

A FORMAÇÃO DO LÉXICO ROMÂNICO E O MITO DA ORIGEM HÍBRIDA DO INGLÊS

Miguel Afonso Linhares (IFRR)
afonsolinhaires@hotmail.com

1. Introdução

O estudo de metaplasmos tem sido uma das atividades fundamentais da linguística histórica desde a própria consolidação do método histórico-comparativo no século XIX. Embora tenha perdido a primazia dentro da linguística para outros enfoques, este estudo continua a ser muito útil para conhecer a formação de uma língua, e, por conseguinte, a sua classificação quanto à origem.

Para tanto, o estudo do léxico pode ser ao mesmo tempo ponto de partida e de chegada da pesquisa. É ponto de partida porque não se estudam os metaplasmos senão nas palavras e é de chegada porque a conformação das palavras assinala a origem da língua.

Como se exporá mais adiante, é comum a dúvida sobre a classificação da língua inglesa quanto à sua origem, se é mesmo uma língua germânica, se pode ser considerada uma língua românica (ou, como se costuma dizer, “latina”) ou, em um meio termo, se é uma língua híbrida, que combina ambas as origens.

O modo mais difundido de estudo da questão tem sido a quantificação da etimologia com base nas entradas de dicionários, o que gera diferentes cifras. A interpretação destes dados é que tem dado azo ao lugar-comum que precede a dúvida sobre a classificação do inglês: a maior parte do léxico inglês tem origem latina.

Embora pareça contraditório à primeira vista, o estudo da formação do léxico das línguas românicas pode lançar luz sobre esta questão de um ponto de vista diferente. Ora, se se diz que a maior parte do léxico tem origem latina e se isto causa dúvidas em relação à classificação da língua, por que não se confrontam a formação do léxico românico e esta parcela do léxico inglês?

Este artigo tem o fim de gerar um entendimento da configuração híbrida do léxico inglês precisamente a partir das várias camadas

de palavras de origem latina de que se compõe o léxico das línguas românicas. Estas camadas são reconhecíveis justamente a partir dos metaplasmos que se observam nas palavras. Este estudo demonstra que o que se costuma julgar indiscriminadamente léxico de origem latina apresenta, na verdade, uma consistência bem mais complexa, do que se conclui que a proposição de que o inglês tem uma origem híbrida é um mito que distancia a língua e, por conseguinte, a cultura das suas raízes germânicas.

O artigo está repartido em cinco seções: na introdução estamos apresentando-o, à qual seguem exposições sobre a formação dos léxicos latino, românico e inglês; nas considerações finais conclui-se a arguição contra o mito da origem híbrida do inglês que deriva do exposto na segunda e terceira seções e se desenvolve na quarta, e que consiste no confronto entre os fatos expostos acerca da formação do léxico latino e românico e textos publicados na Internet.

2. O léxico latino

O latim é uma língua indo-europeia e, como tal, tem suas palavras mais comuns compartilhadas com a maioria das línguas desta família, isto é, palavras que derivam de uma raiz proto-indo-europeia hipotética. Pode-se ilustrar isto com partes do corpo humano como o dente: lat. *dens*, gr. ὀδών¹, sânscr. दत् (dat); o joelho: lat. *genu*, gr. γόνυ, sânscr. जानु (jánu); o pé: lat. *pes*, gr. πούς, sânscr. पद् (pad); relações de parentesco como mãe: lat. *mater*, gr. μήτηρ, sânscr. मातृ (mātr); pai: lat. *pater*, gr. πατήρ, पितृ (pitṛ); irmão: lat. *frater*, gr. φράτηρ “confrade”, sânscr. भ्रातृ (bhrātṛ); elementos naturais como a luz: lat. *lux*, gr. λευκός “branco”, रोचते (rócate “luzir”); a noite: lat. *nox*, gr. νύξ, sânscr. नक्ति (nákti); o sol: lat. *sol*, gr. ἥλιος, सूर्य (súrya); animais como o cavalo: lat. *equus*, gr. ἵππος, sânscr. अश्व (ásva); a ovelha: lat. *ouis*, gr. οἶς, sânscr. अवि (ávi); o urso: lat. *ursus*, gr. ἄρκτος, sânscr. ऋक्ष (ṛkṣá); elementos vitais como a casa: lat. *domus*,

¹ O conhecimento dos valores das letras gregas é bastante estendido entre o público deste artigo, de modo que opta por não as transliterar.

gr. *δόμος*, sânschr. *दम* (dáma); a lâ: lat. *lana*, gr. *λῆνος*, sânschr. *ऊर्ण* (úrṇā); o jugo: lat. *iugum*, gr. *ζυγόν*, sânschr. *युग* (yugá); adjetivos como direito: lat. *dexter*, gr. *δεξιός*, sânschr. *दक्षिण* (dákṣiṇa); novo: lat. *nouus*, gr. *νέος*, sânschr. *नव* (náva); vermelho: lat. *ruber*, gr. *έρυθρός*, *रुधिर* (rudhirá); vários verbos como viver: lat. *uiuo*, gr. *βίος* “vida”, sânschr. *जीवति* (jívati); ver: lat. *uideo*, gr. *εἶδω*, sânschr. *वेत्ति* (vétti); sentar: lat. *sedeo*, gr. *ἕζομαι*, sânschr. *सीदामि* (sídami); comer: lat. *edo*, gr. *ἔδω*, sânschr. *अत्ति* (átti); beber: lat. *poto*, gr. *πίνω*, sânschr. *पिबति* (píbati); conhecer: lat. *nosco*, gr. *γινώσκω*, sânschr. *जानाति* (jánāti); cozer: lat. *coquo*, gr. *πέσσω*, sânschr. *पचति* (pácati); tecer: lat. *texo*, gr. *τέχνη* “arte”, sânschr. *तक्षति* (tákṣati); levar: lat. *fero*, gr. *φέρω*, sânschr. *भरति* (bhárati); dar: lat. *do*, gr. *δίδωμι*, sânschr. *ददाति* (dádāti); seguir: lat. *sequo*, gr. *ἕπομαι*, sânschr. *सचते* (sácate); um pronome como eu: lat. *ego*, gr. *ἐγώ*, sânschr. *अहम्* (ahám); um advérbio como ontem: lat. *heri*, gr. *χθες*, sânschr. *ह्यस्* (hyas); uma preposição como e: lat. *-que*, gr. *τε*, sânschr. *च* (ca).

