

## **ASPECTOS LEXICAIS DO PORTUGUÊS USADO NO ENCLAVE (TIMOR LESTE)**

*Maria José Albarran Carvalho*

Escola Superior de Educação de Beja

(licença especial no apoio à transição em Timor Leste)

**RESUMO:** *Aplicando conceitos da linguística lexical e quantitativa, desenvolve-se uma primeira descrição do léxico do português em Timor, ainda há pouco tempo definido como Língua Oficial, pondo a nu alguns, ainda que poucos, traços de património linguístico próprio, i.e., variedade do português, em itens lexicais demarcados do Português Europeu Contemporâneo. A serem considerados nos manuais em elaboração para o território, tal como nos estabelecimentos de ensino que, em Portugal, recebiam bolseiros timorenses. O estudo centra-se nos dois extremos mais conservadores de traços próprios, a Ponta Leste e o Enclave do Oecusse, zonas onde a autora assumiu a coordenação / execução da Formação de Professores ao nível do Aperfeiçoamento.*

Da baixíssima densidade lexical do desempenho escrito – esperável em utentes do português como língua oficial – de discentes da Formação de Professores, Curso de Aperfeiçoamento (Lautém, Oecusse), e da criatividade de alguns itens desse componente da competência gramatical, surgiu a necessidade de se analisar a variação realmente existente, ou melhor, direcções dessa diversificação, se cotejada com o Português Europeu Contemporâneo. Ou seja, se em 25 anos sem contactos com outros falantes de português, e proibidos de o usar, os membros da comunidade o mantiveram com as características encontradas, e todos têm a 4ª classe e cursos intensivos, do tempo da colonização portuguesa - que os capacitavam a leccionar como professores de posto, auxiliares de monitores escolares e monitores escolares – é legítimo concluir que não só eram eles que divulgavam a língua portuguesa, como foram eles quem preservou o modelo, ou variedade, de então. Desde já se esclarece que o conjunto de docentes/informantes observados, tem consciência da diferença entre o seu desempenho e o nosso, tendo solicitado uma palestra sobre a evolução do português, em Portugal, nestes 25 anos, por «verificarem que não se falava, nem escrevia, como antigamente»...

É baixa a densidade lexical, o que se desenvolve adiante, e é diversa. Por densidade lexical, em harmonia com S. Laviosa-Braithwaiter (1995), entende-se a relação entre tipos e ocorrências no vocabulário usado em unidades textuais, segundo a qual o vocabulário é variado quanto mais elevada é aquela relação, i.e., será mais fraco quanto menos um texto use palavras variadas. A medida resulta da «proportion d'éléments lexicaux ou mots pleins par rapport aux éléments grammaticaux» (M.

Baker: 483). É preocupação, que não carece de demonstração, o facto de a produção de falantes não nativos apresentar mais baixa percentagem de palavras plenas que gramaticais. Contudo, as palavras plenas que constituem o seu léxico activo oferecem algumas particularidades. No presente trabalho apenas se examina a variação face ao Português Europeu Contemporâneo, quer no uso de léxico diferente, quer na frequência no uso de léxico comum.

Na história recente da linguística, desde 1950 sobretudo, o recurso a *corpus* tem sido criticado porque: i) era utilizado como representativo de componentes/planos de língua a descrever e não como metodologia; ii) constitui mero exemplo de desempenhos, escamoteando a competência linguística; iii) primarizava a capacidade explicativa da investigação; iv) excluía da análise as ocorrências ausentes do *corpus*; v) reduzia a análise à mera descrição, censo de ocorrências do nível dos desempenhos; vi) excluía a introspecção de análises com maior poder teórico explicativo, etc. Alguns dos principais críticos foram os clássicos G. Leech (1991), Ch. Hockett (1948), N. Chomsky (sobretudo 1957-1964), Ch. Fillmore (1992). Contudo, decorridos 50 anos, com o desenvolvimento da informática e criação de programas adequados, nomeadamente para estudos lexicais, para desambiguação de constituintes não decomponíveis, de colocações, etc. - com introdução de variáveis sociolinguísticas, pragmáticas e mais - podem ser usadas potencialidades complementares de ambos os extremos apontados (McEnnery, T; Wilson, A:1996).

Optou-se pelo método do corpo de dados, todavia tão questionado, porque: i) se ia investigar uma colecção de dados de formandos adultos, não uma variedade de língua na totalidade dos seus aspectos, i.e., o desenvolvimento escrito, e também oral - não se observou diferença entre estas duas modalidades - de população particular, num domínio especial, pesquisado por anotação manual.

