujeitos que, tendo sido expostos ao *input* do português até aos 10 anos, o usam o também (mas nunca exclusivamente) nas interações em família.

Já a subamostra PLE, como atrás referido, é constituída por 90 informantes adultos e estudantes de língua portuguesa da Faculdade de Letras da Universidade de Coimbra, de LM espanhola, alemã e chinesa (Martins 2015).

Tabela 1. Subamostra PTL: distribuição por idade de exposição ao input (PT) e domínios de uso de PT

	A (≤ 10 anos de idade)	Idade de exposição ao <i>input</i> (PT)		Total
		B (≥ 11 anos de idade)		
Domínio de uso de PT	a (contexto familiar e contextos extrafamiliares)	16 (53%)	2 (7%)	18 (60%)
	b (contextos extrafamiliares)	7 (23%)	5 (17%)	12 (40%)
Total		23 (77%)	7 (23%)	30 (100%)

As idades deste conjunto de aprendentes estão compreendidas entre os 17 e os 55 anos, registando-se um valor médio de 24 anos.²³ Os dados extraídos das respostas válidas obtidas por autotestemunho mostram que nenhum destes aprendentes foi exposto ao *input* do português antes dos 17 anos (valor médio de 22 anos)²⁴, o que permite concluir que estamos, em todos os casos, perante aprendentes tardios de português que, tipicamente, começam por contactar com este idioma em contexto instrucional.

Por outro lado, e uma vez que estes estudantes se encontram em imersão (e, portanto, deslocados do seu contexto familiar), não se afigurou aqui relevante proceder à distinção entre domínios de uso a que, no caso dos falantes de PTL, se procedeu.

A seleção das LM de aprendentes de PLE decorre do facto de estas serem línguas que “recobrem um espectro relevante de proximidade/distância tipológica em relação à LA no que concerne às categorias em análise” (Martins 2015, p. 33): se no espanhol as categorias de género e de número apresentam valores idênticos aos do português, registando-se igualmente fortes afinidades formais na sua manifestação morfológica, o alemão apresenta três valores de género e recorre a marcas morfológicas que, além de género e número, veiculam igualmente valores de caso. Por seu turno, o chinês não só não regista um sistema de variação em número e em género, como não possui um mecanismo de concordância nominal.

²³ 23 é o valor de Mo.

²⁴ Neste caso, o valor de Mo é 21.

Tabela 2. Estrutura da amostra de informantes

Informantes	N.º
LM Espanhol_A2	10
LM Espanhol_B1	10
LM Espanhol_B2/C1	10
LM Alemão_A2	10
LM Alemão_B1	10
LM Alemão_B2/C1	10
LM Chinês_A2	10
LM Chinês_B1	10
LM Chinês_B2/C1	10
PTL	30
Σ	120

Além da LM, variável pela qual os informantes desta subamostra se distribuíram de forma equitativa, foi também considerado, na investigação de Martins, o respetivo nível de proficiência, recorrendo-se, para tal, à escala do *Quadro Europeu Comum de Referência para as Línguas* (QECRL). Teve-se, aqui, em consideração o nível correspondente à turma frequentada na altura de produção dos textos escritos, tendo os sujeitos sido distribuídos por 3 grupos: A2, B1 e B2/C1.

Cada um dos 120 informantes da amostra (*cf.* Tabela 2) produziu um texto, em sala de aula, sem recurso a dicionários nem outro tipo de materiais de apoio, pelo que a base empírica da presente investigação é constituída por 120 textos escritos autênticos.²⁵

A metodologia que presidiu ao levantamento e organização dos dados no estudo da subamostra PLE (Martins 2015) foi replicada no tratamento das produções escritas dos sujeitos timorenses. Assegurou-se, desta forma, a comparabilidade dos dados que, por outro lado, e como já referimos, foram recolhidos seguindo um protocolo muito similar.

Assim, começou por se proceder à contabilização do número de nomes que surgem nos textos, de modo a avaliar a importância relativa desta classe de palavras no momento da produção escrita.

25 Os procedimentos de recolha dos textos produzidos pela subamostra PLE podem ser consultados em <http://teitok2.iltec.pt/peapl2/pt/index.php?action=metodologia> e os que, muito semelhantes a estes, presidiram à recolha dos textos escritos pelos sujeitos da subamostra PTL encontram-se detalhadamente descritos em Xie Shanna (2017).

Contabilizaram-se, seguidamente, todos os nomes que ocorrem em construções de concordância, bem como todos aqueles que constituem núcleos nominais isolados (*bare nouns*). Os valores obtidos serviram de referência para a avaliação da expressão quer das ocorrências convergentes, quer das que são desviantes relativamente à gramática do português europeu.

Nos casos em que se registaram desvios de concordância nominal (CN), em género e/ou em número, procedeu-se ainda à contabilização das ocorrências em função do constituinte afetado: distinguiram-se (i) especificadores (considerando os seus diferentes tipos), (ii) adjetivos (contabilizando separadamente os que desempenham função atributiva daqueles que desempenham função predicativa), (iii) nomes e (iv) outras categorias.²⁶ No caso específico dos desvios de AGN/CGN, procedeu-se, ainda, a uma análise tendo em consideração a estrutura morfológica do nome (em concreto, a respetiva classe temática), bem como o sentido da alteração do valor de género (masculino (alvo) → feminino (desvio) e feminino (alvo) → masculino (desvio)).

