

"Origens e estruturação histórica do léxico português" (1976)

in *Estudos de Linguística Histórica Galego-Portuguesa*, Lisboa, IN-CM, 1989, pp. 9-16.

Joseph-Maria PIEL

1. O léxico de uma língua de civilização como a língua portuguesa é um **organismo vivo**, extremamente complexo na sua composição, pois resulta de um trabalho multissecular de elaboração e de selecção, cujos princípios se situam bastante para além da época em que o português se manifesta como instrumento literário nos primeiros documentos escritos (cerca de 1200). Como sucede com o léxico das demais línguas de cultura, nunca será possível reconstituir todas as fases por ele percorridas e destrinçar a contribuição das muitas gerações que nele colaboraram até se constituir o magno edifício que hoje se nos depara nos grandes dicionários modernos. Este longo e laborioso passado explica a **falta de homogeneidade**, em perspectiva histórica, que caracteriza a sua estrutura. Com efeito, o conjunto vocabular do português nunca se manteve estacionário, antes evoluiu constantemente num ritmo ora mais, ora menos acelerado, evolução que não chegou ainda ao seu termo e que representa o esforço comum de homens procedentes de ambientes geográficos e sociais muito diversos. Entre as formas faladas-populares e as cultas-literárias observa-se uma permuta contínua e fertilizadora, em ambos os sentidos. Ao caudal vocabular do primitivo património afluíu, em ondas sucessivas, uma infinidade de elementos estrangeiros, europeus e extra-europeus, entrando em linha de conta praticamente todas as línguas com que os Portugueses, no decorrer da sua história, estiveram, directa ou indirectamente, em contacto.

2. Concebida em moldes modernos, a disciplina que se ocupa da origem das palavras, ou seja a **etimologia**, já não se contenta com identificar a forma actual de um vocábulo com a sua forma primitiva subjacente, mas esforça-se por reconstituir a sua história externa (fonética) e interna (semântica) através dos séculos. É verdade que este ideal se revela tanto mais difícil de alcançar quanto é certo que a origem de um número ainda demasiado avultado de elementos continua duvidosa ou mesmo totalmente enigmática.

3. Desde que, há pouco mais de um século, os eruditos obtiveram a última certeza de que as hoje chamadas línguas românicas se entroncam no latim falado nos últimos períodos do Império Romano, também o léxico português foi revelando progressivamente, as suas fontes primordiais. Costuma chamar-se, com um termo pouco apropriado, «vulgar» a este latim desprezioso, coloquial, que existiu desde sempre ao lado do latim escrito, o qual usa um vocabulário literário mais apurado e conservador em relação às formas vivas, progressivas, da língua. Por outro lado, sabemos que, não obstante a sua relativa homogeneidade, resultante da homogeneidade da própria civilização romana, o **léxico «popular»** não deixava de apresentar certas particularidades geograficamente condicionadas e circunscritas. Resulta daí que o léxico português, no que respeita ao seu pecúlio mais vetusto, vem a ser o prolongamento do léxico corrente hispano-latino, mais concretamente do tipo próprio do noroeste da Península, que podemos qualificar de «galaico-lusitano». De certo modo a história do vocabulário português começa, pois, com a romanização das regiões que foram o berço do Idioma. O momento crucial, a partir do qual vai adquirindo paulatinamente o seu perfil individual perante as outras línguas românicas, incluindo as hispânicas, é o da ruína do Império (séc. V), resultante da acção militar e política de povos invasores

germânicos, no nosso caso dos Suevos e Visigodos, os quais se estabeleceram no seu solo. Uma vez destruída a unidade geográfica e cultural do Império, as províncias adquirem uma autonomia administrativa e política própria, a qual acelera a desintegração linguística. Deste modo desenvolvem-se e consolidam-se progressivamente diferenças regionais, já anteriormente existentes em estado latente.

4. Dois séculos mais tarde, a conquista e ocupação efectiva da maior parte da Península por populações muçulmanas, de língua árabe, vem modificar radicalmente o curso da evolução linguística, apagando lenta, mas quase completamente, nas partes meridionais e centrais da Península, as formas românicas regionais elaboradas na época da monarquia visigótica (sécs. VI e VII). Ignora-se ainda se estes **falares «moçárabes»**, que porventura puderam resistir até à reconquista do Centro e do Sul de Portugal, tiveram alguma interferência de vulto na constituição do vocabulário da futura língua comum portuguesa. Mesmo que admitamos esta hipótese, não parece que a acção respectiva fosse considerável.