As escolhas linguísticas que presidiram à elaboração do *corpus* foram: os diferentes domínios de comunicação, para além do educacional; o género de texto - comentário, exposição e resumo / diálogo quotidiano, em contexto escolar.

O estrato populacional é quase exclusivamente docente, 95%, dos 36-55 anos, ambos os sexos, mas 80 % homens; ii) quase todo originário da região escolar eleita (98%), Lautém e Oecusse, segundo informação solicitada nas fichas de inscrição; iii) o período de tempo de uso da língua não primeira, o português, foi igual.

Salienta-se que as zonas examinadas correspondem aos dois extremos opostos de Timor Leste, portanto mais isolados e preservadores de traços conservadores, sobre os quais é opinião corrente a ideia de neles se ter vindo a usar mais o português porque, sendo a primeira ponta a mais distante do território indonésio, usou mais o português como língua veicular, até ao ordinário da missa ter passado a ser em tétum,

havendo 2% de pessoas que o mantiveram no domínio familiar; a segunda ponta, por se revestir da forma de enclave, tanto tem recorrido menos ao tétum – ainda hoje, 1/3 da celebração da missa, cânticos e orações, é feito em baiqueno, aliás língua falada em todo Timor Ocidental - como manteve traços da presença portuguesa, inexistentes no restante território, por exemplo nos antropónimos, nomes próprios individuais e apelidos, v.g.: Tarsísio, Lopo, Martinho (M.J. Carvalho: 2001). Neste, roçam os 4 % os indivíduos que o vieram usando na rede familiar. O conjunto de dados é representativo, sob o ponto de vista regional, pela real inventariação da totalidade de docentes de português, já professores sob o domínio colonial português. A legitimidade referencial, instrumental e identitária do conjunto obtido possibilita precisão na identificação e tratamento dos fenómenos linguísticos seleccionados. Impõem-se precauções nomeadamente face a eventual reducionismo no conceito de erro, de interferência cultural, etc., como perspectiva de censura de variedades ou de factores orientadores de variação, mesmo até preconceitos que perturbem a objectividade descritiva. É também factual que o registo social releva de uma *praxis* em comunidade, espelhando a insegurança lexical face ao OUTRO, no caso, o antigo colonizador, cujo modelo se procura seguir, objectivo a não ser colocado. O peso dos modelos de referência será analisado adiante. Partindo da consciência de preocupações cautelares, que colmatem as estreitezas previstas, julga-se que o *corpus* obtido, pela sua extensão (200 000 palavras gráficas) – totalidade de ambas as regiões – e pela homogeneidade do perfil das fontes de informação, constitui o documento claramente estruturado e delimitado que a análise diferencial requer, validando-se, assim, os contrastes, particularismos a caracterizar. Por palavra gráfica entendeu-se todo o item separado por espaços, critério da investigação no Centro de Linguística da Universidade de Lisboa, usado na definição do Português Fundamental, com o qual poderia ser necessário estabelecer algum contraste.

Optou-se por um corpo de dados não muito abundante, para maior número de especificidades serem detectáveis, porque, em experiências anteriores (M.J. Carvalho: 1991) se verificou nem sempre ser verdade essa melhoria de informações e, também, segundo a posição recente, e no mesmo sentido, de S. Lafage (1997: 93).

Exposto o objecto de investigação, aborda-se o objectivo de caracterizar particularismos que: i) destaquem os desfazamentos do livro didáctico, das obras de leitura em português face à problemática do real desempenho discente, não só por não terem sido feitos especificamente para a realidade timorense, também em termos de variedade usada, a norma culta europeia, que não contempla o léxico do território; ii) abram vias para inovação pedagógica junto de público jovem ou adulto; iii) chamem a atenção para variações espaciais, caracterizadas pela interioridade e questões geoculturais, além da variedade dialectal; iv) deixem dados a serem tratados no âmbito das variedades do português.