Por outro lado, considerou-se uma única categoria de não convergências (AGN/CGN), não se distinguindo, portanto, desvios de atribuição de género nominal e desvios de concordância de género nominal. Esta opção decorre do facto de, como salientado na Secção 1, não haver, em português, critério formal e/ou semântico seguro que oriente inequivocamente o aprendente na identificação do valor de género a atribuir aos nomes; nestas circunstâncias, e uma vez que os dados não foram elicitados, tendo sido, antes, extraídos de um *corpus*, a concordância sintática apresenta-se como a única forma sistematicamente válida de identificação do valor de género nominal em dados empíricos como os aqui analisados.

Esta decisão metodológica harmoniza-se com a posição expressa, por exemplo, por Corbett (1991), Aronoff (1994) e Franceschina (2005), que consideram que um sistema de género nominal se distingue de outros tipos de sistemas de classificação nominal precisamente porque desencadeia fenómenos de concordância: “A language will have gender if and only we find in that language (1) some form of agreement with nouns that (2) involves a distinction among noun classes, no matter

26 No levantamento dos dados, e ao contrário do que se faz em Martins (2015), optou-se pela não consideração de desvios que, traduzindo-se em “malformação” morfológica (desvios morfológicos) de classes de palavras relevantes para este estudo, não podiam ser associados a desvios de AGN/CGN e/ou de CNN.

what the semantic basis of the distinction may be” (Aronoff 1994, p. 66). Será, então, pela observação da estrutura formal das classes *target* e não da classe *trigger* (aquela para a qual o género é uma categoria intrínseca) (Franceschina 2005, p. 78) que se determinará, com segurança, o valor de género atribuído ao nome.

3.2. Resultados e discussão

Observou-se, então, que, no conjunto de 9153 palavras produzidas pelos falantes de PTL, ocorrem 2326 nomes (*cf.* Tabela 3); esta categoria corresponde, portanto, a 25% das formas vocabulares produzidas. Do mesmo modo, também no *subcorpus* PLE se regista uma percentagem de 21% de nomes (correspondendo a 4561 ocorrências), num universo de 21231 palavras (*cf.* Tabela 3). Ambos os *subcorpora* atestam assim, em proporção semelhante (facto que reforça a sua comparabilidade), “a importância desta classe na produção textual dos aprendentes”, bem como permitem antever “as repercussões que dificuldades associadas à aquisição das suas propriedades (...) poderão ter na configuração final dos textos que eles produzem” (Martins 2015, pp. 33-34).

Tabela 3. Caracterização da base empírica do estudo

Informantes	N.º textos	# de palavras	# de nomes	% de nomes face ao # palavras
LM Espanhol_A2	10	1343	274	20
LM: Espanhol_B1	10	2666	565	21
LM Espanhol_B2/C1	10	2553	557	22
LM Alemão_A2	10	1645	360	22
LM Alemão_B1	10	2593	554	21
LM Alemão_B2/C1	10	3171	713	22
LM Chinês_A2	10	2063	390	19
LM Chinês_B1	10	2201	470	21
LM Chinês_B2/C1	10	2996	678	23
Total PLE	90	21231	4561	21
Total PTL	30	9153	2326	25
Σ	120	30384	6887	23

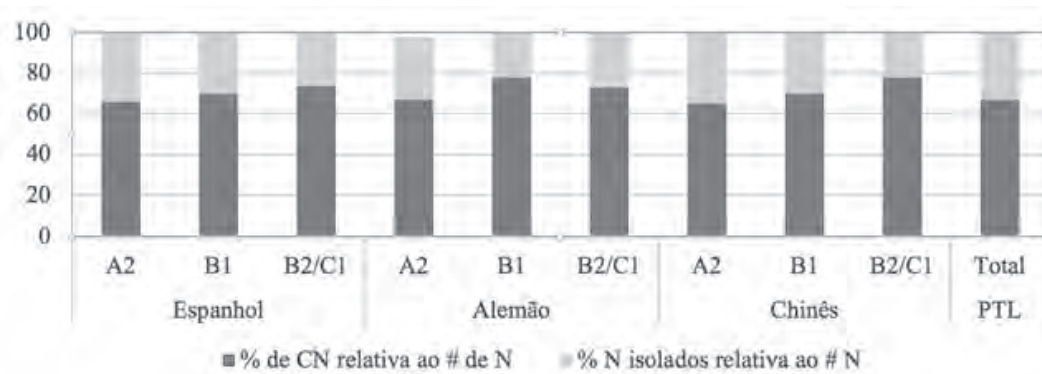


Gráfico 3. Distribuição de nomes em construções de concordância e fora de construções de concordância

Nas produções dos sujeitos timorenses, os nomes aí registados surgem maioritariamente (1564 ocorrências, correspondendo a 67% dos casos) integrados em estruturas que, no português europeu, ativam mecanismos de concordância plena, coocorrendo com especificadores (determinantes e quantificadores) e/ou com adjetivos (desempenhando função atributiva ou predicativa); as restantes 762 ocorrências configuram núcleos nominais isolados (*bare nouns*). O confronto do comportamento registado pelos aprendentes timorenses com o desempenho observado junto dos aprendentes de PLE descrito em Martins (2015) permite detetar um padrão idêntico na distribuição de nomes, tanto em construções de concordância, como fora de construções de concordância (*bare nouns*) (cf. Gráfico 3).

Por outro lado, embora o número total de desvios em estruturas de concordância nominal seja bastante inferior ao de acertos em qualquer segmento da amostra (cf. Gráfico 5), observa-se que ocorrências não convergentes com o português europeu surgem consistentemente em quase todos os textos dos falantes timorenses (cf. Gráfico 4). Não obstante a existência de uma clara tendência comum, *i.e.*, observa-se que os aprendentes “acertam, em qualquer um dos casos analisados, muitíssimo mais do que erram” (Martins 2015, p. 35), encontramos, entre os aprendentes de PLE, no que se refere à expressão deste tipo de desvios, valores mais baixos do que os registados nas produções dos informantes timorenses.