Como argüi Devoto (1940, p. 17), dentro da família indo-europeia o latim se configura pelo seu léxico como uma língua periférica, juízo amparado em várias palavras cognatas com outras sâns-critas, no outro extremo do domínio indo-europeu antigo, mas sem correspondente grego, que ocupa uma posição central. Podem-se citar palavras como o varão: lat. *uir*, sânschr. *वीर* (vīrá); o jovem: lat. *iuuenis*, sânschr. *युवन्* (yúvan); a boca: lat. *os*, sânschr. *आस्* (ās); o nariz: lat. *nasus*, sânschr. *नासा* (nāsā); a língua: lat. *lingua*, sânschr. *जिह्वा* (jihvā); o traseiro: lat. *clunis*, sânschr. *श्रोणि* (śróṇi); um deus: lat. *deus*, sânschr. *देव* (devá); um rei: lat. *rex*, sânschr. *राज्* (rāj); o fogo: lat. *ignis*, sânschr. *अग्नि* (agní); o dia: lat. *dies*, sânschr. *दिन* (diná); o freixo: lat. *fraxinus*, sânschr. *भूर्ज* (bhūrja “bétula”); o carvalho: lat. *quercus*, sânschr. *पर्कटी* (parkaṭī “figueira”); o ninho: lat. *nidus*, sânschr. *नीड* (nīdá); a roda: lat. *rota*, sânschr. *रथ* (rátha “carro”); o eixo: lat. *axis*, sânschr. *अक्ष* (ákṣa); estreito: lat. *angustus*, sânschr. *अंह* (amhú); devasso: lat. *lasciuus*, sânschr. *लषति* (lāṣati “desejar”); temer: lat. *temeo*, sânschr. *तमस्* (támas “trevas”); esquentar: lat. *tepeo*, sânschr. *तपति* (tápiti); vi-

ger: lat. *uegeo*, sânschr. *वज्ज* (*vāja* “vigor”); além: lat. *trans*, sânschr. *तिरस्* (*tirás*).

Depois, abstraindo o grego e o sânscrito, descobre-se a relação entre o léxico latino e o germânico, que caracteriza o latim como língua ocidental dentro da família indo-europeia (cf. id. *ib.*, p. 25): lat. *mentum* “queixo”, ingl. *mouth* “boca”; lat. *hostis* “inimigo”, ingl. *guest* “hóspede”; lat. *gelu* “gelo”, ingl. *cold* “frio”; lat. *faba* “fava”, ingl. *bean* “feijão”; lat. *far* “trigo”, ingl. *barley* “centeio”; lat. *granum* “grão”, ingl. *corn* m.s.; lat. *hordeum* “cevada”, ingl. *gorse* “tojo”; lat. *gallus* “galo”, ingl. *call* “chamar”; lat. *haedus* “cabrito”, ingl. *goat* “cabra”; lat. *piscis* “peixe”, ingl. *fish* m.s.; lat. *porcus* “porco”, ingl. *farrow* “leitão”; lat. *arcum* “arco”, ingl. *arrow* “flecha”; lat. *sordes* “sujeira”, ingl. *swart* “preto”; lat. *cano* “cantar”, ingl. *hen* “galinha”; lat. *capio* “tomar”, ingl. *have* “ter”; lat. *duco* “puxar, conduzir”, ingl. *tow* “puxar”; lat. *sagio* “farejar”, ingl. *seek* “procurar”; lat. *seco* “cortar”, ingl. *saw* m.s.; lat. *tongeo* “saber”, ingl. *think* “pensar”.

Enfim, como é natural, os latinos tomaram palavras dos povos com os quais interagiram, a começar pelos demais itálicos, cujas palavras são reconhecíveis por metaplasmos alheios à mudança do protoindo-europeu para o latim, por exemplo, em *bos* “boi” ou *lupus* “lobo”, cujas formas ancestrais, **g^wou-* e **w^lk^wos*, teriam dado **uos* e **luquus* em latim, ou na vacilação entre *sibilare* com /b/ latino e *sifilare* com /f/ osco-umbro. Por outro lado, é sabido que estes povos chegaram à Itália por migração, o que comportou o empréstimo de palavras dos povos que já habitavam a península. Ainda seguindo Devoto (*ib.*, p. 50-54), sobressaem-se os nomes de elementos da flora e da fauna: *camox* “camurça”, *cortex* “cortiça”; *cupressus* “cipreste”, *feles* “gato”, *figus* “figo”, *fungus* “cogumelo”, *hirundo* “andorinha”, *ibex* “cabra montês”, *larix* “lariço”, *laurus* “loureiro”, *lilium* “lírio”, *papauer* “dormideira”, *papilio* “borboleta”, *passer* “pardal”, *rosa* “rosa”, *talpa* “toupeira”; também há referentes à sociedade, como *famulus* “servo”, *mulier* “mulher”, *plebs* “plebe”, *urbs* “cidade”; verbos como *cupio* e *opto* “desejar”, *licet* “é permitido” e *oportet* “é preciso”, *loquor* “falar”, *iuvoo* “ajudar”.

Mas, sem sombra de dúvida foram os gregos o povo cuja língua influiu mais na latina; se bem que esta influência começa bem

antes de os romanos conquistarem a Grécia estrita. Já nos tempos do Reino entraram palavras gregas no latim, algumas pelas colônias da Magna Grécia, como *ἔλαιον* > *oleum* “azeite, óleo”, *καμάρα* > *camera* “abóbada”, *κέλις* > *celox* “navio leve”, *κραπάλη* > *crapula* “embriaguez”, *κυβερνῶ* > *gubernō* “pilotar”, *μῆλον* > *malum* “maçã”, *μηχανή* > *machina* “máquina”, *στόμαχος* > *stomachus* “estômago”, e há o curioso caso de *πρόσωπον* “face, máscara”, que chega ao latim, *persona*, por meio do etrusco, *phersu*. Dos tempos da República há de se ressaltar *βαλανεῖον* > *balneum* “banho”, *δραχμή* > *dracuma*, *κάλαμος* > *calamus* “cálamo”, *λαμπτήρ* > *lanterna* “lanterna”, *θέατρον* > *theatrum* “teatro”. Perceba-se que estes poucos helenismos constituem verdadeiros emblemas da cultura romana: o azeite, os banhos, o teatro e um dos instrumentos, o cálamo, pelo qual parte desta cultura nos foi legada, através da escrita, o que sugere a profundidade da influência grega (cf. id. ib., p. 88-92).