Elementos pré-latinos

5. Sabido é que o latim triunfou das línguas autóctones graças ao seu prestígio como instrumento de uma civilização e de uma concepção política superiores. Obra de poucas gerações, a assimilação linguística foi certamente mais lenta nas regiões periféricas do N. O. da Península, de acesso mais difícil e daí mais tardiamente colonizadas que as outras. Embora os **idiomas pré-latinos**, falados nesses territórios, se extinguissem obscuramente, não deixaram, porém, antes de sucumbir, de comunicar ao latim provincial, de feição «galaica», um conjunto não desprezável de termos especiais, correspondentes muitas vezes a noções concretas, para as quais faltava, ou parecia faltar, na língua dos dominadores uma expressão equivalente. Referem-se estes elementos, entre outras categorias, a conceitos topográficos: *barranco*, *lapa*, *pala* ‘caverna’, *morro*, *mouta*, *várzea*, *veiga*; a espécies vegetais espontâneas: *tojo*, *carrasco*, *chaparro*, *mato*, a noções do ambiente rústico: *seara*, *broa*, *bezerro*, *bruxa*, etc. Alguns dos termos desta classe aparecem, aliás, expressamente qualificados como termos indígenas em autores da latinidade tardia: *cabana*, *cama*, *sarna* (Santo Isidoro), ou em inscrições romanas: *arroio* < ARRUGIUM, *páramo* < PARAMUS. A repartição geográfica dos referidos pré-romanismos é bastante desigual. Enquanto uns abrangem vastíssimas zonas da Península, prolongando-se a sua área mesmo muito além dos Pirinéus, há outros que se apresentam arrincoados em certas zonas galaico-portuguesas. O fundo linguístico de onde procedem não é menos variado. Ao lado de elementos celtas e de idiomas anónimos, pertencentes à mesma família indo-europeia, identificaram-se outros, tão pouco homogêneos, que têm as suas raízes em línguas aparentadas, possivelmente, com o basco ou com um substrato «mediterrâneo». Todas estas vozes, sobreviventes de um grande naufrágio, testemunham migrações e infiltrações, em tempos pré-históricos, de povos e raças, de que mal reza a história.

Fundo latino

6. O latim, que, tendo atingido uma feição relativamente homogênea e nivelada, se tornou a língua comum das províncias romanas, assimilara, ainda anteriormente à sua expansão fora da Itália, numerosos **elementos estranhos**: «mediterrâneos», etruscos e, principalmente, gregos. Mais tarde admitiu também alguns termos gerais e técnicos, próprios de povos submetidos ao domínio romano: celtas, «ibéricos», germânicos e outros. Quer isto dizer que, na altura em que o latim se fragmentou, dando origem aos dialectos românicos, o seu léxico era constituído por um conjunto de palavras em que o

elemento genuinamente itálico-latino só representava uma parte, embora a mais importante.

7. Como sucede com todas as línguas de civilização, o vocabulário do latim apresentava, segundo já se observou, **dois aspectos**, sociologicamente condicionados: um, conservador-literário; outro progressivo-popular. Foi este último que, depois da ruína da civilização antiga, se tornou decisivo na elaboração do léxico românico, ao passo que muitos termos «nobres», de tradição escolar-literária (como OS, ORIS, que cedeu o lugar a BUCCA, e ENSIS, substituído pelo grecismo SPATHA), se obliteraram para sempre. Uma terceira categoria de palavras pode qualificar-se de neutra, por abranger termos peculiares tanto da língua escrita como da falada: MATER, PANIS, BONUS, AUDIRE, etc. Devido a razões óbvias, o vocabulário «coloquial» latino, principalmente o de feição rústico-popular, é menos bem conhecido que o dos escritores e poetas. Em muitos casos são mesmo unicamente os idiomas românicos a testemunharem, indirectamente, a existência de uma palavra latina, pertencente àquela categoria vulgar, como acontece com *sossegar*, que pressupõe um vulgarismo: *SESS-ICARE.

8. O triunfo do **cristianismo** não só abriu a porta a neologismos e novos grecismos, como ECCLE:SIA 'assembleia dos fiéis', port. *igreja* (cf. FILIUS ECCLESIAE > port. *freguês*), mas impregnou também termos tradicionais latinos com significações novas (FE:RIA(E), port. *-feira*, nos nomes dos dias da semana). Por outro lado, estava definitivamente condenada a terminologia dos cultos pagãos, refugiando-se uma ou outra expressão respectiva no vocabulário rústico e dialectal (cf. o dial. *jã* 'fada' < DIANA).