Há, no entanto, uma diferença considerável no número de produções escritas sem desvios no domínio da concordância nominal, entre os dois grupos da amostra; enquanto só cerca de 7% dos textos produzidos pelos informantes de PTL se apresentam sem desvios desta natureza,

à subamostra PLE corresponde um número substancialmente superior de produções textuais totalmente bem-sucedidas neste domínio: entre estes aprendentes, o valor mais baixo de textos sem desvios atinge os 20% e regista-se junto dos estudantes chineses dos níveis A2 e B1 (*cf.* Gráfico 4).

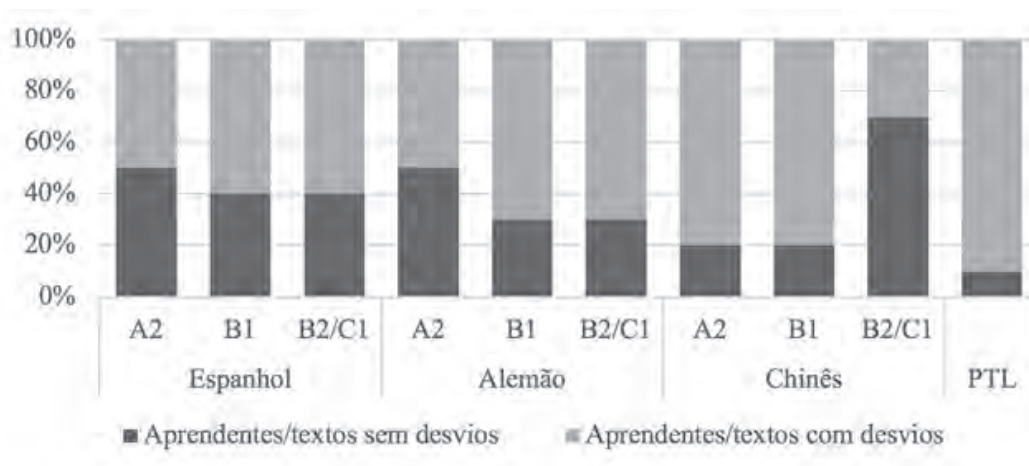


Gráfico 4. Aprendentes/textos sem e com desvios de AGN/CGN e CNN em função do perfil dos informantes (percentagens)

Se, no confronto dos dados, considerarmos o nível de proficiência (em associação com a LM) dos aprendentes de PLE, constatamos ainda que os falantes de PTL se aproximam mais do grupo de aprendentes chineses do nível A2, suplantando-os, no entanto, no número de ocorrências divergentes (*cf.* Gráficos 5 e 6). Este facto é particularmente expressivo quando consideramos que, como se referiu atrás, os informantes timorenses (i) foram, maioritariamente, expostos ao *input* do português, em contexto escolar, a partir do ensino básico (i. e., antes dos 10 anos), (ii) apresentam escolarização formal de nível universitário e (iii) tiveram, predominantemente (77%), o português como língua de escolarização.

A comparação dos dados relativos às duas subamostras revela convergência também no predomínio dos desvios relacionados com AGN/CGN que, assim, se configura como uma categoria idiossincrática, mais resistente à aquisição convergente com a LA. Os dados mostram, no entanto, uma percentagem global de desvios dos timorenses superior à registada nos aprendentes de PLE, inclusive nos níveis de proficiência mais baixos, apesar da idade de exposição ao *input* da maioria dos sujeitos falantes da VNN (*cf.* Gráficos 5 e 6).

A análise segmentada do comportamento da subamostra PTL revela-nos ainda que os valores mais elevados de desvios se registam nas produções escritas do grupo exposto ao *input* do português durante a infância (A) e do grupo que usa o português também em contexto familiar (a) (cf. Gráficos 5 e 6).²⁷ Os dados sugerem, portanto, a reprodução de modelo não nativo, configurando-se, deste modo, as condições para a consolidação do padrão de concordância variável.

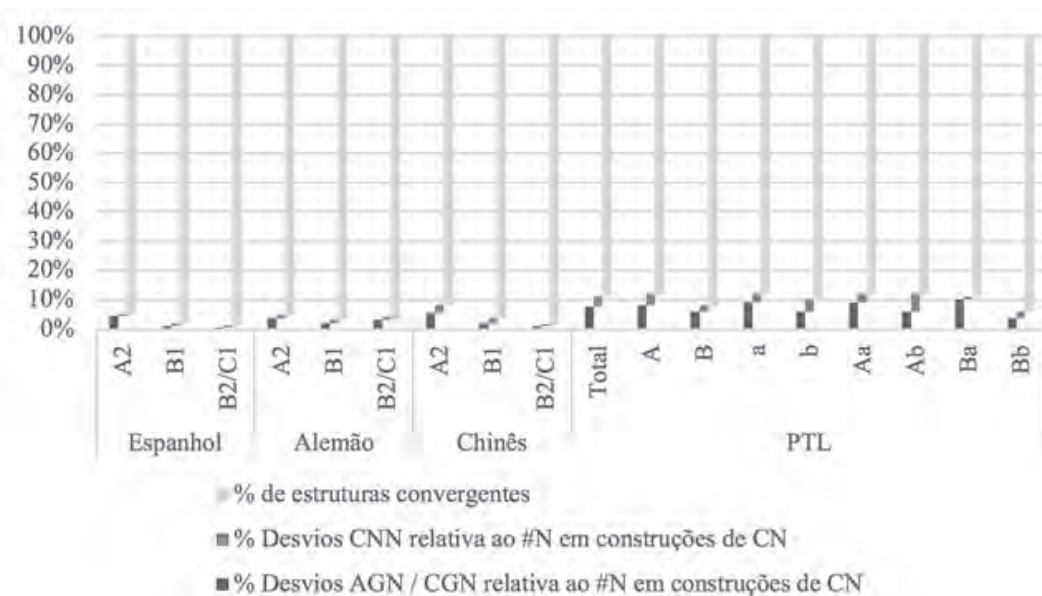


Gráfico 5. Estruturas convergentes e desvios de timorenses (PTL) e de aprendentes de PLE: distribuição

