

Artigos

Processos de formação lexical das línguas românicas e germânicas: uma nova perspectiva teórica

Lexical formation processes of Romance and Germanic languages: a new theoretical perspective

Aldo Bizzocchi*

Resumo: Este artigo discute a dinâmica de formação das unidades que compõem o léxico, notadamente das línguas europeias ocidentais (românicas e germânicas), a partir da identificação dos diversos processos lexicogênicos existentes. Para tanto, reconhecem-se dois tipos básicos de processos: os autogenéticos (herança ou criação interna) e os alogenéticos (empréstimos), cujos subtipos são aqui descritos. Com base na etimologia dos elementos temáticos (radicais e afixos), os itens lexicais de línguas europeias costumam também ser divididos em cultos, semicultos e vulgares. Essa classificação decorre do reconhecimento da profunda influência cultural e, conseqüentemente, linguística que os idiomas literários da Antiguidade Ocidental (grego e latim clássicos) exerceram sobre as modernas línguas de cultura da Europa, dentre as quais o português. Esta pesquisa procura identificar, por meio de análise empírica sobre um *corpus* formado de textos de vários gêneros, os processos lexicogênicos presentes nas principais línguas românicas e germânicas, tanto aqueles já descritos anteriormente na literatura especializada quanto alguns novos, que os estudos anteriores ignoravam ou classificavam erroneamente em outras categorias.

Palavras-chave: léxico; etimologia; neologia; línguas românicas; línguas germânicas.

Abstract: This paper discusses the formative dynamics of the constituent units of the lexicon, notably in the Western European languages (Romance and Germanic), starting from the identification of the various existing lexicogenical processes. To do so, two basic types of processes are recognized: autogenetic (inheritance and internal creation) and allogetic (loans), whose subtypes are described herein. Based on the etymology of their thematic elements (radicals and affixes), lexical items of European languages also tend to be divided into cultivated, semicultivated, and vulgar. This classification stems from the recognition of the profound cultural and, consequently, linguistic influence that the ancient Western literary languages (classical Greek and Latin) brought about the modern languages of culture in Europe, including Portuguese. This research seeks to identify, through empirical analysis on a *corpus* composed of texts of various genres, the lexicogenical processes present in the main Romance and Germanic languages, both those previously described in the specialized literature and some new ones, which previous studies used to ignore or erroneously classify into other categories.

Keywords: lexicon; etymology; neology; Romance languages; Germanic languages.

* Pesquisador do NEHiLP-USP (Núcleo de Apoio à Pesquisa em Etimologia e História da Língua Portuguesa da Universidade de São Paulo) e do GMHP (Grupo de Morfologia Histórica do Português, cadastrado junto ao CNPq).

Introdução

O léxico de uma língua é uma entidade à qual só se pode, em grande medida, ter acesso por abstração. É impossível na prática listar todos os itens lexicais de uma língua, quer porque o léxico é dinâmico e a todo momento surgem novos itens, que só muito tempo depois virão – ou não – a ser catalogados em obras de referência, quer porque a maior parte dos itens é de uso restrito a determinadas áreas profissionais, o que dificulta sobremaneira o trabalho lexicográfico. Em todo caso, sabe-se que o léxico pode ser submetido a três níveis diferentes de análise, que segundo Coseriu (1979), são *sistema*, *norma* e *fala*, produzindo assim três recortes diferentes da realidade e revelando unidades lexicais que apenas aproximadamente se correspondem (BARBOSA, 1996): o lema (unidade de sistema, polissêmica, polissemêmica), o vocábulo ou *type* (unidade de norma, polissêmica, monossemêmica) e a palavra-ocorrência ou *token* (unidade de fala, monossêmica, monossemêmica).

Por consequência, o estudo etimológico do léxico produz resultados diferentes conforme aplicado a cada um dos níveis de análise acima, pois o mesmo lema (digamos, um verbete do dicionário) compõe-se de vários *types* (acepções) que, por sua vez, são atualizados em situações de discurso, isto é, atos de fala ou de escrita (ocorrências concretas da palavra no texto) que tendem ao infinito. Diferentes acepções de um lema podem apresentar etimologias diferentes (por exemplo, *estrela*₁, “corpo celeste”, é vocábulo herdado, mas *estrela*₂, “atriz famosa”, é decalque do inglês *star*).

Por outro lado, podem-se estudar os processos de formação do léxico de uma língua europeia sob duas perspectivas diferentes: a neológica e a etimológica. A primeira diz respeito ao modo como a palavra foi criada, e a segunda, à origem da matéria-prima com que foi criada.

Assim, do ponto de vista neológico, as palavras ou são hereditárias (e, portanto, nunca tiveram caráter neológico na língua, pois já formavam sua base lexical quando o idioma se individualizou) ou foram em algum momento neologismos. Estes se dividem em fonológicos, sintagmáticos, semânticos e alogenéticos (GUILBERT, 1975; BARBOSA, 1996; ALVES, 2007). Os neologismos fonológicos podem ser do tipo *ex nihilo* (criação de um novo significante a partir do nada: *popperô*) ou onomatopaicos (imitação do som do significado: *plim-plim*, *tilintar*). Os sintagmáticos englobam a composição e a derivação; os semânticos consistem na atribuição de novo significado a palavra já existente, e os alogenéticos resultam do empréstimo de palavra estrangeira.

Por outro lado, do ponto de vista etimológico, os vocábulos costumam ser classificados em cultos, ou eruditos (cultismos); semicultos, ou semieruditos (semicultismos); e vulgares, ou populares (vulgarismos), além dos inclassificáveis nas categorias anteriores.¹

Nas línguas europeias, cultismos são as palavras tomadas do grego, do latim ou resultantes da combinação de elementos greco-latinos. Vulgarismos são os vocábulos herdados, seus compostos e derivados, bem como os empréstimos de línguas que não o grego ou latim (estrangeirismos). Já os semicultismos apresentam hibridismo de elementos cultos e vulgares, seja por combinação ou por vulgarização da forma greco-latina. São, em princípio, inclassificáveis os neologismos fonológicos, as derivações de siglas (*ufologia, petista, aidético*) ou de nomes próprios (*amperímetro, kantiano*), bem como as constituídas de fragmentos de morfemas, ou fractomorfemas (*reprografia, informática, metrô, minissaia, showmício*).