A influência grega se torna ainda mais sensível na literatura da Idade de Ouro. Devoto (ib., p. 225) a expõe nestes termos:

Si sono fatti calcoli per determinare la percentuale dei grecismi negli autori di questo tempo: 10% di parole greche presso Catullo, qualche cosa di più presso Ovidio e Tibullo, 11% nelle *Satire* e nelle *Epistole* di Orazio, 12% in Properzio, 14% nelle *Bucoliche* di Virgilio e quindi, discendendo nel tempo, 15% in Giovenale e 20% in Persio (Marouzeau, op.cit., p. 163). Ma queste statistiche dicono poco: il valore del numero è pieno quando si confrontano forme del tutto equivalenti come le percentuali di *forem* e di *essem*, non quando si contano elementi di un lessico che difficilmente si possono inquadrare in unità primordiali e che dipendono dal contenuto delle singole opere. La penetrazione del grecismo in questo tempo si pesa invece efficacemente attraverso la fortuna di caratteri greci isolati.

De fato, voltando a observar os helenismos arrolados percebe-se que em geral houve adaptação à fonologia do latim, o que sugere que a incorporação se deu pelo uso comum, enquanto que os helenismos incorporados desde a Idade de Ouro não só conservam fonemas gregos estranhos ao latim (o que, a propósito, demandou o acréscimo das letras y e z ao alfabeto latino), mas em parte também a morfologia, e é por isto que as gramáticas latinas contemporâneas trazem junto aos modelos comuns de declinação os de palavras de origem grega. A palavra grega *δραχμή*, citada, é exemplo disto: a sua forma antiga, *dracuma*, mostra a perda da aspiração e uma epêntese para adaptar tanto o fonema /k^h/ como um segmento /km/ que seria

igualmente estranho, porém na época clássica se fixa a forma *drachma*, atestadas inclusive flexões helenizantes, como o genitivo *drachmum*, calcado sobre *δραχμῶν*, o que sugere um grau alto de afeição à cultura grega.

Porém, nem sempre se lançou mão do empréstimo. Como é natural, também se aproveitou a capacidade de a língua gerar palavras pelos seus próprios elementos. Assim, do particípio presente de *esse* “ser” mais o sufixo substantivador *-ia* se deriva a palavra *essentia* “essência”, correspondente ao grego *οὐσία*; da raiz de *homo* “homem” mais o sufixo adjetivador *-an(us)* se deriva *humanus* “humano”, correspondente a *φιλάνθρωπος*; da raiz do supino de *nascor* “nascer” mais o sufixo substantivador *-ur(a)* se deriva *natura* “natureza”, correspondente a *φύσις*; da raiz do supino de *reor* “conto, penso” mais o sufixo substantivador *-io* se deriva *ratio* “razão”, correspondente a *λόγος*; o particípio passado de *cado* “cair”, *casus*, tomado como substantivo tanto com o significado primitivo de *queda* com o metafórico de *caso*, correspondentes a *πτῶσις*. Por outro lado, destas palavras derivam-se outras: mais o sufixo *-al(is)* deriva-se o respectivo adjetivo, ou seja, *essentialis* “essencial”, *naturalis* “natural”, *rationalis* “racional”, *casualis* “casual”, e do mesmo modo como de *humanus* se deriva o substantivo *humanitas* “humanidade”, de *naturalis* pode-se voltar a um substantivo: *naturalitas* “naturalidade”.

3. O léxico românico

A maior parte do léxico das línguas românicas, de todas elas, tem origem no léxico latino. Observa-se isto em palavras panromânicas derivadas de palavras latinas de origem vária:

dens,-tis > port. *dente*, cast. *diente*, cat. *dent*, fr. *dent*, it. *dente*, rom. *dinte*

lingua > port. *língua*, cast. *lengua*, cat. *llengua*, fr. *langue*, it. *lingua*, rom. *limbă*

nox,-ctis > port. *noite*, cast. *noche*, cat. *nit*, fr. *nuit*, it. *notte*, rom. *noapte*

piscis > port. *peixe*, cast. *pez*, cat. *peix*, fr. *poisson*, it. *pesce*, rom. *pește*

lupus > port. *lobo*, cast. *lobo*, cat. *llop*, fr. *loup*, it. *lupo*, rom. *lup*

iugum > port. *jugo*, cast. *yugo*, cat. *jou*, fr. *joug*, it. *giogo*, rom. *jug*

balneum > port. *banho*, cast. *baño*, cat. *bany*, fr. *bain*, it. *bagno*, rom. *baie*

nouus > port. *novo*, cast. *nuevo*, cat. *nou*, fr. *neuf*, it. *nuovo*, rom. *nou*

uiuere > port. *viver*, cast. *vivir*, cat. *viure*, fr. *vivre*, it. *vivere*, rom. *vive*

ego > port. *eu*, cast. *yo*, cat. *jo*, fr. *je*, it. *io*, rom. *eu*

Evidentemente, nem todas as palavras latinas passaram às línguas românicas, nem passaram por igual. Assim como o léxico latino é resultado das vicissitudes do povo romano, o léxico românico deixa ver os traços da formação dos povos que as falam.

Ora, como as línguas românicas são a continuação de dialetos do latim, é natural que em umas esteja presente certo étimo e em outra, outro, por exemplo, a forma simples *edere* cedeu à derivada *comedere*, com o prefixo *com-* “companhia”, que ficou em português, *comer*, e castelhano, *comer*, enquanto que nas demais línguas ficaram continuações de *manducare* “mastigar”: cat. *menjar*, fr. *manger*, it. *mangiare* (prov. do norte), rom. *mânca*. O mesmo quanto a port. *irmão*, cast. *hermano*, cat. *germà*, de um lado, de *germanus* “irmão cujos pais são os mesmos”, e fr. *frère*, it. *fratello*, rom. *frate*, de outro, de *frater* “irmão cujos pais não são necessariamente os mesmos”; também quanto a port. *maçã*, cast. *manzana*, de um lado, de *mattiana* “de Mattium (Germânia)”, e cat. *poma*, fr. *pomme*, de outro, de *poma*, por *pomum* “fruta”, e ainda it. *mela*, rom. *măr*, de *melum*, -a por *malum* “maçã”. E notável: a conservação tanto de *sibillare*, com /b/ latino, como de *sifilare*, com /f/ osco-umbro: do primeiro port. *assoviar*, cast. *silbar*, cat. *xiular*, rom. *șuiera*; do segundo cast. *chiflar*, fr. *siffler*, it. *zuffolare*.²

Ademais, do mesmo modo como o léxico latino contém palavras tomadas às línguas faladas no chão onde apareceu à luz da história antes das migrações indo-europeias e às línguas à volta de seu domínio territorial, o léxico românico também contém palavras das línguas faladas em seus territórios antes da conquista romana e com as quais os dialetos emergentes do latim conviveram antes de as substituir, bem como das línguas com que travaram contato.