27 Embora em ambos os casos se apresentem os mesmos dados, o Gráfico 5 permite visualizar a relação entre acertos e desvios, ao passo que o Gráfico 6 torna clara a relação entre os valores dos desvios nos diferentes segmentos da amostra

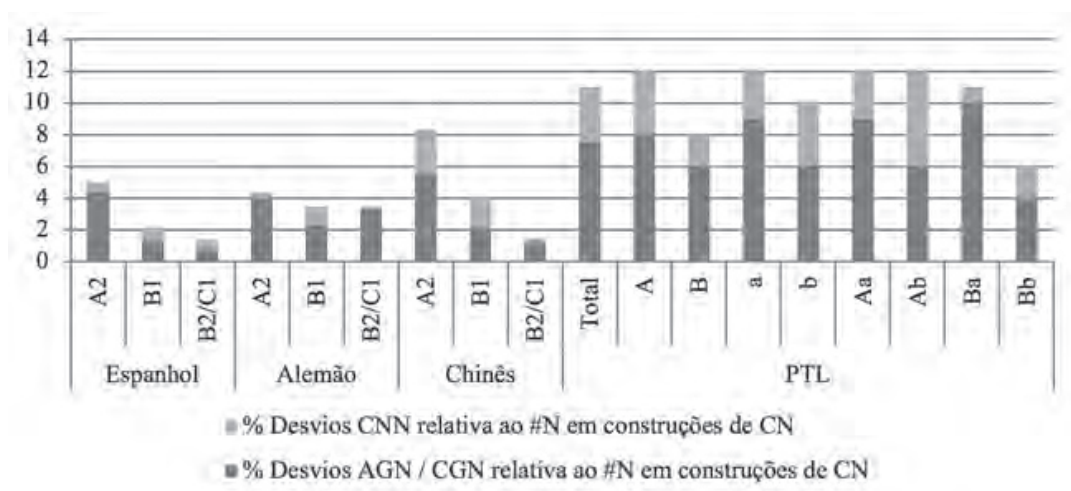


Gráfico 6. Desvios de timorenses (PTL) e de aprendentes de PLE: distribuição

Ao confrontarmos os dados relacionados com o constituinte sintático afetado pelo desvio, constatamos, novamente, um comportamento afim no que diz respeito às categorias AGN/CGN, com o predomínio da marcação do desvio no determinante (*cf.* Gráfico 7).

Já na categoria CNN, encontramos mais especificidades entre os dois grupos em estudo. Embora possamos destacar a afinidade que se traduz em valores percentuais idênticos, na marcação do desvio no nome, chama a atenção a assimetria que se observa no determinante: se, no caso dos alunos de PLE, não é esse o constituinte que mais ocorrências desviantes regista, no caso dos sujeitos timorenses, o determinante apresenta-se como o constituinte mais afetado, optando estes informantes tendencialmente pela marcação do valor de plural no nome (*cf.* Gráfico 8). Não obstante a limitação dos dados, é interessante o facto de este comportamento não convergir com o que ocorre noutras VNN, onde os especificadores na adjacência esquerda do nome surgem como o elemento privilegiado de marcação do número plural no SN (*cf.* Secção 2).

Neste trabalho de confronto dos dados relativos a grupos distintos de aprendentes de português, procedeu-se ainda à análise dos resultados em função da estrutura morfológica do nome afetado por desvios de AGN/CGN, observando-se igualmente a direção (masculino → feminino ou feminino → masculino) em que se processa o desvio. Assim, considerando a respetiva classe temática, foi possível observar que são os nomes de tema em -a que, em

ambos os grupos, se revelam mais vulneráveis (cf. Gráfico 9).²⁸

Ao mesmo tempo, as situações de desvio de AGN/CGN mostram-nos que o masculino parece funcionar, em ambos os casos, mas com mais expressividade entre os timorenses, como a categoria “default”. Vejam-se casos como *um cabeça* (PTL) e *o vida* (PLE – Martins 2015, p. 40).

Constituinte afetado pelo desvio (AGN/CGN) (exemplos)

1. Determinante

(PTL: *desta lugar*

PLE: *nesta dia*)

2. Quantificador

(PTL: *muito actividade*

PLE: *dois semanas*)

3. Adjetivo com função atributiva

(PTL: *as pessoas estrangeiros*

PLE: *uma paisagem bonito*)

4. Adjetivo com função predicativa

(PTL: *a cidade de Díli é muito limpo*

PLE: *a água do mar é muito limpo*)

5. Nome (PTL: *no pre-secundária*

PLE: *a irmão mais nova*)

6. Outros

(PTL: *a minha mãe foi falecido*

PLE: *por volta das 3,5 milhões*)