Considerando-se ainda que uma palavra pode ter origem numa língua sem qualquer contribuição, direta ou indireta, de outra língua, ou pode, ao contrário, ter-se formado a partir de elementos mórficos externos ao sistema ou mesmo com elementos autóctones mas segundo um modelo estrutural estrangeiro, podemos reconhecer dois processos básicos de formação lexical: os autogenéticos (hereditariedade ou criação intralinguística) e os alogenéticos (empréstimo ou criação a partir de elementos alógenos). Pode-se dizer que são autogenéticas as palavras que não contenham nenhum elemento estrangeiro quer no plano da expressão quer no do conteúdo, e que são alogenéticas as que contenham pelo menos um elemento estrangeiro, seja ele de natureza morfológica ou semântica.

Podemos sistematizar esses conceitos da maneira a seguir.

1. Processos autogenéticos (autogenia):

- a. herança (léxico vernáculo);
- b. neologia fonológica;
- c. ressemantização de palavras autogenéticas;
- d. composição ou derivação a partir de palavras autogenéticas.

2. Processos alogenéticos (alogenia):

- a. empréstimo de palavra estrangeira;

¹ Cultismos, semicultismos e vulgarismos são categorias etimológicas e não sociolinguísticas; dizem respeito, portanto, à origem do vocábulo e não ao seu uso. Por sinal, há palavras herdadas, como *escorreito*, que só ocorrem no registro ultraformal, assim como há latinismos, como *operário*, que são de amplo uso em todos os níveis de linguagem, do ultraformal ao informal e popular.

- b. ressemantização de palavra autogenética com significado importado;
- c. ressemantização de palavra alogenética com significado vernáculo;
- d. composição ou derivação a partir de palavras alogenéticas.

Os cultismos e semicultismos enquadram-se na categoria alogenética, já que resultam de elementos importados do grego ou latim.

É preciso considerar ainda que a fronteira entre as duas famílias de processos é fluida, visto que palavras vernáculas podem receber significados estrangeiros, bem como, no caso das línguas românicas, ser restauradas, isto é, ter seu significante herdado substituído pelo significante latino que lhe deu origem, como é o caso de *mosteiro*, que hoje vem sendo progressivamente substituído por *monastério* (do latim *monasterium* < grego *monastérion*).

Como cada acepção de um lema pode ter etimologia diferente (no exemplo citado mais acima, *estrela*₁ é vocábulo autogenético, ao passo que *estrela*₂ é alogenético), a análise etimológica e consequente classificação das palavras em cultismos, semicultismos etc. devem ser feitas com base nas unidades lexicais de norma, isto é, os vocábulos ou *types*.

Unindo as perspectivas neológica e etimológica, podemos classificar preliminarmente os vocábulos de uma língua europeia nas diversas subcategorias dos quatro tipos básicos já mencionados: cultismo, semicultismo, vulgarismo ou outro. Algumas dessas subcategorias já são clássicas nos estudos etimológicos; outras foram propostas por nós em trabalhos anteriores (BIZZOCCHI, 1998; 2009) e algumas, ainda, estão sendo propostas agora, com base em levantamento feito sobre *corpora* em seis idiomas – português, espanhol, francês, italiano, inglês e alemão –, levantamento este cujos procedimentos metodológicos de coleta e análise serão descritos mais adiante.

Vulgarismos

2.1. Herança

Vocábulos vernáculos ou hereditários são aqueles herdados diretamente do sistema linguístico anterior à consolidação de uma língua como idioma literário, e, portanto, como língua de cultura autônoma, não mais como simples dialeto.

No caso das línguas românicas, é vernáculo tudo quanto já existia no latim vulgar e continuou a existir, sem solução de continuidade, nos diversos romances, de modo que, quando

estes ganharam o *status* de línguas distintas do latim, esse acervo de material lexical (e também gramatical, evidentemente) passou a constituir o léxico de base dessas línguas.

É vernáculo nas línguas germânicas o léxico que já existia nos diversos dialetos germânicos² (gótico, anglo, saxão, juto, franco, lombardo, nórdico etc.) e foi herdado pelas línguas germânicas modernas.

O vocabulário herdado costuma restringir-se ao universo da casa, do comércio e da agricultura, atividades mais comumente praticadas pelo povo. Por isso, termos referentes à atividade intelectual, existentes nas línguas clássicas, só entraram nas línguas europeias modernas mais tarde, como empréstimos.

É preciso ressaltar que, devido ao intercâmbio de palavras entre o latim e o germânico durante os últimos séculos do Império Romano e início da Idade Média (séculos I a VII d.C.), há também palavras vernáculas nas línguas românicas que são de origem germânica (por exemplo, português *guardar* < frâncico *wardan*), assim como palavras vernáculas nas línguas germânicas que são de origem românica (por exemplo, inglês *dish* < lat. *discus*). Por terem esses intercâmbios se dado entre as línguas-mães, do ponto de vista das línguas-filhas essas palavras são herdadas e não empréstimos.

2.2. *Estrangeirismo* (empréstimo de vulgarismo estrangeiro, nacionalizado ou não)

Chamamos aqui de estrangeirismo o empréstimo de palavra de qualquer língua que não o grego ou o latim clássicos, ainda que esses idiomas também sejam “estrangeiros” às línguas modernas. No entanto, tendo elas sido por muito tempo a única fonte de termos de cultura das línguas vulgares, as palavras greco-latinas quase nunca são sentidas como alienígenas nessas línguas, tanto que não passam pelos estágios naturais de aclimação dos empréstimos (xenismo e peregrinismo) a que estão sujeitos vocábulos de outras origens. Ao contrário, já entram na língua com todas as adaptações fonológicas, morfológicas e ortográficas que lhe garantem, desde o início, o estatuto de unidades genuinamente pertencentes ao léxico da língua importadora. Dito de outra maneira, salvo uns poucos casos (*status*, *habitat*), já ingressam plenamente nacionalizados, ao passo que alguns estrangeirismos já estão nacionalizados na grafia e pronúncia (*futebol*, *abacaxi*) enquanto outros ainda não (*pizza*, *bonbonnière*).

² Para a noção de *germance*, cfr. Bizzocchi, 2003.