² Os dados etimológicos das línguas românicas e do inglês estão fundados nos dicionários arrolados na bibliografia.

Destarte, da lista que se arrolou como palavras do patrimônio indo-europeu comum há a palavra *equus*, que não passou às línguas românicas (a não ser a forma feminina a algumas: *equa* > *égua*), que têm derivações de *caballus* (> *cavalo*), palavra de origem gaulesa que se integrou ao latim vulgar e daí a romances onde nunca houve presença céltica, como em sardo (> *caddu*) e em romeno (> *cal*).

Cada língua românica ou grupo de línguas românicas afins possui palavras de origem pré-romana que lhe são exclusivas. Assim, o lugar onde um enxame de abelhas se instala é uma *colmeia* em português e uma *colmena* em castelhano, palavra de origem pré-romana que se supõe céltica; um *rusc* em catalão e uma *ruche* em francês, palavra de origem gaulesa; em italiano é uma *arnia*, palavra também presente em castelhano e catalão, *arna*, e que tem origem incerta, pela sua extensão provavelmente céltica. Curiosamente, em romeno tem-se uma palavra de etimologia greco-latina bem segura: *stup*, de **stypus*, pelo grego *στύπος* “tronco” (o latim literário tem *stipes*, *-itis* m.s.). Na verdade, no caso do romeno é difícil averiguar étimos pré-romanos devido quer ao desconhecimento das línguas faladas na Dácia antes da conquista romana, quer à incerteza de se a língua romena surgiu mesmo neste território (e não ao sul), quer ao movimento intenso de povos nos Bálcãs desde a romanização.

Depois, assim como os falantes do latim tomaram palavras do povo que os conquistaram culturalmente, os gregos, os falantes do romance tomaram palavras dos povos que os conquistaram politicamente, os germanos no ocidente e os eslavos no oriente, e, depois, os árabes (e berberes) no ocidente e os turcos no oriente. Assim, das palavras que se citaram estudando-se a formação do léxico latino tem-se que *uir* foi substituído no ocidente pela palavra germânica latinizada *baro*, *-onis*, donde port. *barão*, cast. *barón*, cat. *baró*, fr. *baron*, it. *barone*, que do significado de “varão” (que é sua variante) passou a referir-se a um varão dignitário, e, daí, a designar um título de nobreza. Por outro lado, no oriente, que tem outra história, por *uir* diz-se *bārbat*, de *barbatus* “barbado”, enquanto que as dignidades feudais têm nomes de origem eslava, como *boier*, de *болярин* (*bojárin*) “nobre”, e *voievod*, de *voevoda* (*voevóda*) “príncipe”. Assim, no ocidente têm-se port. *ganhar*, cast. *ganar*, fr. *gagner*, it. *guadagnare*, de *waðanjan* “colher”; port. cast. cat. it. *guerra*, fr. *guerre*, de *werra* “peleja”; port. cast. cat. *guardar*, fr. *garder*, it. *guardare*, de

**wardōn* “vigiar”. No oriente, *câştiga*, do lat. *castigare* (e este de *castus* “casto”); *război*, do esl. *razbōu* (razbój) “matança”; *păstra*, do búlg. *пастря* (pastrjá) m.s.

Por outro lado, percebe-se que em castelhano e em português, os romances perdurantes mais expostos a influência da invasão muçulmana da península Ibérica, há ao lado da palavra derivada de *oleum*, “óleo” (> fr. *huile*, it. *olio*) uma de origem árabe, زيت (*zeyt*), donde port. *azeite* e cast. *aceite*, com o *a(l)*-típico das palavras desta origem, corresponde ao artigo definido. Já o romeno, deixa ver as imbricações características da história linguística dos Bálcãs: uma das formas, *olei* ou *ulei*, deriva do eslavo *олѣу* (*olěi*), e a outra, *olai* ou *ulai*, do húngaro *olaj*, e tanto a forma eslava como a húngara da latina *oleum*. A implicação do léxico ibero-românico com o árabe é ilustrada pelo fato de não haver em português e espanhol palavra de origem latina que designe a raiz *Daucus carota*, isto é, a *cenoura* em português e *zanahoria* em castelhano, do árabe اسفنارية (*isfanariyya*), de que também deriva a forma catalã usada em Valência, *safanòria*, enquanto que na Catalunha e nas Ilhas Baleares se tem uma derivação latina: *pastenaga* < *pastinaca*. Palavras de origem árabe de grande difusão são *مخازن* (*maḥāzin*) > port. *armazém*, cast. *almacén*, cat. *magatzem*, fr. *magasin*, it. *magazzino*, rom. *magazie* (< gr. mod. *μαγαζῖ*); *سكّر* (*sukkar*)³ > port. *açúcar*, cast. *azúcar*, it. *zuccher* (> fr. *sucrer*); *صفر* (*ṣifir*) “zero” > lat. med. *cifra* > port. cast. it. *cifra* (it. > rom. *cifră*), cat. *xifra*, fr. *chifre*.

Ainda tratando de empréstimos, cumpre observar que as línguas românicas influenciaram umas as outras, sobretudo o francês, expressão da cultura românica mais valorada, as demais: fr. ant. *joye* > port. cat. *joia*, cast. *joya*, it. *gioia*; fr. *courage* > port. *coragem*, cast. *coraje*, cat. *coratge*, it. *coraggio*, rom. *curaj*. Aliás, o francês exportou o sufixo *-ge* (de *-ticum*, que em português dá *-dego*, como em *achádego*, e *-zgo*, *hallazgo*; a forma catalã, *-tge*, é vernácula), de modo que de *armazém* se deriva *armazenagem*, de *almacén*, *almacenaje*, de *magatzem*, *magatzematge*, de *magasin*, *magasinage*, de *ma-*

³ Na verdade, esta palavra árabe tem origem persa, *شکر* (*šakar*), que a seu turno deriva do sânscrito *शर्करा* (*śárkarā*), que significa primitivamente “saibro”. A forma romena, *zahăr*, tem origem no grego moderno *ζάχαρη*, e este do grego antigo *σάκχαρη*, adaptação do persa.

gazzino, *magazzinaggio*, de *magazin*, *magazinaj*. Perceba-se que o empréstimo medieval o romeno não tem: “joia” é *giuvaier*, do turco *civahir*. O galicismo nas demais línguas românicas tem de fato uma dimensão insuspeita: palavras tão comuns quanto *controle* e *hotel* têm origem francesa, de *contrôle* (< *contre-rôle*) e *hôtel* (< fr. ant. *hostel* < lat. *hospitalis*, de *hospes* “hóspede”, de um antigo composto de *hostis* “forasteiro” e *potis* “amo”).