9. Em confronto com os demais léxicos românicos, o do português apresenta, como era de esperar, um flagrante paralelismo com o do castelhano, o que não exclui divergências importantes. Vozes peculiares aos dois idiomas, como *pássaro/pájaro* < PASSER (propriamente 'pardal'), *medo/miedo* < METUS, *comer* < COMEDERE, etc., mais ou menos correspondentes, na sua significação, ao fr. *oiseau*, it. *ucello* < AVICELLUS, fr. *peur* < PAVO:RE, it. *paura* < *PAVU:RA, fr. *manger* it. *mangiare* < *MANDUCARE, deixam entrever um **fundo hispânico comum**. Este pode ser de carácter conservador, como nos exemplos apontados, ou inovador, como em *irmão/hermano* < (FRATER) GERMA:NUS, em relação ao fr. *frère* < FRATER, *cabeça/cabeza* < *CAPITIA, em confronto com o fr. *chef* < CAPUT, *Coração/corazón* < *COR-ATIO, -O:NE, em relação ao fr. *coeur* < COR.

10. No entanto, não raro, e amiúde solidariamente com o galego-asturiano, o léxico português apresenta uma **individualidade inconfundível**, mesmo em comparação com o castelhano. A atitude conservadora dos falares periféricos respectivos reflecte-se em formas como *colmo* < CULMUS, *adro* < ATRIUM, *gume* < ACU:MEN e outras. Estes traços individuais podem ascender ainda ao período romano, ou resultar de inovações regionais. De modo geral, os falares do N. O. da Península, em Portugal sobretudo os do Minho com as zonas limítrofes, revelam, no âmbito rústico, uma riqueza e especialização lexical e semântica surpreendentes, facto que nos permite inferir que a romanização das referidas regiões, embora mais tardia, não foi menos eficaz e profunda que no resto da Península. Ao lado de muitos arcaísmos puderam identificar-se não poucos neologismos populares da época romana, como em *branha* 'pastos de verão' < VERA:NEA, *encedoiro* < *INCITORIUM. Por via de regra trata-se, na verdade, de formas regionais pouco conhecidas ou mesmo ignoradas da língua comum, o que não impede que ofereçam um real interesse para a história da cultura rústica do latim.

Elemento germânico

11. Dos termos que o latim foi buscar a vários **dialectos germânicos**, só poucos foram transmitidos às línguas românicas, como CARPA 'carpa' e COFEA 'coifa'. A estes vieram juntar-se mais tarde, na época da dominação dos visigodos, alguns outros germanismos, oriundos do idioma deste povo, do ramo oriental. Os exemplos mais seguros de visigotismos do português são: *aleive, bando, espeto, espora, espia, escanção, luva, roca, ufano, arrear*, ou seja, vocábulos referentes essencialmente a actividades militares, conceitos jurídicos e objectos caseiros, assim como dois nomes de aves (manifestamente de agouro): *mejengra* 'chapim' e *laverca* 'cotovia'. Além destas palavras, directamente perfilhadas, existem ainda germanismos «de segunda mão», ou seja, galicismos de origem frâncica, que, na época do feudalismo, invadiram o léxico peninsular: *bastir, elmo, estala, guerra; guardar, guarnecer, jardim, rico, sala, trégua, venda* 'fita', etc. Há casos em que é difícil, se não impossível, decidir se estamos em Presença de um visigotismo ou de um elemento frâncico, trazido simultaneamente com outros galicismos.

Aluvião lexical árabe

12. Entre os elementos que, nas épocas obscuras em que se foi constituindo o Idioma, vieram avolumar o património vocabular latino, não há nenhuns que, quantitativa e qualitativamente, se possam comparar com aqueles, cuja aceitação se deve ao convívio e relações culturais das populações hispânicas com as muçulmanas. O «**superstrato**» árabe revela-se, com efeito, incomparavelmente mais avultado que o visigodo, pois abrange praticamente todos os sectores da vida material. Contam-se por várias centenas os vocábulos árabes, comuns, regionais ou antigos, que o português, antes ainda de merecer este nome, fez seus, adaptando na medida do possível os sons da língua semita ao sistema fonológico próprio. Por muito importante que seja, esta contribuição limita-se, na verdade, quase exclusivamente, a substantivos, sendo virtualmente inexistentes expressões respeitantes a qualidades morais e outras noções abstractas. De um ponto de vista formal, convém lembrar que a grande maioria dos arabismos assimilados apresenta a aglutinação do artigo semita *al*, cuja função gramatical não teria sido sentida quando os vocábulos em causa foram adoptados e adaptados, evidentemente por via oral, pela gente de fala hispânica. Daí o artigo pleonástico: *o alguidar, a alface*, etc.