Caso digno de nota entre os estrangeirismos são aqueles importados juntamente com suas desinências originais, que se incorporam ao radical na língua de chegada. Por exemplo, os verbos alemães *marschieren* e *garantieren* provêm do francês *marcher* e *garantir*, tendo a desinência francesa de infinitivo *-er* passado a fazer parte dos radicais *marschier-* e *garantier-* (a desinência de infinitivo em alemão é *-en*). Igualmente, os verbos ingleses *flatter* e *render* incorporam as desinências dos infinitivos franceses *flatter* e *rendre*.

2.3. Tradução de estrangeirismo (isto é, de vulgarismo estrangeiro)

A tradução, isto é, substituição dos morfemas de um vocábulo estrangeiro por equivalentes nacionais (ing. *skyscraper*, *hot-dog* > português *arranha-céu*, *cachorro-quente*) também é chamada de decalque ou *clipping*. No entanto, há uma diferença entre a tradução, tal como definida aqui, e o decalque, pois, se tomarmos o fr. *naturel* em face do lat. *naturalis*, houve a substituição do sufixo latino *-alis* pelo vernáculo francês *-el*. A literatura não considera isso um decalque, mas é, sem dúvida, uma forma de tradução. Portanto, constitui tradução de estrangeirismo a substituição de qualquer morfema de um estrangeirismo por outro, de qualquer outra origem, com significado equivalente. Em decorrência dessa definição, as traduções podem ser totais (todos os morfemas, como em ing. *skyscraper*, *hot-dog* > port. *arranha-céu*, *cachorro-quente*; fr. *chou-fleur* > port. *couve-flor*) ou parciais (apenas alguns morfemas, como em fr. medieval *estrangier* [atual *étranger*] > port. *estrangeiro*; ing. *goalkeeper* > port. *goleiro*; espanhol *cañón* > port. *canhão*).

Um caso particular de tradução total é o chamado empréstimo semântico, ou empréstimo de sentido, como em ing. *star* (“estrela” – corpo celeste > atriz famosa) > port. *estrela* (ambas as acepções) ou em ing. *mouse* (“rato” – animal > periférico de informática) > esp. *ratón* (ambas as acepções).

Deve-se alertar para os casos de falsa tradução, ou tradução duvidosa, em que duas ou mais línguas apresentam para o mesmo conceito palavras de mesma estrutura semântico-morfológica, que podem, no entanto, ter surgido de maneira independente e motivadas pelo próprio conceito e não uma pela outra. Por exemplo, não se pode afirmar que o port. *saca-rolhas* seja uma tradução do fr. *tire-bouchons*, nem que *acendedor de cigarros* seja tradução do ing. *cigarette-lighter*. Como o significante é fortemente motivado pelo significado (assim como *laranjeira* é motivado por *laranja*, o que leva praticamente todas as línguas a nomearem a árvore a partir do nome da fruta), é muito difícil afirmar que uma língua tenha influenciado

outra na escolha da designação. Todavia, tampouco se pode afastar essa hipótese: quando um artefato de origem estrangeira é introduzido numa sociedade, é natural que, junto com a coisa, venha o nome. Portanto, quem cunhou *saca-rolhas* ou *acendedor de cigarros* em português certamente não desconhecia as denominações originais desses objetos. Estes são, pois, casos em que a determinação exata da etimologia do vocábulo depende de dados experimentais que raramente estão disponíveis, o que leva tais itens lexicais a permanecerem indefinidamente como de classificação problemática ou insolúvel.

2.4. *Transcrição de estrangeirismo*

Muitas vezes, ao importar um significado, uma língua nem importa nem traduz fielmente o significante, mas cria um novo termo, apenas parcialmente motivado pelo original. Ao lat. *exemplum*, “exemplo”, corresponde em alemão *Beispiel*. Essa palavra é uma criação culta, já que não existia nos dialetos germânicos e, além disso, expressa um conceito abstrato e não corriqueiro. Entretanto, *exemplum* deriva do verbo *eximo*, de *ex-* (“fora”) + *emo* (“tomar”), “pôr de parte, eximir, isentar”. Já *Beispiel* derivaria de um suposto verbo *beispielen*, formado de *bei-* (“junto”) + *spielen* (“jogar”). Pode-se dizer que *Beispiel* foi motivado por *exemplum*, mas não é tradução deste. Igualmente, o al. *Beziehung*, “relação”, foi motivado pelo lat. *relatio*, mas há certas divergências semânticas entre o prefixo latino *re-* (“para trás, de novo”) e o al. *be-* (“em torno”), assim como entre *-latio*, derivado de *ferre*, “levar”, e *-ziehung*, proveniente de *ziehen*, “puxar”. Trata-se de uma “quase tradução”, portanto de uma transcrição da palavra em outro idioma, tal qual fazem os tradutores que, ao traduzirem um texto, principalmente literário, por vezes o transcriam³, parafraseando o original em vez de aterem-se ao significado literal.

2.5. *Composição com radicais vulgares*

Trata-se da palavra formada pela composição entre dois ou mais radicais vulgares (vernáculos ou estrangeiros). Só entram nesta categoria as palavras que tiverem efetivamente sido criadas por esse processo na própria língua que se está analisando. Assim, *puxa-saco*, *pernilongo* e *cabisbaixo* são compostos vulgares legitimamente portugueses. Já *cachorro-*

³ O termo *transcrição* foi cunhado por Haroldo de Campos para explicar a criação em português de um novo poema como “tradução” de um poema estrangeiro.

quente, citado no item anterior, não é resultado de composição em português, mas, como vimos, tradução do ing. *hot-dog*.

2.6. Derivação a partir de radical vulgar, seja com afixos cultos, semicultos ou vulgares

Trata-se da palavra formada pela derivação (prefixação, sufixação, parassíntese, derivação regressiva ou conversão) a partir de radical vulgar, seja herdado ou estrangeiro (*cabecear*, *mesário*, *saudosismo*). Não entram aqui palavras que, embora erradamente classificadas em dicionários etimológicos do português como derivadas, são na verdade traduções, como é o caso de *desafiar*, do fr. *défier*.