O italiano forneceu às demais línguas românicas sobretudo vocabulário referente à vida cortesã e militar nos séculos XVI e XVII, como a própria palavra *cortesão*, cast. *cortesano*, cat. *cortesà*, fr. *courtisan*, de *cortigiano*; *capriccio* > port. cast. *capricho*, cat. *capritx*, fr. *caprice*; *infanteria* > port. *infantaria*, cast. *infantería*, cat. *infanteria*, fr. *infanterie*. As formas romenas, por falta de contato com o resto da România até o século XIX, têm transmissão indireta: *curtizan* e *capriciu* através do francês e *infanterie* através do russo.

O castelhano e o português sobressaem por terem transmitido palavras das realidades que a Europa passou a conhecer a partir das grandes navegações, como cast. *batata*, de origem aruaque, donde port. *batata*; do cruzamento de *batata* e *papa*, de origem quíchua, surge a forma *patata*, que passa ao catalão e ao francês *patata* e ao francês *patate* (em romeno tem-se *cartof*, do al. *Kartuffel*, do it. *tartufolo*, diminutivo de *tartufo* “trufa”, do lat. *terrae tuber*, com /f/ osco-umbro); cast. *tomate*, do náhuatl *tomātl*, donde port. e fr. *tomate*, cat. *tomàquet*, it. do norte *tumata* (em italiano tem-se *pomodoro* “fruto d’ouro”, e em romeno, *roşie* “vermelha”; port. *crioulo* > cast. *criollo*, daí fr. *créole*, e deste it. *creolo* e rom. *creol*).

No entanto, o léxico das línguas românicas possui uma espécie de empréstimo muito peculiar devido ao fato de que a língua que lhes deu origem cessou para que existissem e, ao mesmo tempo, continuou a ser usada. A contradição é apenas aparente: como língua histórica, isto é, um diassistema coeso histórica e socialmente, o latim desde ao menos o século IX, ou seja, desde quando sabemos que havia consciência de que não se falava mais o latim, e sim a “rustica romana lingua”⁴. Por outro lado, a norma literária do latim continuou a ser usada de tal modo que somente após o século XII a maioria das

⁴ No Art. 17 dos cânones do Concílio de Tours, de 813 (cf. BASSETTO, 2005).

línguas românicas é escrita com alguma normalidade social. Consequência desta situação de diglossia durante tanto tempo é que o léxico das línguas românicas possui várias camadas de origem latina, a tal ponto que uma palavra latina pode ter assumido até três ou quatro formas em uma língua românica conforme a época em que voltou a ser usada. Este é o caso da derivação em português da palavra *macula*: o latim falado legou as formas *malha* e *mancha*, que pressupõem **mac'la* e **manc'la*; quando a síncope deixara de operar, a palavra voltou a ser usada, do que resulta a forma *mágoa*, que já tem um sentido metafórico; mais tarde foi retomada de novo, restringindo-se a mudança à adaptação da forma literária latina ao sistema fonológico, daí *mácula*, com um sentido ainda mais abstrato. Palavras como *mácula* são classificadas como cultismos e aquelas como *mágoa*, semicultismos (cf. id. ib., p. 171-176).

Todas as línguas românicas, salvo o romeno, possuem cultismos e semicultismos ao lado das formas herdadas do latim vulgar. O cultismo teve tal força na formação do léxico românico que não é raro descobrir que a forma padrão atual suplantou a forma vernácula ou semiculta ou a marginalizou como vulgarismo, como é o caso de *avondar*, refeito como *abundar*; *contrairo*, como *contrário*; *enfamar*, como *inflamar*; *esprementar*, como *experimentar*; *moimento*, como *monumento*; *ojeito*, como *objeto*; *reverença*, como *reverência*; *soterranho*, como *subterrâneo*⁵.

Cabe observar que esta latinização do romance varia conforme as diferentes fonologias. Assim, em uma língua como o catalão, em que ocorrem bastantes segmentos de consoantes oclusivas ou por síncope na mudança a partir do latim, como *dubitare* > *dubtar*, ou fonética sintática, *un cop d'ull* “olhada”, um cultismo como *efecte* apresenta esta forma desde os textos mais antigos; em castelhano, que possuía estes segmentos em sua fase antiga, como *dubitare* > ant. *dubdar* > mod. *dudar*, vacilava-se entre *efeto* e *efecto* até a Academia fixar a forma latinizante; em francês e português, cujas fonologias os acomodavam menos, *dubitare* > fr. *douter* e port. *duvidar*, têm-se o semicultismo *effet* e *efeito*, embora haja atestações medievais de *effect* e *efecto* e o adjetivo tenha forma culta: *effectif* e *efeti-*

⁵ As formas não padrão estão atestadas por Houaiss (2001).

vo; em italiano, que tem uma fonologia ainda mais restrita a isto tem-se a assimilação como regra: *effetto*, *effettivo*. O romeno constitui um caso bem específico, posto que seu léxico latino culto ou deriva do francês ou passou pelo seu crivo.

Enfim, é curioso e surpreendente como nas línguas românicas se forjaram palavras que têm um aspecto perfeito de palavras latinas ou gregas latinizadas, mas que ou não estão atestadas nem no léxico latino nem no grego ou não está atestado o significado que têm hoje. É que se trata de termos recentes com morfemas do latim ou do grego ou palavras destas línguas para designar objetos novos. Assim, não se tem notícia de uma palavra **ciuilizatio*, mas ela serviu de modelo para a palavra francesa *civilisation*, que, a seu turno, condiciona formações correspondentes: port. *civilização*, cast. *civilización*, it. *civilizzazione*, rom. *civilizație*. Por outro lado, a palavra *societas* está bem atestada em latim, porém o seu significado estava bem atrelado ao nome *socius* “sócio”, enquanto que nas línguas modernas port. *sociedade*, cast. *sociedad*, cat. *societat*, fr. *société*, it. *società*, rom. *societate* se somou um significado mais abstrato no qual a ligação com *sócio* é fraca.

Vejamos agora como a formação do léxico românico pode dissipar qualquer dúvida sobre a classificação genética do inglês.

4. O léxico inglês

Parte-se de uma afirmação cabal: o inglês é uma língua germânica. Faz-se, pois, uma análise dedutiva para comprovar isto.