13. No que toca à **significação** dos arabismos do português, são as seguintes as categorias semânticas, em que se integram principalmente: 1º) designações de cargos e dignidades: *alcaide, alferes, almoxarife*; 2º) termos castrenses: *arraial, arrebate, alcácer, alcáçova, atalaia*; 3º) de administração: *aldeia, arrabalde, alfoz, alfândega, alvará, almoeda*; 4º) de plantas cultivadas e silvestres: *arroz, algodão, alcachofra, cenoura, laranja, açúcar, alfarroba, alecrim, açucena, alfazema*; 5º) de profissões e indústrias: *alfaiate, alveitar, almocreve, alvanel, algoz, azenha, atafona, adobe*; 6º) de unidades de medida: *almude, arrátel, alqueire, arroba*; 7º) de animais: *atum, alcatraz, alforreca, alacrau, javali*; 8º) de particularidades topográficas: *albufeira, alverca, algar, lezíria, Recife*; 9º) de artigos de luxo e instrumentos de música: *almofada, alcatifa, marfim, alfinete, adufe, rabeca, anafil, alaúde*; 10º) de produtos agrícolas e industriais: *azeite, álcool, alcatrão*; 11º) da vida pastoril: *zagal, alfeire, rês, tabefe, almece*; 12º) de arquitectura: *aljube, chafariz, açoteia, alvenaria*; 13º) das ciências exactas: *algarismo, álgebra, cifra, auge*, etc.

14. São extremamente raros os adjetivos: *azul, zarco, forro*. Quanto a **outras categorias nocionais**, podem apontar-se: um verbo, *afagar*, um pronome indefinido, *fulano*; uma preposição, *até*, e uma interjeição, *oxalá*. Nalguns casos, o arabismo

apresenta-se com um matiz semântico superior ao de um termo vernáculo sinónimo: *alcatifa e aljôfar*, em relação a *tapete e pérola*, respectivamente.

Noutros casos observa-se uma depreciação do sentido primário, como sucede com o próprio nome dos árabes: *alarve*, e o da sua fala: *algaravia*. O número de arabismos foi antigamente consideravelmente maior do que hoje, e muitos dos que figuram nos dicionários actuais têm um cunho pronunciadamente literário-histórico, regional ou profissional. Por arabismos devem entender-se também vocábulos originariamente não árabes (orientais, berberes, etc.), incorporados no léxico dos muçulmanos peninsulares.

Latim medieval, humanista e científico-erudito

15. Uma vez constituído, ainda em tempos proto-históricos e com base numa massa lexical oralmente seleccionada - à qual se poderia chamar a «armação» do vocabulário português, de índole preponderantemente material-concreto - o **latim**, que na sua forma **medieval** continua a ser escrito e mesmo falado nos meios mais ou menos instruídos, vem trazer-lhe novas achegas. Língua não só da liturgia, mas também das escolas conventuais e das chancelarias, a sua acção vai crescendo com o incremento da cultura geral. Formas vindas daqueles ambientes aumentam o caudal das vozes alatinadas já anteriormente integradas no léxico comum. Os latinismos mais antigos mostram em geral uma adaptação maior, em relação aos posteriores, à fonética do romance, sendo costume qualificá-los de termos «semicultos». Trata-se de palavras como *virgem, anjo, diabo, cabido, regra, reino*, etc., cuja feição se aproxima bastante da dos elementos genuínos, quer dizer, dos que nunca deixaram de fazer parte do vocabulário patrimonial de base.

16. Ao lado destes latinismos antigos e fortemente assimilados, existe uma infinidade de outros, recebidos na época do **Humanismo e Renascimento**, que foram o maior veículo da propagação de elementos eruditos. Nesta classe de latinismos, a sua adaptação, superficial, limita-se em geral à terminação, p. ex., em *angélico*, o que permite reconhecer facilmente o seu carácter adventício. Desde então, a acção do latim, a que se vem juntar a do grego, nunca mais deixou de fazer sentir-se até aos nossos dias, sendo porventura a língua portuguesa entre as românicas a que maior rendimento tirou e continua a tirar deste inesgotável manancial que são as línguas clássicas. É também surpreendente a facilidade com que certos latinismos, pertencentes, a princípio, a determinados meios intelectuais e eruditos, conseguiram propagar-se não só na língua comum, como até na pronunciadamente popular, onde se usam hoje em pé de igualdade com os elementos lexicais do fundo primitivo, avoengo.