2.7. Reanálise por etimologia popular

Algumas palavras têm sua morfologia alterada pelo fenômeno da etimologia popular. Isso pode mascarar sua verdadeira etimologia. É o caso do inglês antigo *formest*, herdado do germânico **furmistaz*, superlativo de **furmaz*, que foi reanalisado como *foremost* (*fore* + *most*). Outros exemplos interessantes são:

- o fr. *camion*, que passou ao port. *camião*, depois reanalisado como *caminhão* por cruzamento com *caminho*;
- lat. vulgar *foresta* > português *foresta* > *floresta* (cruzamento com *flor*);
- fr. *neveu* > ing. *nephew* (cruzamento com ing. ant. *nef*, vernáculo);
- fr. *mousseron* > ing. *mushroom* (cruzamento com *mush* e *room*);
- esp. *cucaracha* > ing. *cockroach* (cruzamento com *cock* e *roach*);
- lat. *salsicia* > fr. ant. *saussiche* > ing. ant. *sausige* > ing. moderno *sausage* (por influência do sufixo *-age* < fr. *-age* < lat. *-aticum*);
- lat. vulg. **abantiare* > fr. *avancer* > ing. *advance* (cruzamento com o prefixo lat. *ad-*);
- lat. vulg. **allocare* > port. *alugar*, que gera o derivado *aluguer* (e sua variante *aluguel*), por cruzamento com *alquiler*, do ar. *al-kirâ*.

2.8. Retroversão ou retroviagem de vulgarismo

Esse fenômeno ocorre quando uma palavra, pertencente a uma língua A, é importada por uma língua B e, tempos depois, retorna à língua A, quando já havia caído em desuso ou sofrido

evolução nesta. Neste segundo caso, há a possibilidade de a forma recém-chegada conviver como alótopa da outra, mais antiga.

Por exemplo, o fr. med. *challengier* passou ao ing. *challenge*; a seguir, desapareceu do próprio francês, retornando ao francês moderno como *challenger* a partir de empréstimo do inglês. Igualmente, o fr. ant. *estage* resultou no ing. *stage*, que posteriormente gerou o fr. mod. *stage*. Esta última forma, significando “estágio”, convive atualmente em francês com a forma herdada *étage*, “andar”, evoluída a partir de *estage*.

Também são exemplos de retroversão o port. *fetiche* (< fr. *fétiche* < port. *feitiço*) e o fr. *sport* (< ing. *sport* < fr. med. *desport*), *budget* (< ing. *budget* < fr. med. *bougette*) e *test* (< ing. *test* < fr. med. *test* – atual *têt*).

Cultismos

3.1. Cultismo direto

Considera-se como cultismo direto o empréstimo de palavra diretamente do grego ou latim sem nenhum processo de vulgarização, mas apenas com o enquadramento morfológico (isto é, com eventual adaptação das desinências) e ortográfico no sistema da língua receptora. Por exemplo, os termos latinos *theatrum*, *temperatura* e *status* passaram diretamente ao francês na forma *théâtre*, *température* e *status*. Houve substituição das desinências latinas *-um* e *+ -a* pela vernácula *-e* nos dois primeiros casos e, no último, a desinência *-us* foi mantida, tendo-se incorporado ao radical.

3.2. Cultismo indireto

Cultismo indireto é o empréstimo de cultismo a partir de outra língua vulgar. No exemplo do item anterior, o fr. *théâtre*, *température* e *status*, provenientes do latim literário, passam ao português *teatro*, *temperatura* e *status* com nova adaptação morfológica e ortográfica. A partir do estudo das abonações dessas palavras em francês e português, sabe-se que, ao menos na Idade Média e Moderna, a maior parte dos grecismos e latinismos existentes nas línguas europeias chegam primeiramente ao francês, e posteriormente, graças à influência cultural dessa língua, passam às demais. Hoje, quando o inglês é a língua europeia de maior prestígio, grande parte dos cultismos passa primeiramente pelo inglês para depois atingir os outros idiomas.

3.3. *Composição com radicais cultos*

Grande parte dos cultismos existentes nas línguas europeias não existia em grego ou latim, tendo sido criada nas próprias línguas europeias, especialmente no francês e no inglês e, secundariamente, no italiano e alemão, por serem essas as línguas mais influentes culturalmente. Como boa parcela desses novos cultismos são termos técnico-científicos, é natural que surjam nas línguas dos países onde ocorre a maior parte das inovações científicas e tecnológicas.

Os cultismos são criados basicamente de duas maneiras: por composição entre radicais cultos ou por derivação a partir de um radical culto com afixos igualmente cultos.

Como exemplos de composição culta, temos o fr. *hélicoptère*, cunhado a partir dos radicais gregos *hélix* e *ptéron*, e *social-démocrate* (= *social* + *démocrate*).

3.4. *Derivação a partir de radical culto com afixos cultos*

Um exemplo de derivação a partir de um radical culto com afixos igualmente cultos é o fr. *initiative*, criado a partir de um hipotético lat. **initiativa*, derivado de *initiare*. Igualmente, o inglês produz *fractal* a partir do lat. *fractus* + *-alis*, e *genome* a partir de *gene* (< alemão *Gen* < gr. *génos*) + *-ome* (< gr. *-oma*).

Muitos cultismos criados por este processo pressupõem uma forma latina hipotética, jamais atestada, como no caso do ing. *privacy*, que não pode ser explicado como derivação direta de *private*, mas como incorporação de um lat. hip. **privatia*, supostamente derivado de *privatus*. Outro exemplo inglês é *legacy*, de um lat. hip. **legatia*, de *legatus*. Derivados cultos legitimamente portugueses são *fisiologismo* e *globalitarismo*.

Neste processo também entra a derivação regressiva, incluindo casos “exóticos”, como o italiano *legislazione*, que gera *legislare* (a rigor, não existe um radical **lare* em latim) ou o lat. *mixtus*, que passa ao ing. *mixt* (depois reescrito como *mixed*), que, por assemelhar-se a um participio, gera o verbo *mix*.

3.5. *Empréstimo de cultismo cunhado em língua vulgar estrangeira*

A partir dos exemplos dos itens anteriores, temos que o fr. *hélicoptère* e *initiative* passam ao port. *helicóptero* e *iniciativa*, assim como o ing. *fractal* e *genome* resultam no port. *fractal* e *genoma*. Trata-se da importação de cultismos estrangeiros.