A primeira evidência, a mais lógica, é que as mesmas dez palavras que demonstram a unidade das línguas românicas ilustram a pertença do inglês à família germânica. Confrontem-se o inglês e as três línguas germânicas mais faladas depois dele:

***tanþz** > ingl. *tooth*, hol. *tand*, al. *Zahn*, sueco *tand*

***tungôn** > ingl. *tongue*, hol. *tong*, al. *Zunge*, sueco *tunga*

***naht-** > ingl. *night*, hol. *nacht*, al. *Nacht*, sueco *natt*

***fiskaz** > ingl. *fish*, hol. *vis*, al. *Fisch*, sueco *fisk*

***wulfaz** > ingl. *wolf*, hol. *wolf*, al. *Wolf*, sueco *ulf*

***yukam** > ingl. *yoke*, hol. *juk*, al. *Joch*, sueco *ok*

***baþam** > ingl. *bath*, hol. *bat*, al. *Bad*, sueco *bad*

***newjaz** > ingl. *new*, hol. *nieuw*, al. *neu*, sueco *ny*

***libæ-** > ingl. *live*, hol. , al. *leben*, sueco *leva*

***ika** > ingl. *I*, hol. *ik*, al. *ich*, sueco *jag*

Todas estas palavras têm origem indo-europeia, mas percebe-se que o inglês, como língua germânica, herda do proto-germânico as mutações consonânticas que caracterizam esta família: a espirantização das oclusivas surdas (comparem-se *peixe*, de *piscis*, e *fish*, de **fiskaz*) e o ensurdecimento das oclusivas sonoras (comparem-se *ju-go*, de *iugum*, e *yoke*, de **yukam*).

De fato, das cem palavras que compõem a chamada lista Swadesh, que o linguista norte-americano Morris Swadesh concebeu como modo fiável de se classificarem geneticamente as línguas, dado que se trata de um vocabulário básico resistente ao empréstimo, só uma não tem origem germânica: *mountain*, do francês *montaigne*.

Apesar destas obviedades, vige certo senso comum que produz o juízo de que o inglês é uma língua “latina”. De fato, se se lançar no *site* de busca *Google* a frase “English comes from Latin” (com aspas, que garantem a busca da frase), ou seja, “o inglês vem do latim”, obtêm-se 8.210 ocorrências, o que dá uma ideia da repercussão que de alguma maneira gera este lugar-comum. Neste sentido, os usuários do fórum *How-to-learn-any-language* criaram um tópico intitulado *Popular misconceptions about languages* “Equívocos populares em relação a línguas”, em cuja sétima intervenção se lê o testemunho seguinte:

Regarding the main topic at hand, most people (at least in the anglo-sphere) are far too ignorant about languages to have any conceptions at all, including misconceptions. ‘English is a Romance language’ is a good example, though.⁶

Este testemunho recebe confirmação na intervenção 53, na qual outro usuário comenta: “I’ve also been told that English is a

⁶ “A respeito do tópico em questão, a maioria das pessoas (ao menos no mundo anglófono) é de tal modo ignorante em relação a línguas que tem quaisquer ideias, inclusive equívocos. ‘O inglês é uma língua românica’ é um bom exemplo, a propósito”.

Romance language...” (“também venho dizendo que o inglês é uma língua românica...”).

Neste sentido, como complemento fortuito destes testemunhos, um usuário do *site WordReference* abriu um tópico com a indagação “English isn’t a Romance language. Why?” (“O inglês não é uma língua românica. Por quê?”), após a qual segue esta inquietação:

I know it’s about history and classification and that English is classified as Germanic. I guess that will never change, however, other Germanic languages seem completely foreign! Swedish, Norwegian, Dutch & Icelandic are impossible for me to understand. However, if I look at French and Spanish it seems much easier for me, I can understand the gist just by learning a few rules. I mean think how similar English and Spanish are.⁷

É de se perguntar que semelhança é esta que mexe com o senso comum das pessoas ao encarar a língua inglesa. A resposta não é nada difícil de se encontrar: um usuário do portal Yahoo! do Reino Unido e da Irlanda lançou a pergunta na seção *Yahoo! Answers*: “Why so many people thinks English is a romance language?” (“Por que tanta gente acha que o inglês é uma língua românica?”), e uma das respostas bem votadas é esclarecedora:

*The reason is very simple. There has been a major Latin influence in English. Although English is Germanic in grammar, its vocabulary is mostly Italic. Sixty percent of the English vocabulary has its roots in Latin (although much of this is indirect, mostly via Anglo-Norman and French). In the medieval period, much of this borrowing occurred through ecclesiastical usage established by Saint Augustine of Canterbury in the 6th century, or indirectly after the Norman Conquest, through the Anglo-Norman language.*⁸

⁷ Eu sei que há mais ou menos uma história e uma classificação e que o inglês é classificado como germânico. Suspeito que isto nunca muda, no entanto as outras línguas germânicas parecem completamente estranhas! O sueco, o norueguês, o holandês e o islandês são para mim impossíveis de compreender. No entanto, quando olho o francês e o espanhol, parecem muito mais fáceis para mim: posso entender o essencial tendo aprendido só umas poucas regras. Quero dizer quão acho semelhantes o inglês e o espanhol.

⁸ A razão é bem simples. Houve uma maior influência latina no inglês. Embora o inglês seja germânico na gramática, o seu vocabulário é sobretudo itálico. Sessenta por cento do vocabulário inglês tem suas raízes no latim (embora muito delas sejam indiretas, sobretudo via anglo-normando e francês. No período medieval, muitos desses empréstimos se deram através do uso eclesiástico estabelecido por Santo Agostinho de

Esta intervenção no fórum mencionado contém dois aspectos chave da questão. No primeiro momento o autor argui que a língua inglesa não sofreu influência mais forte do que a latina; consequência disto a maior parte do léxico inglês tem origem “itálica”. De fato, observa-se que a maioria das proposições que contêm a frase “English comes from Latin” arroladas pelo Google gira em torno da quantificação da etimologia do léxico inglês:

45% of English comes from Latin so that makes them latin too. (na resposta a uma pergunta no site *Answers.com*, lançada pelo Google na primeira página da lista de resultados)

I believe that something like 60 to 70 percent of vocabulary in English comes from Latin. (na entrevista à professora Esther Guerrero, da University of California, no site *The Island of Alameda*, lançado pelo Google na terceira página da lista de resultados)

50% of English comes from Latin. (em um comentário ao blogue de Andrew Barlett, lançado pelo Google na quinta página da lista de resultados)

Some 80 % of English comes from Latin and Greek. (na seção *Tutoring* do site *Free advanced English*, lançado pelo Google na quinta página de resultados)

60% of English comes from Latin. (em um comentário ao blogue *Eugapae Latina*, lançado pelo Google na nona página de resultados)⁹

Vê-se que o lugar-comum de que o inglês é uma língua românica ou uma língua híbrida latino-germânica (veja-se outra pergunta feita no *Yahoo! Answers*: “Why isn’t English a mixed Germanic-Romance language?” (“Por que o inglês não é uma língua mista germano-românica?”) deriva de uma noção estendida de que o léxico inglês possui uma grande quantidade de palavras, de modo que a filiação da língua passa a ser considerada de um ponto de vista quantitativo, segundo as proporções das contribuições ao seu léxico, ao que cabe indagar: qual é a natureza do léxico “latino” do inglês?