### Descrição de corpus

O corpo de dados foi constituído a partir da totalidade dos trabalhos escritos que os destinatários dos Cursos de Formação de Professores realizara, quer no nível de Aperfeiçoamento, quer no nível da Iniciação, em Lautém e Oecusse. Prosseguindo especialistas como Ch. Muller (1977, reimp. 1992; 1973, reimp. 1992; 1977; 1979; 1993), convém introduzir termos que, doravante, serão usados na acepção dos teóricos da área da linguística quantitativa e lexicográfica: i) Palavra, como unidade de texto e de léxico (1979: 125) – uma palavra, face a questões morfológicas e semânticas, equivale a diversas formas, e vice-versa. Em 1977 (reimp. 1992: 4-5), como unidade de enunciado ou texto, a palavra é «une unité graphique, séparée des unités voisines par un blanc ou une ponctuation». Definição útil por se ir lidar com enunciados escritos. Trata-se de unidade de um enunciado ou texto (1977, reimp. 1992: 4). ii) Enunciado, ligado a texto e *corpus* constituem um conjunto de palavras (1977, reimp. 1992: 5), reservando-se a designação texto para «tout énoncé ou toute succession d'énoncés, tout discours ou fragment de discours, sans limitation d'étendue, provenant d'un même locuteur ou sculpteur et présentant une certaine unité nommée corpus» (id). iii) Frequência e variação de frequência (1979: 145) são conceitos inter-relacionados. O primeiro surge em função de uma base, a frase, o verbo, as palavras (id: 146), (1977, reimp. 1992), sendo a frequência de efectivos dada relativamente ao total da população estudada. A frequência relativa incide sobre um tipo de efectivos, ou face ao total / ou ao total de apenas uma categoria, v.g.: verbos face ao total de palavras, verbos face ao total de verbos (1973, reimp. 1992: 45). O sentido de uma frequência resulta da comparação das frequências observadas em diferentes textos para os mesmos vocábulos, ou «entre celles d'un texte et celles d'un corpus de référence, ou encore comparaison des sousfréquences fournies par les différentes parties d'un texte» (1993). Daqui resulta o vocabulário característico, portanto significativo (1993: 51) de qualquer colecção de dados. Quanto ao cálculo de um vocabulário (1979:167), o vocábulo, como unidade do discurso, difere do lexema, unidade da língua, ou seja, dos sistemas linguísticos (1977, reimp. 1992: 5-6). Assim, o vocabulário de um texto tem por forma a palavra e a extensão do vocabulário de um texto constitui apenas uma parte da estrutura lexical do texto (1979: 168), que requer cálculo sobre o vocabulário e a sua distribuição. Na contagem das palavras entram todas as formas, na do vocabulário só palavras diferentes, normalmente em menor número do que o vocabulário (1993: 51). iv) O léxico (1977, reimp. 1992: 6), retomando o postulado saussuriano de a língua preexistir ao discurso, sua «actualization contingente» (id), é situado por Ch. Müller na língua, como conjunto de lexemas. Cada «mot du texte est une occurrence d'une forme d'un lexème » (id: 7 ), evitando-se a confusão de ocorrências com frequência. Portanto «l'ensemble des lexèmes qui ont des occurrences dans un texte est dit vocabulaire de ce texte ; les éléments de cet ensemble seront dits vocables» (id.: 7). O linguista sintetiza, no quadro seguinte, a terminologia aqui adoptada (id.: 8):

Niveau	Ensemble	Élément	Définition
Langue	Lexique	Lexème	Lemme + cat, gramm
Discours	{ vocabulaire	Vocable (+Flex)+cont,	sém
{ texte	Mot	Forme graph.+place	

À frente se sistematizam, em alíneas, conclusões necessárias à pesquisa : i) «tout mot est une forme d'un vocable» ; ii) «l'ensemble des vocables (quel que soit le nombre, différent de 0, de leurs occurrences) est le vocabulaire du texte» ; iii) «les vocables, dans leur majorité ; sont identiques à des lexèmes du lexique ; ceux qui ne réalisent pas cette relation forment en sous-ensemble particulier du vocabulaire, et fournissent une sorte de supplément au lexique» ; iv) «le vocabulaire d'un texte se présente donc comme une liste de lemmes, soit dans leur ordre alphabétique, soit classés par fréquence» (ib).

Na estatística lexical (1979: 232 ss), procede-se por estabelecimento de index de formas - dos vocábulos e de um quadro de distribuição das frequências (cada frequência exemplifica o número de vocábulos afectados por texto). É uma estrutura significativa que nada acrescenta ao conteúdo. Estas operações possibilitam a estipulação da riqueza lexical, noção relativa, classificando-se o vocabulário, de um texto, de RICO / POBRE numa relação entre si, mais objectiva se a extensão discursiva for idêntica (id: 235).

Aqui se insere a análise dos hapax , vocábulos que só aparecem uma vez (1993: 59). Obviamente, para determinadas extensões requer-se leitura humana directa e atenta, foi o que se fez no caso da colecção de dados em análise.