3.6. *Restauração ou refecção total*

Às vezes, uma palavra herdada acaba com o tempo sendo substituída pelo vocábulo latino que lhe deu origem (MAURER JR., 1951, p. 62). Essa substituição de vernáculo pelo cultismo correspondente chama-se restauração ou refecção e pode ser parcial (refecção de apenas alguns morfemas da palavra) ou total (todos os morfemas). Neste último caso, constitui um cultismo. Por exemplo, o lat. *silentium* e *florem*, que haviam evoluído para o port. ant. *seenço* e *chor*, foram refeitos na Renascença para *silêncio* e *flor*. Atualmente, a palavra portuguesa *mosteiro* vem sendo progressivamente substituída por *monastério*, o que constitui um exemplo recente de restauração.

Embora o inglês não seja língua românica, pode-se dizer que o ing. *advice* e *adventure* são restaurações a partir dos empréstimos franceses vernáculos *avis* e *aventure*, que existiam em inglês medieval.

3.7. *Restituição total*

Denomina-se empréstimo de restituição (BIZZOCCHI, 1998, p. 124) o empréstimo de vulgarismo ou semicultismo estrangeiro com substituição de seus morfemas por correspondentes gregos ou latinos). Quando todos os morfemas da palavra estrangeira são substituídos por elementos greco-latinos, tem-se a restituição total. Por exemplo, o vulgarismo inglês *feed back* foi importado pelo português na forma *retroalimentar*, em que *feed* > *alimentar* e *back* > *retro-*, sendo ambos os elementos, *retro-* e *alimentar*, cultismos. Igualmente, o fr. *opérationnel*, semicultismo formado a partir do latinismo *opération* com o sufixo herdado *-el*, passa ao português como *operacional*, em que o radical se mantém culto e o sufixo é relatinizado.

Quando alguns morfemas são convertidos em elementos cultos e os demais permanecem semicultos ou vulgares, tem-se a restituição parcial, processo gerador de semicultismo, do qual falaremos adiante, em 4.7.

3.8. *Transcrição de cultismo*

Algumas vezes, uma língua vulgar cria um cultismo para substituir um termo culto que já existia em grego, latim ou em outra língua vulgar. Por exemplo, o lat. *consecrare* deveria ter passado ao fr. **consécrer*, mas, na verdade, passou a *consacrer* por influência do verbo *sacrer*,

igualmente culto. A forma **consacrare* não existia em latim, portanto *consacrer* é uma criação culta do francês.

Igualmente, o italiano substituiu o lat. *signatarius* por *firmatario* sob a influência de *firma* (< lat. *firma*), bem como substituiu o lat. *judaismus* por *ebraismo* a partir de *ebreo* (< lat. *hebraeus*).

Finalmente, o inglês cria o verbo *synthesise* a partir do fr. *synthetiser* por influência de *synthesis*.

3.9. *Empréstimo de cultismo estrangeiro sem adaptação grafo-fonética*

Alguns cultismos, por serem indiretos, entram na língua com a forma fonética do idioma a partir do qual foram importados. Isso produz uma divergência entre a forma assumida por esse cultismo e a que seria esperada se ele proviesse diretamente do grego ou latim. Assim, o fr. *frénésie*, *sirène* e *domino* passaram ao português nas formas *frenesi*, *sirene* e *dominó*, que reproduzem parcialmente a pronúncia francesa dessas palavras, quando as formas genuinamente greco-latinas em português seriam **frenesia*, **sirena* e **dômino*.

Da mesma forma, o ing. *privacy* passou ao it. *privacy* (com manutenção da pronúncia inglesa), quando seria de esperar a forma **privazia*.

Outros exemplos, ainda, são o fr. *express* (< ing. *express*), que concorre com a forma *exprès*, também existente em francês, e o al. *Service* (plural *Services*), do fr. *service(s)*, quando o esperável em alemão seria **Serviz* (plural **Servize* ou **Servizien*).

É importante observar que, diferentemente do que ocorre com o fenômeno descrito em 4.12, adiante, neste caso, não há vulgarização da palavra, visto que a grafia permanece greco-latina, ainda que segundo um padrão ortográfico estrangeiro.

3.10. *Retroversão ou retroviagem de cultismo*

Como já explicado em 2.7, a retroversão ocorre quando uma palavra pertencente a uma língua é importada por outra e, tempos depois, retorna à língua de origem. Quando a palavra objeto desse processo é culta ou sofre restituição na língua de chegada, tem-se como resultado um cultismo. Foi o que aconteceu com o fr. med. *parformer*, semicultismo formado a partir do prefixo herdado *par-* e do verbo culto *former*, que passou ao ing. *perform* (portanto com restituição do prefixo para *per-*) e, depois que já havia desaparecido do francês, retornou por intermédio do inglês na forma *performer*.

Semicultismos

4.1. Tradução de cultismo

Conforme vimos em 2.3, tradução é a substituição dos morfemas de um vocábulo estrangeiro por equivalentes nacionais. Quando esse processo se dá a partir de palavra grega, latina ou de um cultismo criado em outra língua, tem-se a tradução de cultismo, isto é, a substituição de qualquer morfema da palavra culta por outro, de natureza semiculta ou vulgar, com significado equivalente. Neste caso, as traduções também podem ser totais (todos os morfemas, como em lat. *superponere* > port. *sobrepor*; lat. *accipere* > al. *annehmen*) ou parciais (apenas alguns morfemas, como em lat. *interrumpere*, *perfectus*, *emotio* > port. *interromper*, *perfeito*, *emoção*).

Em relação ao cultismo, também existe o empréstimo semântico, ou empréstimo de sentido. Um exemplo disso é a palavra latina *nucleus*, que significava originalmente “caroço” e, posteriormente, no âmbito da ciência, passou a significar “núcleo” (da célula, do átomo etc.). Em alemão, a acepção científica de núcleo é representada pela palavra vernácula *Kern*, “caroço” (*Kernphysik* é física nuclear em alemão).

4.2. Transcrição de semicultismo

A transcrição de semicultismos é semelhante em sua lógica à de cultismos, com a diferença de que agora o resultado do processo é uma palavra semiculta. Por exemplo, com base no lat. *favere*, o italiano criou *favorire* e suas formas conjugadas *favorisco*, *favorisci* etc. a partir do cultismo *favore* mediante o sufixo herdado *-isc-*. O inglês fez o mesmo que o italiano, criando o verbo *to favour* a partir do substantivo semiculto *favour* (< fr. *faveur*).