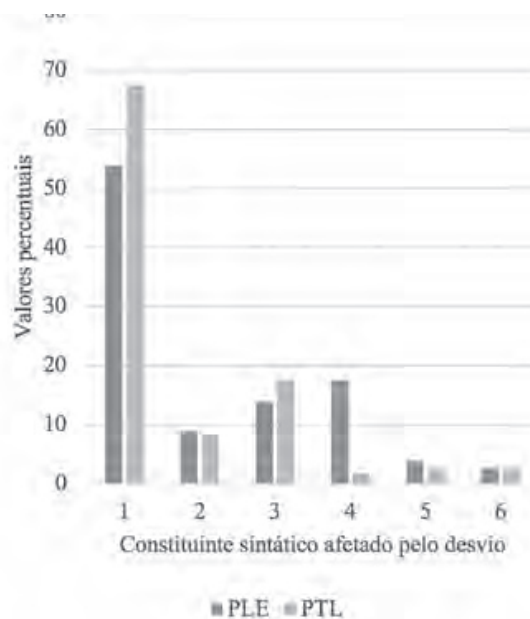


Gráfico 7. Constituinte afetado pelos desvios de AGN/CGN em função do perfil dos informantes

28 É aqui particularmente relevante o notado por Ferreira (2011), que analisa os dados disponibilizados no Corlex (Léxico Multifuncional Computorizado do Português Contemporâneo): “Numa análise da distribuição, por classe temática e por género, dos valores de frequência dos lemas nominais na base de frequências lexicais do português europeu (...), Ferreira (2011) apurou um peso de 38,71% de ocorrências de nomes não pertencentes às classes *-o* e *-a* (...). Nesse mesmo estudo observou-se que as frequências relativas dos nomes masculinos que integram a classe temática *-o* e dos nomes masculinos com IT *-a* correspondem a um peso de 59,1% no total das ocorrências” (Martins 2015, p. 28).

Constituinte afetado pelo desvio (CNN) (exemplos)	
1. <u>Determinante</u>	(PTL: <i>no tempos livres</i> PLE: <i>no tempos livres</i>)
2. <u>Quantificador</u>	(PTL: <i>nenhuma pessoas</i> PLE: <i>não registado</i>)
3. <u>Adjetivo com função atributiva</u>	(PTL: <i>horas livre</i> PLE: <i>as coisas nova da vida</i>)
4. <u>Adjetivo com função predicativa</u>	(PTL: <i>os animais vivem (...) contente</i> PLE: <i>a gente do sul me parecem mais alegres</i>)
5. <u>Nome</u>	(PTL: <i>as nossas vida</i> PLE: <i>muitas pessoas de país diferentes</i>)
6. <u>Outros</u>	(PTL: <i>não registado</i> PLE: <i>os livros deles [=do autor]</i>)

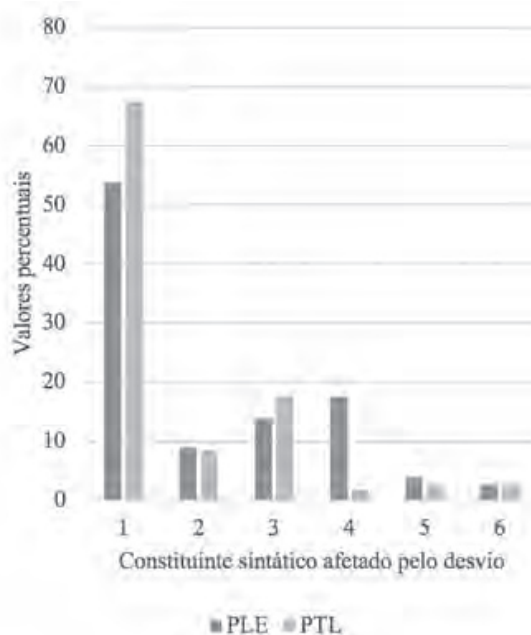


Gráfico 8. Constituinte afetado pelos desvios de CNN em função do perfil dos informantes

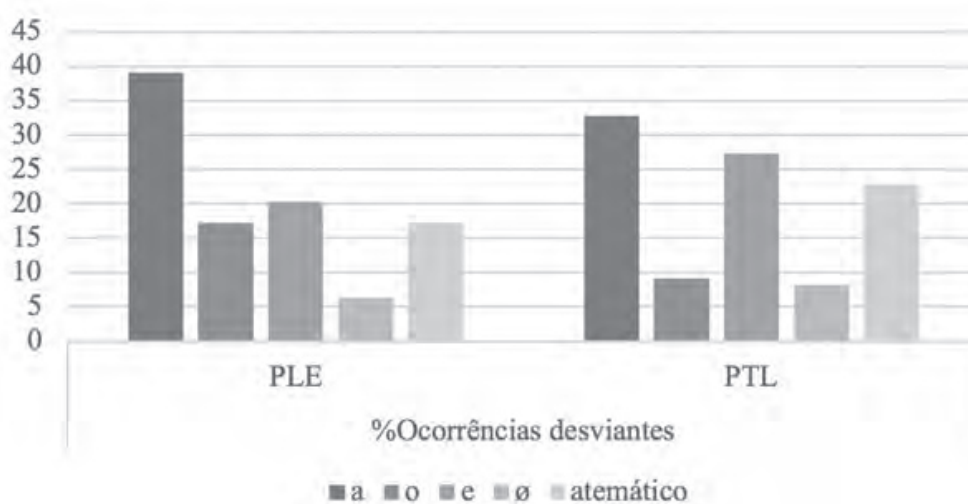


Gráfico 9. Distribuição percentual das ocorrências desviantes de AGN/CGN por classe temática do nome afetado