Cantuária no século VI, ou indiretamente após a conquista normanda, por meio da língua anglo-normanda.

⁹ Respectivamente “45% do inglês vem do latim, o que também o torna latino”; “creio que em torno de 60 a 70 por cento do vocabulário em inglês em do latim”; “50% do inglês vem do latim”; “uns 80% do inglês vem do latim e do grego”; “60% do inglês vem do latim”.

Considerando que até agora se tem partido do discurso do senso comum, aproveita-se uma lista de palavras que o autor da última referência mencionada aporta em sua arguição, as quais julga que “are used naturally in colloquial speech” (“são usadas naturalmente na fala coloquial”):

*change - bottle - people - common - use - fame - quiet - silen(c)e - air - sure - front - just - actually - move - ridiculous - pain - line - decide - advice - face - really - annoying - pay - joke - realize - party - touch - visit - normal - idiot - bastard - place - (use) - arm - table - clear*¹⁰

Considerando então tudo o que se expôs sobre a formação do léxico latino e bem mais sobre a do léxico românico, não se pode evitar levantar algumas questões: sendo a palavra latina *cambiare*, por que em inglês se tem *change*, com os metaplasmos característicos do francês: (-)c(c)- > (-)ch- (que se confirma em *touch*, ainda que **toccare* seja hipotético); -b̄i- > -g^{e.i.}-, -j-? O mesmo em relação *sure*, com uma síncope característica do francês: *securus* > ant. *sëur* > mod. *sûr* (confirmada em *pay*: *pacare* > **pagar* > *payer*), ou *pain*, com a ditongação do *ē* longo vacilante entre *ei* e *oi*: *poena* (*pēna*) > *peine*, ou ainda *party*, com outra síncope caracteristicamente francesa: *partita* > *partie*.

Ora, os metaplasmos apenas refletem os fatos, a saber, das 35 palavras 26 têm clara origem francesa: *cambiare* > *changer* > *change*, tard. *boticula* > ant. *boteille* (mod. *bouteille*) > *bottle*, *populus* > anglo-normando *people* (mod. *peuple*) > *people*, *communis* > *commun* > *common*, **usare* > *user* > *use*, *fama* > *fame* > *fame*, *quietus* > ant. *quière* > *quiet*, *silentium* > *silence* > *silence*, *aer* > *air* > *air*, *securus* > ant. *sëur* (mod. *sûr*) > *sure*, *frons,-tis* > *front* > *front*, *iustus* > *juste* > *just*, *mouere* > anglo-normando *movir* (mod. *mouvoir*) > *move*, *poena* > *peine* > *pain*, *uisum* (mais a preposição *à*) *avis* > *advice*, *facia* > *face* > *face*, tard. *innodiare* > *ennuyer* > *annoy*, *pacare* > *payer* > *pay*, *partita* > *partie* > *party*, **toccare* > ant. *touchier* (mod. *toucher*) > *touch*, tard. *bastardus* > ant. *bastard* (mod. *bâtard*) > *bastard*, *platea* > *place* > *place*, *arma* > *arme* > *arm*, *tabula* > *ta-*

¹⁰ Mudança, garrafa, povo, comum, usar, fama, quieto, silêncio, ar, seguro, frente, justo/apenas, atualmente, mover, ridículo, dor, linho/linha, decidir, conselho, face, realmente, enojar, pagar, piada, realizar, festa/partido, tocar, visitar, normal, idiota, bastardo, lugar, arma, mesa, claro.

ble > *table*, *clarus* > ant. *cler* (mod. *clair*) > *clear*; em duas a ausência de metaplasmos característicos, por serem cultismos, impede dizer-se com certeza se são empréstimos ao francês ou ao latim: *decidere* > *décider* (>?), *decide* e *uisitare* > *visiter* (>?), *visit*; cinco são cultismos consistentes em palavras latinas ou gregas ou formadas de morfemas latinos ou gregos, das quais a forma francesa serve de condicionante: *actualis* (em latim clássico ligado diretamente a *actus* “ato” e nas línguas modernas mais abstrato) > *actual*, tard. *realis* (derivado de *res* “coisa”) > *real*, modelo *réaliser* (< **realizare*) > *realize*, *normalis* (em latim clássico atestado como ligado a *norma* “esquadra”, e nas línguas modernas a “regra”, e daí a “costume”), *idiota* (< gr. *ιδιώτης*) > *idiot*¹¹; somente duas palavras derivam diretamente do latim: *ridiculous* (mesmo assim leva o sufixo *-ous*, da forma anglo-normanda homomorfa (mod. *-eux*), e esta do latim *-osus* > port. *-oso*) de *ridiculus* e *joke* de *jocus* “jogo”; e, por fim, uma coincidência: o autor julgou que *line* tem origem no latim *linum* ou *linea* quando tem origem no proto-germânico **līnon*, portanto palavra do patrimônio indo-europeu.

O estudo dos metaplasmos que caracterizam a mudança do latim para as línguas românicas, especificamente a francesa, ilustra o que disse o autor da resposta à pergunta “Why so many people thinks English is a Romance language?”, feita no *Yahoo! Answers*: “Sixty percent of the English vocabulary has its roots in Latin (although much of this is indirect, mostly via Anglo-Norman and French)” (o sublinhado é nosso). O problema repousa sobre o fato de que a quantificação (que, aliás, varia) da etimologia está divulgada, porém não se leva em conta que cifras não dizem nada por si mesmas, mas demandam uma interpretação dentro de certo contexto. Destarte, o conjunto que se nomeia “palavras de origem latina” é, na verdade, um complexo vário, cujo único elemento compartilhado é uma referência de alguma espécie à língua latina, e, não sendo de uma, mas alguma espécie, compromete a precisão da qualidade “de origem latina”. Mais precisa, mas não de todo, é a quantificação informada no

¹¹ Estes (pseudo)latinismos não são característicos do inglês fora da família românica: o holandês tem *actueel* e o alemão e o sueco *aktuell*; o holandês tem *reëel* e o alemão *real*; o holandês tem *normaal* e o alemão e o sueco *normal*; o holandês tem *idiot* e o alemão e o sueco *idiot*.

portal dos dicionários Oxford, *Oxford Dictionaries*, com base em um estudo de Thomas Finkenstaedt e Dieter Wolff a partir das etimologias da terceira edição do *Shorter Oxford Dictionary*:

28,24% tem origem latina (incluem-se criações modernas calcadas sobre a morfologia do latim, características do discurso técnico e científico nas línguas europeias de modo geral);

28,3% tem origem francesa (incluem-se as palavras de origem anglo-normanda);

25% tem origem germânica (incluem-se as palavras de origem propriamente vernácula, ou seja, anglo-saxã, e empréstimos das demais línguas germânicas);

5,32% tem origem grega (desconhece-se o critério que distingue este conjunto e o de origem latina);

4,03% não tem origem certa;

3,28% deriva de nomes próprios;

1% tem origem em outras línguas.