4.3. Tradução de semicultismo estrangeiro

Os mesmos processos de tradução (total, parcial e semântica) também se dão a partir de semicultismos estrangeiros. Por exemplo, o fr. *aéroport*, composto híbrido, deu em inglês *airport*, em que, curiosamente, ambos os radicais são franceses. O fr. *désordre* deu o port. *desordem*, em que o elemento semiculto *ordre* (< lat. *ordinem*) foi substituído pelo igualmente semiculto *ordem*.

4.4. *Transcrição de semicultismo estrangeiro*

O it. *favorire*, visto acima, foi traduzido pelo fr. *favorir*, *favoris* etc. com os sufixos *-i-* ou *-iss-*. O português e o espanhol fazem o mesmo: *favor* + *-ecer* = *favorecer*.

4.5. *Metamorfismo*

Trata-se da substituição, acompanhada da adaptação ortográfica correspondente, de fonemas da palavra culta por outros, típicos da língua vulgar, seja por evolução fonética ao longo do tempo ou por mutação no momento da introdução do vocábulo: o lat. *includere* passou ao port. med. *includir*, que posteriormente evoluiu para *incluir*; já o lat. *secta* entrou em port. como *seita* e não **secta* – que seria o esperável em se tratando de empréstimo de palavra latina –, por analogia com palavras herdadas que sofreram a mesma evolução fonética (por exemplo, lat. *directus* > port. *direito*).

O metamorfismo é espontâneo quando resulta da evolução fonética natural, acompanhado das correspondentes adequações gráficas (por exemplo, lat. *canonicus*, *clericus*, *capitulum* e *articulus* > port. *cônego*, *clérigo*, *cabido* e *artigo*) e deliberado quando ocorre por analogia com palavras herdadas (lat. *scepticus* e *includere* > port. *cético* e *incluir*, por analogia com *septem* > *sete* e *sudare* > *suar*).

Em certos casos, o metamorfismo é obrigatório, como em *estátua*, do lat. *statua*, em que o acréscimo do *ē* inicial é exigido pela própria estrutura fonológica da língua portuguesa. Em outros casos, o metamorfismo é facultativo, como nas palavras latinas *doctor* e *conceptus*, que originaram *doutor* e *conceito* quando poderiam ter resultado em **doctor* e **concepto*.

Um caso particular de metamorfismo é aquele que afeta, imediata ou posteriormente, um vocábulo restaurado. Como vimos em 3.6, retro, restauração ou refecção é a substituição de termo herdado pelo vocábulo latino que lhe deu origem. Ocorre que, às vezes, o termo restaurado também sofre metamorfismo, seja por mutação fonética e adaptação ortográfica naturais pela evolução histórica, seja por analogia com termos vernáculos no momento mesmo da refecção.

Por exemplo, o port. antigo *chor* (< lat. *florem*) foi feito a certa altura para *flor*, que logo evoluiu para *fror* por rotacismo e a seguir para *frol* por dissimilação. (Também se pode dizer que houve simplesmente metátese entre *l* e *r*.) Portanto, *frol* é metamorfismo de um vocábulo que, por sua vez, era refecção de outro.

Já o fr. *égal*, que substituiu o antigo *ivel* (< lat. *aequalem*), é uma refecção que já nasceu metamorfozizada, pois a forma culta esperada em francês seria **égal*, que jamais foi documentada.

4.6. *Restauração parcial*

É a substituição de alguns morfemas de um vulgarismo pelos equivalentes latinos, com a manutenção dos demais: lat. *inimicum*, que deu o port. ant. *eOemigo*, depois refeito para *inimigo*, mantendo-se o sufixo herdado *-igo*; lat. *felicem* > port. ant. *fiiz*, restaurado para *feliz* (a restauração total seria **felice*).

4.7. *Restituição parcial*

Como vimos em 3.7, restituição é o empréstimo de vulgarismo ou semicultismo estrangeiro com substituição de seus morfemas por correspondentes gregos ou latinos, a qual pode ser total ou parcial. Quando alguns morfemas semicultos ou vulgares são traduzidos por correspondentes cultos e outros não, tem-se uma restituição parcial, como no ing. *starship*, que passou ao fr. *astronef*.

4.8. *Composição híbrida*

É aquela que se dá entre um radical culto e um semiculto, entre um radical culto e um vulgar, entre um radical semiculto e um vulgar ou, ainda, entre dois radicais semicultos (por exemplo, port. *auriverde*, *rubro-negro*, *bafômetro*).

4.9. *Derivação híbrida*

A derivação híbrida ocorre a partir de radical culto com afixo semiculto ou vulgar (port. *agricultável*, *deseeducar*) ou a partir de radical semiculto com afixos de qualquer natureza (fr. *désordre*).

4.10. *Composição ou derivação “anômala” de elementos cultos*

Em alguns casos, um cultismo pode ser substituído por palavra igualmente formada de elementos cultos, porém combinados de modo não admissível pelo sistema fonológico do grego ou do latim. Por exemplo, o espanhol fez corresponder aos vocábulos latinos *diminuere*, *immortalis* e *commotio* as formas *disminuir*, *inmortal* e *conmoción*, ou seja, eliminou as

assimilações *sm* > *m* e *nm* > *mm* do latim, como se essas palavras voltassem à sua forma pré-histórica, antes de tais metaplasmos se produzirem. Ressalte-se que formas pré-históricas latinas como **disminuere*, **inmortalis* e **conmotio* jamais foram atestadas, sendo dedutíveis por reconstrução. No entanto, o surgimento dessas formas no espanhol não deve ter-se dado por um desejo de reconstruir a forma primitiva dessas palavras (afinal, elas existem em espanhol desde bem antes do advento da linguística histórico-comparativa) e sim por efeito da analogia com outras palavras (por exemplo, *disponer*, *intenso* e *contracción*).

4.11. Reanálise por etimologia popular

O fenômeno da etimologia popular, já tratado em 2.6, pode gerar vulgarismos, semicultismos e, teoricamente, até cultismos. Do semicultismo fr. *dromedier* (< lat. *dromedarius*) resultou o al. *Trampeltier*, literalmente, “animal que patea”, a partir das palavras herdadas *trampeln* (patear) e *Tier* (animal). Outros exemplos são:

- lat. *corporalis* > it. *caporale* (cruzamento com it. *capo*, “cabeça”);
- lat. *impedio* > port. *impeço* (cruzamento com *peço* < lat. *petio*);
- lat. *impressa* > port. *imprensa* (cruzamento com *prensa* < lat. *prehensa*).