#### Considerações prévias:

- a abordagem seguida não divide o léxico da gramática, pois «l'unité prise en considération ne l'est pas en tant qu'élément constitutif d'un système lexical mais en tant que constituant d'un syntagme, ce qui revient à dire que le lexème n'est pas au centre de ces travaux», segundo F. Cusin-Berche (1999: 13). Mantendo-se unidade básica do léxico, oposta ao vocabulário, integrado na «parole», o lexema remete para a «langue», numa linha saussuriana . É pois, em termos instrumentais que aqui se perfilha a terminologia de Ch. Müller (1973 –1993).
- Não se aguardam indícios de qualquer riqueza lexical, conhecedores do facto de a alta frequência das palavras implicar: i) empobrecimento semântico ; ii) variedade polissémica ; iii) leque policategorial (J. Rey-Debove: 1998: 214). Prevê-se abundante frequência de escasso número de vocábulos.
- A repartição do vocabulário segundo níveis dos utentes (id: 215), no *corpus* em causa, dos indivíduos envolvidos, será homogénea, nas hipóteses prévias que antecedem as conclusões desta pesquisa.

- A palavra, constituindo unidade lexical, compõe-se de três elementos : forma, sentido e categoria gramatical (cf. A. Lehmann ; F. Martin-Berthet: 1998: 1). Deles decorrem, para além das triviais questões de flexão, polissemia e diversidade categorial, fontes de individualização do utente, fenómenos como: sinónimos, criatividade derivacional, sentido não literal, etc. Nesta zona, não se vai deter a investigação pela qual se optou, na linha de relacionar vocábulo, sentido e referência em fase posterior.

## CONCLUSÕES

A opção por colectânea de dados potenciadora das previsões acima avançadas encontra justificação no meio envolvente. A população discente observada radica quase exclusivamente – poucos apresentam outras origens – em Lautém e Oecusse – cuja realidade quanto a geografia, fauna, flora, descritores socioculturais, etc diverge do contexto integrador das gentes portuguesas.

Sob o ponto de vista linguístico, a região caracteriza-se por boa amostragem de traços arcaizantes: vestígios que encontram explicação no mero isolamento imposto pela insularidade, a interioridade rural (esta agudizada pela situação de Enclave), como seja a sobrevivência de: 1) formas morfologicamente medievais – Tersia, Lopo ; 2) léxico em desuso – Sancha, Ágata e outros antropónimos; 3) realizações fonologicamente conservadoras – [k]antidade (quantidade); valores semânticos correntes no passado – cruzado (moeda desaparecida). Ainda que esporádica e dispersamente, tais formas, entre mais, ocorrem hoje, no discurso oral, da tradição cultural ou não, sobretudo na população mais velha, mesmo em Portugal (M.J. Carvalho: 1997-2000). A ilusão de se obter «bom emprego», «vencimento alto», via grau de licenciatura, conduziu ao objectivo de ser «Sr. Dr.», a todo o custo. Mas a distribuição da «riqueza» não respeita as regras da democratização do ensino e o desejo de obter uma proficiência igual à dos portugueses, inclusive ao nível lexical, não facilitará a divulgação da língua que os timorenses escolheram como oficial, nem dará acesso ao emprego.

Segue-se a anunciada enumeração de itens, organizada por áreas temáticas, sintoma identitário de património linguístico autónomo, até na Língua Oficial, ainda que numericamente não muito significativo.

LISTA DE LÉXICO PORTUGUÊS (mais frequente ou específico do território, confrontando-o com o Português Europeu Contemporâneo - pesquisa do Centro de Linguística da Universidade de Lisboa - ou aportuguesado de empréstimo inicial)

*Vestuário*

Cambalique, cabaia, bajú, táis, lipa, salenda

*Instrumentos de trabalho*

Machado, enxada, pá, azagaia / zagaia, lança, zarabatana, seta, catana, parão, lapiseira (= esfográfica), teares manuais

*Danças e instrumentos*

Tebedai, tebe, lorosae, dança do milhafre, dança da cobra, silate, coremetam, surubeque, tambores, babadoques, gongos, pratos, guitarra, viola, cavaquinho, violino, ferrinhos, bombo, corneta