O que distingue, pois, o inglês das demais línguas germânicas não é uma maior influência do latim, equívoco que o autor da equilibrada resposta à pergunta “Why so many people thinks English is a romance language?” no *Yahoo! Answers* não evitou, mas uma influência anormal da língua francesa. Chamamo-la de anormal em uma consideração de talhe sociolinguístico, posto que tal influência é consequência da invasão comandada por Guilherme II da Normandia, que despojou a aristocracia anglo-saxã e tomou a coroa da Inglaterra em 1066, a partir de quando o francês em sua variante normanda e depois o padrão emergente baseado no franciano¹² se tornou a língua culta do reino, gerando uma situação de diglossia. Esta situação perdeu até pelo menos 1362, quando o Parlamento autorizou o uso do inglês nos seus atos, mas como se expôs na seção sobre a formação do léxico românico, o galicismo teve um papel importante na formação do léxico das línguas europeias modernas, fornecendo palavras feitas ou modelos semânticos para a derivação com morfemas próprios.

¹² O dialeto da região da Ilha-de-França, onde fica Paris.

5. *Considerações finais*

As palavras não têm a sua história gravada apenas nos documentos em que foram escritas. A história de uma palavra é presumível também por sua própria forma, isto é, pelas transformações desde o seu étimo: os metaplasmos.

Na introdução a este artigo afirmou-se que há certo senso comum que julga que o inglês é uma língua românica ou uma língua mista latino-germânica, o que se comprovou na penúltima seção, em que se constatou que tal juízo se funda em diferentes quantificações da etimologia do léxico inglês, que indicam proporções altas de palavras de origem “latina”. Mostrou-se então que a história gravada nestas palavras contradiz este juízo, posto que os metaplasmos não remetem a um étimo latino, mas a um étimo românico, especificamente francês.

É claro que o inglês possui palavras tomadas diretamente do latim, mas curiosamente elas contradizem ainda mais o juízo de que houve uma influência maior e notável sobre o inglês, dado que estão presentes em outras línguas germânicas: *cuprum* “cobre” > ingl. *copper*, hol. *koper*, al. *Kupfer*, sueco *koppar*; *strata* “estrada” > ingl. *street*, hol. *straat*, al. *Straße*, os três significando “rua”; *discus* “disco” > ingl. *dish* “prato”, hol. *disch*, al. *Tische*, ambos significando “mesa” etc. Estas palavras passam despercebidas pelo senso comum porque remontam a interação entre romanos e germanos na Antiguidade, portanto antes da invasão germânica da Bretanha, e por isto sofreram metaplasmos que impedem a sua pronta identificação.

Destarte, mostrou-se que o estudo da mudança do latim para as línguas românicas ilustra com bastante precisão a natureza do léxico inglês de origem “latina”. Ora, assim como os falantes do romance sentiram a necessidade e por ela passaram a distinguir o que é latim (*latine*) e o que é romance (*romanice*), cumpre manter esta distinção, a distinção entre uma palavra propriamente latina, como *popular*, do latim *popularis*, e uma propriamente romance, como *people*, do francês anglo-normando *people*.

Se os discursos do senso comum sobre a classificação da língua inglesa quanto à origem não se verificam, qual será a sua motivação? Explicamo-lo por uma ideologia tão velha quanto o Ocidente:

a ideia do Império, absolutizado assim como a dignidade do imperador de Roma. A história desta ideia na política é bem conhecida e não cabe aqui. A história dela nos estudos das línguas foi mencionada indiretamente na seção sobre a formação do léxico românico: quando se latinizava uma palavra com forma popular como *avondar* – *abundar* ou semiculta como *contrairo* – *contrário*, os letrados não perseguiam, consciente ou inconscientemente, outro fim senão enaltecer, enobrecer, dignificar a língua vulgar aproximando-a da língua sábia. Cultivar o legado linguístico latino é legitimar o cultivo ou mesmo reivindicar o legado político romano.

BIBLIOGRAFIA CONSULTADA

ALCOVER, Antoni Maria; MOLL, Francesc de Borja. *Diccionari català-valencià-balear*. 2. ed. Barcelona: Institut d'Estudis Catalans, 2002. Disponível em: <http://dcvb.iecat.net>

BASSETTO, Bruno Fregni. *Elementos de filologia românica*. 2. ed. São Paulo: Edusp, 2005.

BOURCIEZ, Édouard. *Éléments de linguistique romane*. 5. ed. Paris: Klincksieck, 1967.

DEVOTO, Giacomo. *Storia della lingua di Roma*. Bolonha: Licino Cappelli, 1940.

GAFFIOT, Félix. *Dictionnaire latin-français*. Paris: Hachette, 1974. Disponível em: <http://www.lexilogos.com/latin/gaffiot.php?q>

HARPER, Douglas. *Online etymology dictionary*. Disponível em: <http://www.etymonline.com>

HOUAISS, Antônio *et al.* *Dicionário eletrônico Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2001.

PIAGENTINI, Ottorino. *Vocabolario etimologico della lingua italiana*. Roma: Albrighi & Segati, 1907. Disponível em: <http://www.etimo.it>

PIERREL, Jean-Marie (org.). *Trésor de la langue française informatisé*. Disponível em: <http://www.cnrtl.fr/etymologie>

REAL Academia Española. *Diccionario de la lengua española*. 22. ed. Madri: Real Academia Española. Disponível em: <http://buscon.rae.es/draeI>

TAGLAVINI, Carlo. *Le origini delle lingue neolatine*. 6. ed. Bologna: Pàtron, 1982.

Vários autores. *Dicționar explicativ al limbii române*. Disponível em: <http://dexonline.ro>

WIKIMEDIA Foundation. *Wiktionary, the free dictionary*. Disponível em: <http://www.wiktionary.org>