4.12. Empréstimo de semicultismo estrangeiro

Quando uma língua importa um semicultismo, este mantém seu caráter semiculto. Assim, o fr. *nécessaire*, metamorfismo do lat. *necessarius*, passou ao port. *nécessaire* na acepção de bolsa ou estojo feminino para utensílios de toalete. Da mesma forma, o lat. *societas* deu o fr. *société* por tradução do sufixo *e*, a seguir, esta forma passou ao ing. *society*.

4.13. Retroversão ou retroviagem de semicultismo

O lat. *humor*, “líquido”, passou nessa acepção ao fr. med. *humour* (atual *humeur*) por tradução do sufixo. Este passou por empréstimo ao ing. *humour*, que mais tarde adquiriu a acepção de “senso de humor, graça, humorismo”. Esta acepção retornou ao francês como *humour*, que hoje convive com *humeur* (formas divergentes).

Igualmente, o fr. *entrevue*, “entrevista”, passou ao ing. *interview* por restituição do prefixo *e*, posteriormente, voltou ao francês como *interview*. Hoje, ambas as formas convivem em

francês com significados diferentes: *entrevue* é uma entrevista para tratar de negócios (uma oferta de emprego, por exemplo) e *interview* é a entrevista jornalística.

4.14. *Empréstimo de cultismo estrangeiro com reprodução gráfica da pronúncia original*

Normalmente, os cultismos entram nas línguas vulgares por via escrita e só depois chegam à fala. Por isso, as línguas importadoras costumam conservar a grafia greco-latina (salvo adaptações ortográficas obrigatórias), ainda que em detrimento da pronúncia, já que é muito difícil manter ambas. Entretanto, às vezes, quando uma língua vulgar importa um cultismo indireto, isto é, a partir de outra língua que não o grego ou latim, pode ocorrer de esse empréstimo se dar pela via oral. Nesse caso, tanto pode ser mantida a grafia que o cultismo tem na língua exportadora, com prejuízo ou não da pronúncia (cfr. 2.9, acima), quanto pode ser reproduzida a pronúncia dessa língua (que quase nunca coincide com a do grego ou latim) com adaptação da grafia. Dois exemplos tornarão isso mais claro.

O sueco importou o termo para “estátua” do fr. *statue* e não diretamente do lat. *statua*. Como o empréstimo se deu por via oral, o sueco manteve a pronúncia francesa, adaptando conseqüentemente a grafia. Como resultado, “estátua” em sueco é *staty* (o *y* tem em sueco a mesma pronúncia que o *u* do francês; além disso, *staty* é palavra oxítone como *statue* em francês).

De modo análogo, o fr. *estime* deu em inglês *esteem* e não **estime*, como seria de esperar se a palavra em questão tivesse penetrado por via literária. (O inglês também tem *estimate*, empréstimo direto do latim, com outro significado.) A grafia *ee* reproduz a pronúncia do *i* francês.

4.15. *Empréstimo de cultismo estrangeiro acompanhado de desinências*

Assim como acontece com palavras vulgares, cultismos também podem ser emprestados acompanhados de desinências, que se incorporam ao radical da palavra na nova língua. Os cultismos franceses *privé* e *habitué* passaram sem alteração ao port. *privé* e *habitué*, respectivamente, portanto com manutenção da desinência de participio *-é*. Por ser uma desinência em francês, ela não retira da palavra o seu caráter culto. Já em português, que não possui tal desinência, *-é* faz parte do radical da palavra. Por ser um elemento francês e não latino, ele confere aos termos portugueses um caráter híbrido, isto é, semiculto. O mesmo ocorre com o ing. *occupy*, *privy* e *deputy* (< fr. *occupé*, *privé*, *député*), em que *-y* é elemento francês,

e com o al. *informieren* (< fr. *informer*). Aliás, a terminação *-ieren* tornou-se um sufixo produtivo em alemão, entrando na formação de diversos verbos, a maioria de origem latina.

Vocábulos inclassificáveis nas categorias anteriores

5.1. Epônimos e compostos ou derivados de nomes próprios

Epônimos são nomes próprios que se tornam comuns. Muitos deles homenageiam o criador do objeto que nomeiam. Por exemplo, *ampère*, *macadame*, *gilete* e *zepelim* são epônimos dos sobrenomes *Ampère*, *Gillette*, *MacAdam* e *Zeppelin*, respectivamente. No caso de *gilete*, o substantivo comum não derivou diretamente do nome do inventor da lâmina de barbear e sim da marca do produto, originalmente fabricado pelo próprio inventor. Outro caso de marca registrada que se tornou nome comum é *maisena* (< *Maizena*®).

Como nomes próprios, especialmente os gentílicos, não fazem parte do léxico da língua, palavras epônimas, bem como compostos e derivados de nomes próprios, não podem ser classificados nem como cultismos nem como semicultismos nem como vulgarismos.

Dentre os compostos e derivados desse tipo temos: port. *abreugrafia*, fr. *voltmètre*, ing. *Newtonian*, al. *Kantismus*, e, de modo geral, todos os derivados de nomes de pessoas e países, como *hitlerista* e *zimbabuense*.

5.2. Siglas, seus compostos e derivados

Siglas, acrônimos e abreviaturas tampouco são palavras da língua, embora estas últimas possam ser reconstituídas em sua forma por extenso (por exemplo, *Prof.* por *Professor*). Muitas siglas são passíveis de lexicalização e chegam a ter uma pronúncia silábica, enquanto outras permanecem pronunciadas de modo soletrado (BARBOSA, 1993). De qualquer maneira, elementos como *USP*, *AIDS*, *PT*, e seus derivados *uspiano*, *aidético*, *anti-AIDS*, *petista*, são inclassificáveis do ponto de vista etimológico.

5.3. Vocábulos formados de fractomorfemas (palavras truncadas e palavras-valise)

Chama-se truncação (ALVES, 2007, p. 68) a eliminação de uma parte, geralmente a final, da sequência lexical. Exemplos em português são *preju* (*prejuízo*), *micro* (*microcomputador*), *vice* (*vice-presidente*), *ex* (*ex-presidente*, *ex-marido*) etc. Em francês, temos *métro* (*métropolitain*), *photo* (*photographie*), *cinéma* (*cinématographe*), que passaram ao port. *metrô*, *foto* e *cinema*. O resultado desse processo é um elemento que não chega a ser um morfema: ele

é parte de um morfema ou a sequência formada por um morfema e parte de outro. Mesmo em casos como *auto* (truncação de *automóvel*), não se pode dizer que se trate do elemento de composição grego *auto-* (significando “próprio”) pura e simplesmente. Neste caso, *auto* significa “automóvel”. Elementos desse tipo são chamados de fractomorfemas ou quase-morfemas (LINO, 1990, p. 30-31).