Elementos franceses e provençais

17. As influências lexicais do **francês**, avultadíssimas e constantes, acompanham, como os latinismos, toda a história do português, desde os primeiros contactos dinásticos, culturais e literários (séc. XII) até à época moderna. À acção da Ordem de Cluny e dos colonos «francos» devem-se, p. ex., *granja e rua*, como *monge e freire*; aos romances de cavalaria e às instituições feudais de além-Pirinéus, *torneio, arauto, duque, dama, palafrém, vergel, batalha, jaula, proeza, ligeiro, cobarde, coragem, linhagem, vilanagem, mensagem, viagem*, etc.; à lírica dos **provençais**: *trovar, trovador, refrão, segrel, jogral*. Em épocas sucessivas, a moda francesa reflecte-se em termos de indumentária: *boné, chapéu, colete, blusa, paletó*; assim como de culinária: *croquete, filete, fricassé, puré e restaurante*. Não faltam termos de música: *charamela, fabordão* (cf. *forrobodó*) e *oboé*. Existem outros sectores, em que certos galicismos se aninharam definitivamente, como no das comunicações: *comboio, trem, cais, bilhete*, etc. Se certos vocábulos franceses, que a língua acolhera, foram eliminados com o tempo (*reproche*,

afares), outros, principalmente no campo das noções técnicas (automóvel, maquinaria, etc.), encontram-se tão solidamente estabelecidos no uso geral que dificilmente os puristas conseguiriam extirpá-los. Diga-se ainda que boa parte dos componentes eruditos do vocabulário (latinismos e grecismos) foram adoptados primeiro pelo francês, chegando só através deste canal ao português.

Elementos exóticos

18. Os descobrimentos e a expansão dos Portugueses, ao revelar-lhes mundos novos com povos, civilizações e idiomas até então ignorados, haviam fatalmente de repercutir-se no vocabulário. São, com efeito, bastantes as **vozes exóticas** que, a partir do séc. XVI, se vão infiltrando no léxico português, seja por via literária (cronistas), seja por via oral (convivência com povos indígenas; comércio), aumentando assim consideravelmente o caudal de termos orientais recebidos anteriormente por intermédio do árabe. Alguns foram totalmente assimilados, outros não perderam o seu ar e feição estranhos, outros, finalmente, não passam hoje de curiosidades enterradas nos dicionários. Não se podem enumerar aqui todos os idiomas de África, Ásia e América que, em períodos e através de caminhos diversos, forneceram termos especiais ao léxico português. A Índia contribuiu, p. ex., com *bengala, andor, pagode, chita, xaile*; a China com *chá e chávena*; o Japão com *biombo e leque*; a África com *batuque e soba, ananás e inhame*; o Brasil com *mandioca e tapioca*; as Antilhas com *batata, cacau e tabaco*.

Elementos italianos, espanhóis e outros.

19. A contribuição do **italiano** manifesta-se principalmente nas terminologias artísticas. Em termos de música: *piano, contralto, ópera*; de arquitectura: *balcão, fachada*; de pintura: *aguarela, pitoresco*, de literatura: *soneto, burlesco*; de conceitos militares: *brigada, infantaria, escopeta*; de comércio: *banco, fiasco*; de indumentária: *calção*. Trata-se, de um modo geral, de termos de civilização que, na época do Renascimento, em que a Itália desempenhou o papel cultural-histórico que se sabe, se propagaram a quase todas as línguas românicas e outras europeias. Nem todos os italianismos do português foram importados directamente, tendo às vezes servido de intermediários o espanhol ou o francês.

20. Além dos italianismos, é considerável, embora em grau menor, o volume das contribuições do **espanhol**, que essencialmente datam da época do predomínio político e literário da Espanha. Alguns dos termos em causa situam-se no ambiente cortesão: *cavalheiro, lhano, airoso*; outros referem-se a noções militares: *cabecilha, caudilho, guerrilha*; outros à terminologia taurina: *ganadaria, bandarilha, muleta*; outros a costumes e vestuário tipicamente espanhóis: *tertúlia, chiste, boina, mantilha*, abstraindo de muitos outros vocábulos, pertencentes a campos semânticos diversos, como *faina, trecho, tijolo, moçoila, hediondo, moreno* e o próprio étnico. *castelhano*, que usurpou o lugar do ant. *castelão*. Diga-se ainda que existem não poucos castelhanismos perfeitamente integrados na fonética do português, e daí difíceis de identificar.

21. O que o português deve a outras línguas europeias (germanismos, nordismos, eslavismos, etc.) pouco pesa na balança dos empréstimos, podendo ser desprezado na presente apreciação, forçosamente sumária, dos elementos estrangeiros do léxico português. No entanto, mereceriam ser postas em relevo algumas palavras de origem inglesa, como *bife, clube, lanche, pudim*, etc., com outras de origem neerlandesa, p. ex., *amarrar, bombordo, bacalhau*, etc.