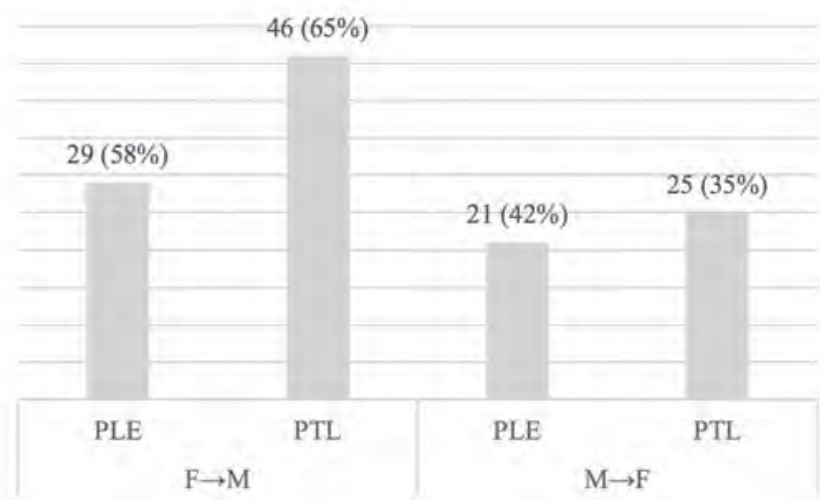


Gráfico 10. Nomes afetados por desvios de AGN/CGN em função do perfil dos informantes e da natureza do desvio: valores absolutos e percentuais

De facto, são de 58% e 65% (respetivamente para aprendentes de PLE e timorenses) as percentagens de atribuição indevida do género masculino a nomes que, na língua-alvo, são de género feminino; para a situação inversa registam-se os valores mais baixos de, respetivamente, 42% e 35% (*cf.* Gráfico 10).

Uma análise de maior pormenor, intersetando a natureza do desvio (i. e, o valor de género incorretamente atribuído) e a estrutura morfológica do nome afetado (*cf.* Gráfico 11), confirma-nos a vulnerabilidade dos nomes de tema em *a* e um comportamento simétrico entre timorenses e aprendentes de PLE nos casos de atribuição indevida do valor de masculino: os resultados mostram, aqui, posições relativas totalmente coincidentes. Há, no entanto, que destacar a divergência entre os dois grupos de informantes no que diz respeito aos casos em que o desvio se traduz na atribuição, indevida, do valor de feminino: aqui, os nomes de tema em *-a* revelam-se, junto dos aprendentes de PLE, como os menos afetados, registando os timorenses valores relativamente mais elevados de desvios nessa classe morfológica de nomes.

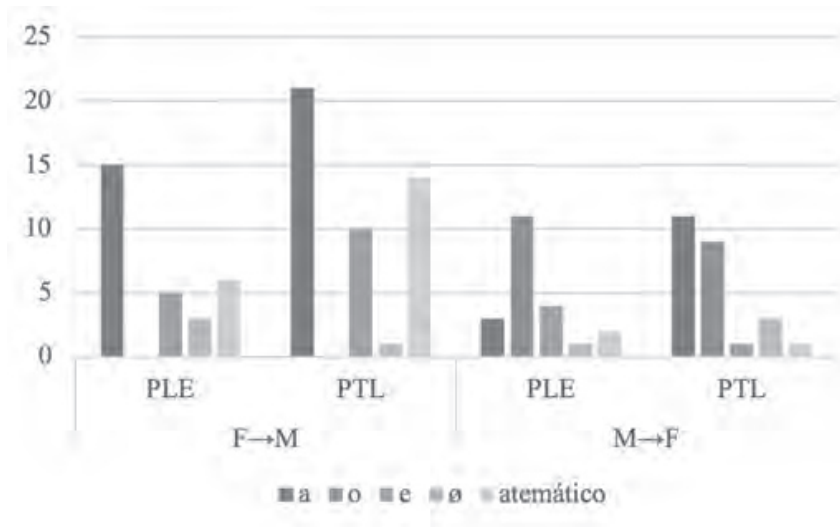


Gráfico 11. Número de nomes afetados por desvios de AGN/CGN por classe temática e em função da natureza do desvio

4. Conclusões

A concordância nominal variável é um fenómeno estruturante do PTL, variedade que assim converge, por um lado, com o atestado noutras VNN e, por outro, com o observado nos dados de aprendentes de PLE. A diferença que, apesar das tendências convergentes, se deteta por via do confronto dos dados recolhidos junto dos dois grupos de falantes de português parece residir, então, no grau de ocorrência do fenómeno. Na verdade, é no PTL que surgem com mais frequência desvios deste tipo e essa maior expressão regista-se quer no domínio da AGN/CGN, quer no âmbito da CNN.

Por outro lado, e de entre as categorias gramaticais que envolvem o nome, este estudo confirma a maior vulnerabilidade, junto dos aprendentes timorenses, da AGN/CGN, relativamente ao verificado com a CNN. Estes resultados estão em consonância com o observado na generalidade dos estudos sobre outras VNN e das pesquisas sobre o desempenho de aprendentes de PLE. No caso da AGN/CGN, confirma-se ainda, na subamostra de PTL, a tendência “masculino por defeito” e a preferência pela marcação do valor de género no determinante. No que se refere à CNN, e não obstante as afinidades assinaladas ao longo do texto entre o comportamento das subamostras PTL e PLE, a variedade não nativa timorense singulariza-se pela distribuição sintática dos des-

vios, favorecendo a não marcação do plural tendencialmente no determinante.

A inexistência de dados rigorosos sobre a proficiência linguística dos falantes timorenses constitui uma limitação deste trabalho. No entanto, a consideração do momento de exposição ao português e dos domínios de uso deste idioma permitem uma análise segmentada deste conjunto de informantes, revelando-se assim que são os grupos A (aquele que integra os indivíduos que aprenderam português antes dos 10 anos) e a (o que acolhe os indivíduos que usam o português também em contextos familiares) que apresentam os valores globais mais altos de desvios de AGN/CGN.