*Alimentação*

Batata-doce, canja, inhame, maçaroca, mandioca, batata brava, batata europa; pentes (=cacho) de bananas, papaias, jaca, maracujá, cajú, abacate, diospiros, areca, tamarindo, jambalão, fruta-pão, manga, coração de bananeira, coração de boi, anona, ata, goiaba, maracujá, ananás; pipocas, milho, maçaroca, espiga de milho, arroz, painço, chuchu, abóbora, folhas de mandioqueira, de batateira, de feijoeiro, de aboboreira, flores e rebentos de papaieira; catupa, sassates, carne de búfal, lémur, javali, porco bravo, tubarão, sardinha, enguia; doces (bolos e patéis de banana frita, de farinha de trigo e açúcar, etc); matabicho (comum a bolsas do português transmuntano, de moçambique, Angola, Brasil, entre mais)

*Bebida*

Tuaca, leite, água de coco, cerveja, vinho (de palma, de tua, de acadiro, de coqueiro)

*Jogos e apostas*

Luta de galos, corrida de beiros, de cavalos, cartas

*Adornos*

Luas, meias luas, cordões de moedas, mutissala, fechos de cabelo, anéis, coral, botões, argolas, cintos

*Manufacturas*

Tabaqueiras, bainhas de catana, beiros, cestaria (porta-moedas, sacos para café, para sal, para dinheiro, para bétel, areca e cal, pirogas, olaria (potes, panelas, alambiques, etc), esteiras cadeiras de rota, lantém

*Pássaros*

Loricos, cacaoques, pombos, rolas, milhafre, águia, corvo, codorniz, catatua

*Animais domésticos*

Macaco, ovelha, carneiro, cordeiro, cabra, bode, pato, ganso, búfalo

*Outros animais (de caça ou não)*

jacaré / crocodilo, cobra (verde, verde e vermelha, pitão / jibóia), macaco, cavalo, búfalo, lémur, lacrau, aranha (vermelha, etc), veado, javali, porco bravo

*Materiais domésticos*

Lantern, geleira, esteira, terrina, pote, panela (comum a Moçambique) / tacho

*Árvores e plantas*

Bambu, palapeira, arequeira, tuaqueira, tabaco, casuarina, acácia, palavão (branco e preto), madres del cacao, acadiro, coqueiro, bananeira, romanzeira, ateira, sumaumeira, jaqueira, papaieira, teca, gondões / gondoeiros, sandaleiro, tamarindeiro, mogno. Pau rosa, pau preto, pau ferro, pau branco

*Natureza (trabalhada / não trabalhada)*

Coilão, picos, sucacos, vázeas, hortas, mangal, cafezal

*Máquinas*

Lagarto (maquinaria agrícola)

*Divisão do Tempo*

1h de tempo (comum a Moçambique)

## BIBLIOGRAFIA

BAKER, M (1998). «Réexplorer la Langue de la Traduction : une Approche par corpus», META, vol 43 :481-485.

CARVALHO, M.J. (1991), Aspectos Sintáctico-semânticos do Português Oral de Maputo – Verbos Locativos, Lisboa / Luanda, ICALP-Angolê

CARVALHO, M.J. (2001) «Apontamentos Linguísticos sobre a Antroponímia no Enclave», Palavras, 20 (prelo)

CUSIN-BERCHE, F (1999), «La Notion d' 'Unité Lexicale' en Linguistique et son Usage en Lexicologie». LYNX ,n40 N11-39.



LAFAGE, S.(1997). «Extensité et Cohérence». C. FREY, D. Latin Le Corpus Lexicographique et de Gestion. Lovaina: Duculot.

LAVIOSA-BRAITHWAITE, S. (1995) «Comparable Corpora : towards a Corpus Linguistic Methodology for the Empirical Study of Translation». LEHMANN, A ;

MARTIN-BERTHET (1998). Introduction à la Lexicologie – Sémantique et Morphologie Paris : Dunod.

McENERY, T. ; WILSON, A (1996). Corpus Linguistics. Edimburgo : Edinburgh University Press.

MÜLLER, Ch. (1968, reimp1973) Initiation à la Statistique Linguistique. Paris: Larousse

(1977). Principes et Méthodes de Statistique Lexicale. Paris: Hachette. (Reimp 1992). Paris: Champion.

(1979 a). Le Vocabulaire du Théâtre.... Etude de Statistique Lexicale. Genève: Slatkine.

(1979 b). Langue Française et Linguistique Quantitative. Genève: Slatkine.

(1992a). Principes et Méthodes de Statistique Lexicale. Paris: Champion.

(1993 ). Langue Française, Débats et Bilans-Recueil d'Articles. Paris: Champion-Slatkine.

(1993 c). Initiation aux Méthodes de la Statistique Linguistique. Paris: Champion.

REY-DEBOVE,J.(1998). La Linguistique du Signe. Paris : Colin.