Por outro lado, a comparação do comportamento das duas subamostras revela-nos (e apesar das diferenças que separam os dois grupos de informantes quanto ao modo e momento de exposição ao português, línguavalvo) que os falantes de PTL apresentam um perfil linguístico mais próximo do observado nos aprendentes do nível A2 de PLE. Este facto sugere, tal como se discutiu na Secção 1, que não só o grau de exposição ao *input*, mas fundamentalmente a qualidade desse mesmo *input*, constitui um fator determinante na configuração das interlínguas dos aprendentes de LNM.

Dada a expressão muito residual de falantes de PTL como LM/L1, dificilmente se pode falar, para esta variedade do português, de um processo de “nativização”. Há, no entanto, que ter presente que a maioria dos sujeitos desta subamostra foi exposta ao *input* do português antes dos 10 anos de idade, desde o ensino básico, em contexto instrucional e que estas são condições que previsivelmente inibiriam a expressão tão acentuada da variabilidade dos padrões de concordância nominal. Assim, os dados sugerem que, à semelhança do que sucede noutras VNN em estádios de desenvolvimento mais avançados (*i.e.*, “nativizadas” em maior grau), a concordância nominal variável poderá vir a constituir-se como uma característica do PTL que o diferenciará do português europeu. Nas suas atuais condições de funcionamento, a escola timorense (domínio de uso em que ocorre a exposição mais relevante ao português) parece não fornecer nem *input* suficientemente robusto nem *feedback* corretivo que promovam o reforço da regra (semi)categorica de concordância nominal plena e a inibição das tendências observadas no conjunto dos aprendentes de PTL.

Referências

- Albuquerque, D. B. (2011). O português de Timor-Leste: Contribuições para o estudo de uma variedade emergente. *Papia*, 21(1), 65-82.
- Almeida, L. & Flores, C. (2017). Bilinguismo. In M. J. Freitas & A. L. Santos (Eds.), *Aquisição de língua materna e não materna: Questões gerais e dados do português* (pp. 275-304). Berlim, Alemanha: Language Science Press.
- Aronoff, M. (1994). *Morphology by itself*. Cambridge, Estados Unidos da América: MIT Press.
- Batoréo, H. J. & Casadinho, M. (2012). Perception and evaluation of language identity in East-Timorese speakers of Portuguese: Language and cognition in a multilingual community. In G. Mininni & A. M. Manuti (Eds.), *Applied Psycholinguistics* (pp. 57-61). Milão, Itália: Franco Angeli.
- Baxter, A. (1992). Portuguese as a pluricentric language. In M. Clyne (Ed.), *Pluricentric languages: differing norms in different nations* (pp. 11-43). Berlim, Alemanha: Mouton de Gruyter.
- Brandão, S. F. & Vieira, S. R. (2012). Concordância nominal e verbal: contribuições para o debate sobre o estatuto da variação em três variedades urbanas do português. *Alfa*, 56(3), 1035-1064.
- Brandão, S. F. (2013). Patterns of plural agreement within the noun phrase. *Journal of Portuguese Linguistics*, 12(2), 51-100.
- Cavele, S. A. (1999). *Cancelamento da marca de género feminino no adjetivo do português oral do Maputo* (Monografia para obtenção do grau de licenciado, Universidade Eduardo Mondlane).
- Chavagne, J. P. (2005). *La langue portugaise d'Angola* (Tese de doutoramento, Université Lumière Lyon 2).
- Corbett, G. (1991). *Gender*. Cambridge, Inglaterra: Cambridge University Press.
- Ferreira, T. (2011). *Padrões na aquisição/aprendizagem da marcação do género nominal em português como L2* (Dissertação de mestrado, Universidade de Coimbra).
- Figueiredo, C. F. G. (2009). A configuração do SN plural do português reestruturado de Almojarife – S. Tomé. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, 1(1), 28-55.
- Fogaça, J. (2011). *Aspetos gramaticais da língua makasae de Timor-Leste: Fonologia, morfologia e sintaxe* (Dissertação de mestrado, Universidade de Brasília).
- Godinho, A. P. (2010). A aquisição da concordância de número e a sua relação com a aquisição da concordância de género: Um estudo realizado com aprendentes chineses de português L2. In M. J. Marçalo *et al.* (Eds.). *Língua portuguesa: ultrapassar fronteiras, juntar culturas* (pp. 28-55). Évora, Portugal: Universidade de Évora.

- Gonçalves, P. (1997). Tipologia de “erros” do português oral de Maputo: Um primeiro diagnóstico. In C. Stroud & P. Gonçalves (Eds.), *Panorama do Português Oral de Maputo - Vol. II: A construção de um banco de “erros”* (pp. 37-70). Maputo, Moçambique: Instituto Nacional do Desenvolvimento da Educação.
- Gonçalves, P. (2010a). *A génese do Português de Moçambique*. Lisboa, Portugal: INCM.
- Gonçalves, P. (2010b). Perfil linguístico dos estudantes universitários: áreas críticas e instrumentos de análise. In P. Gonçalves (Ed.), *O Português escrito por estudantes universitários: descrição linguística e estratégias didáticas* (pp. 15-49). Maputo, Moçambique: Texto Editores.
- Gonçalves, P. (2013). O português em África. In E. B. P. Raposo, M. F. B. Nascimento, M. A. C. Mota, L. Segura & A. Mendes (Eds.), *Gramática do português*, vol. I (pp. 157-178). Lisboa, Portugal: Fundação Calouste Gulbenkian.
- Himmelman, N. (2005). The Austronesian Languages of Asia and Madagascar: Typological Characteristics. In A. Adelaar & N. Himmelman (Eds.), *The Austronesian Languages of Asia and Madagascar* (pp. 110-181). Londres, Inglaterra: Routledge.
- Hull, G. & Eccles, L. (2005). *Gramática da língua tétum*. Lisboa, Portugal: Lidel.
- Inverno, L. (2009). *Contact-induced restructuring of Portuguese morphosyntax in interior Angola. Evidence from Dundo (Lunda Norte)* (Tese de doutoramento, Universidade de Coimbra).
- Jon-And, A. (2011). *Variação, contato e mudança linguística em Moçambique e Cabo Verde: A concordância variável de número em sintagmas nominais do português* (Tese de doutoramento, Universidade de Estocolmo).
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge, Inglaterra: Cambridge University Press.
- Labov, W. (2003). Some sociolinguistic principles. In C. B. Paulston & G. R. Tucker (Eds.), *Sociolinguistics: the essential readings* (pp. 234-250). Hoboken, Estados Unidos da América: Blackwell Publishing.
- Lacsán, V. (2015). *The acquisition of gender agreement in L2 Portuguese by adult Hungarian speakers* (Dissertação de mestrado, Universidade de Lisboa).
- Leiria, I. (2004). Português língua segunda e língua estrangeira: investigação e ensino. *Idiomático*, 3. Disponível em: <<http://cvc.instituto-camoes.pt/idiomatico/03/portuguesLSeLE.pdf>>. Acedido em: 4 set. 2018.
- Leiria, I. (2006). *Léxico, aquisição e ensino do Português europeu língua não materna*. Lisboa, Portugal: FCG/FCT.
- Lucchesi, D. & Baxter, A. (2009). A transmissão linguística irregular. In D. Lucchesi, A. Baxter & I. Ribeiro (Eds.), *O português afro-brasileiro* (pp. 101-124). Salvador, Brasil: EDUFBA.

- Mariotto, E. & Lourenço-Gomes, M. C. (2013). Análise de erros na escrita relacionados à aprendizagem da concordância de gênero por falantes nativos do inglês, aprendentes de português europeu como língua estrangeira. In *Anais do IV Simpósio mundial de estudos de língua portuguesa (SIMELP). Língua portuguesa: Ultrapassando fronteiras, unindo culturas* (pp. 1278-1285). Goiás, Brasil: UFG.
- Mariotto, E. (2014). *Processamento da concordância de gênero por aprendentes de português como língua estrangeira. Evidências de um estudo de leitura automonitorada* (Dissertação de mestrado, Universidade de Lisboa).
- Martins, C. (2015). Número e gênero nominais no desenvolvimento das interlínguas de aprendentes do português europeu como língua estrangeira. *Revista Científica da Universidade Eduardo Mondlane, Série: Letras e Ciências Sociais*, 1(1), 24-49.
- Miguel, M. & Mendes, A. (2013). Syntactic and semantic issues in sequences of the type (Adjective)-Noun-(Adjective). *Journal of Portuguese Linguistics*, 12(2), 151-185.
- Nascimento, M. F. B. do., Pereira, L., Estrela, A. & Gonçalves, J. B. (2008). Aspectos de unidade e diversidade do português: as variedades africanas face à variedade europeia. *Veredas*, 9(1), 35-60.
- Peres, J. A. (2013). Semântica do sintagma nominal. In E.B.P, Raposo., M.F.B. Nascimento, M.A.C. Mota, L. Segura & A. Mendes (Eds.). *Gramática do Português*, vol. I (pp. 71-81). Lisboa, Portugal: FCG.
- Pinto, J. (2015). A aquisição do gênero e da concordância de gênero em português língua Terceira ou língua adicional. In P. Osório & M. J. Grosso (Eds.), *Teorias e usos linguísticos – Aplicações ao português língua não materna* (pp. 91-110). Lisboa, Portugal: Lidel.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209-231. Disponível em: <<http://pt.scribd.com/doc/88547455/Selinker-Interlanguage#scribd>>. Acedido em: 7 set. 2018.
- Schmitz, J. R. (2014). Looking under Kachru's (1982, 1985) three circles model of World Englishes: The hidden reality and current challenges. *Revista Brasileira de Linguística Aplicada*, 14(2), 373-411.
- Timor-Leste Population and Housing Census 2015: Population by mother and age, Timor-Leste*. Disponível em: <<http://www.statistics.gov.tl/pt/category/publications/census-publications/>> Acedido em: 11 set. 2018.
- VanPatten, B. & J. Williams (Eds.) (2008). *Theories in Second Language Acquisition: An introduction*. Nova Iorque, Estados Unidos da América: Routledge.
- Vieira, S. R. & Brandão, S. (2014). Tipologia de regras linguísticas e estatuto das variedades/línguas: a concordância em português. *Linguística*, 30(2), 81-112.
- Villalva, A. (2000). *Estruturas morfológicas. Unidades e hierarquias nas palavras do português*. Lisboa, Portugal: FCG/FCT.
- Villalva, A. (2003). Aspectos morfológicos da gramática do português. In M. H. M. Mateus et al. (Eds.), *Gramática da língua portuguesa* (pp. 917-983). Lisboa, Portugal: Caminho.

Xie Shanna (2017). *O corpus PEAPL2-Timor. Edição e disponibilização de uma infraestrutura de investigação em PLS* (Dissertação de mestrado, Universidade de Coimbra).

[recebido em 15 de setembro de 2018 e aceite para publicação em 14 de fevereiro de 2019]