Compostos ou derivados de fractomorfemas são as chamadas palavras-valise (ALVES, 2007, p. 69). Criações portuguesas desse tipo são: *showmício* (*show* + *comício*), *brasiguaiio* (*brasileiro* + *paraguaiio*) e *portunhol* (*português* + *espanhol*). Em francês, temos *informatique* (*information* + *automatique*, que passou ao port. *informática*), e em inglês, *reprography* (*reproduction* + *photography* > port. *reprografia*), *autopart* (*automobile* + *part* > port. *autopeça*) e *miniskirt* (*miniature* + *skirt* > port. *minissaia*).

5.4. Neologismos fonológicos (*ex nihilo* ou onomatopaicos)

Conforme explicado na Introdução, os neologismos fonológicos podem ser do tipo *ex nihilo* (criação de um novo significante a partir do nada: *poperô*, *chinfrim*, *pirlimpimpim*, *zureta*) ou onomatopaicos (imitação do som do significado: *plim-plim*, *miau*, *zunzum*, *ziguezague*, *tique-taque*, *tilim*, e seus derivados *miar*, *zunzunar*, *ziguezaguear*, *tilintar*).

5.5. Corruptelas

A corruptela pode ser a deformação acidental de um significante, estando assim em grande parte relacionada à etimologia popular (*bicho no corpo inteiro* > *bicho carpinteiro*), mas também pode ser proposital, sendo este processo mais frequente na criação de gírias. Em português, temos, por exemplo, *carango* (< *carro*), *tchurma* (< *turma*), *neura* (< *neurose*) e *busão* (*ônibus*).

5.6. Empréstimo de não classificável

Grande parte das palavras truncadas e dos compostos e derivados de fractomorfemas são, na verdade, importações. O port. *foto*, *cinema*, *informática* e *metrô* são, como dito mais acima, empréstimos do fr. *photo*, *cinéma*, *informatique* e *métro*. Do inglês vieram, dentre outros, *motel* (*motor* + *hotel*) e *minissaia* (tradução do fractocomposto *miniskirt*). Também chegaram ao português vários neologismos fonológicos, como *zoom* (do inglês) e *gás* (do holandês). Mas o que mais encontramos em nossa língua são os empréstimos de epônimos e de seus compostos

e derivados, como *ampère* e *amperímetro* (< fr. *ampèremètre*), *kantiano* (< fr. *kantien* < al. *Kantisch* < *Kant*) e *tupperware* (< ing. *tupperware* = *Tupper* + *ware*).

5.7. Composto ou derivado de não classificável importado

A partir de palavra inclassificável estrangeira, é possível formar um composto ou derivado nacional. Por exemplo, a partir do verbo inglês *zap* obtivemos em português *zapear*, com acréscimo do sufixo *-ear*.

Coocorrência de processos

Dois ou mais desses processos podem coocorrer num mesmo vocábulo, como no caso da palavra *instrução*: a queda do *c* do lat. *instructio* é um metamorfismo analógico (cf. lat. *fructum* > port. *fruto*, palavra herdada), e a substituição de *-tio* por *-ção* é tradução parcial.

Outros exemplos são: lat. *coordinator* > port. *coordenador* (metamorfismo e tradução do sufixo); lat. *infectare* > port. *infectar* > *desinfectar* (derivação) > *desinfetar* (metamorfismo). Em outras línguas, temos vários casos, como o fr. *égalité*, resultante de empréstimo do lat. *aequalitas* por metamorfismo do radical (*aequalis* > *égal*) e tradução do sufixo *-tas* por *-té*.

Material e método

O levantamento dos processos aqui elencados se deu sobre um *corpus* composto de textos retirados da *internet* redigidos em norma padrão nas seguintes línguas: português, espanhol, francês, italiano, inglês e alemão.

Além de embasar a fundamentação teórica, permitindo o reconhecimento de novos processos, que ainda não haviam sido descritos na literatura especializada, a utilização de *corpus* amostral possibilita a realização de um mapeamento estatístico preliminar desses processos, que permite antever quais deles são mais frequentes em cada idioma e, por conseguinte, fazer comparações entre as composições etimológicas dos léxicos desses idiomas. Embora, dadas as dimensões desta pesquisa e o tamanho do *corpus* coletado, o levantamento estatístico não tenha sido exaustivo, ele indica claramente as diferenças de comportamento entre as línguas quando se trata de introduzir novos itens lexicais. Além disso, ele confirma em grande parte as tendências já detectadas em estudo anterior (BIZZOCCHI, 1998).

Foram coletados textos dos seguintes gêneros e subgêneros discursivos:

- Jornalístico:
 - Atualidades;
 - Política;
 - Economia;
 - Esporte;
 - Cultura.
- Institucional:
 - Governo;
 - Fundações;
 - Indústria;
 - Comércio;
 - Tecnologia.
- Religioso.

Foram excluídos por princípio textos literários, quer em prosa ou em versos, por ser a literatura gênero desviante da norma ou, no dizer de Roland Barthes, do “grau zero da escrita”, não retratando os usos efetivos da língua e sim as preferências e idiossincrasias estilísticas do autor.

Ao final da coleta, obteve-se um acervo de 18.306 palavras-ocorrência. As amostras tiveram as seguintes dimensões (em palavras-ocorrência ou *tokens*):

- Português: 2.433
- Espanhol: 3.271
- Francês: 3.124
- Italiano: 3.496
- Inglês: 3.250
- Alemão: 2.732

Foram coletadas dos textos as palavras lexicais (dotadas de lexemas), isto é, substantivos, adjetivos, verbos e advérbios de modo (o chamado léxico aberto, aquele sujeito a inovações neológicas). As palavras gramaticais (léxico fechado) foram desprezadas nesta pesquisa por apresentarem evolução histórica muito mais lenta que a das palavras lexicais, já que estão

diretamente ligadas às relações estrutural-funcionais do sistema, além de pertencerem, em sua maioria, ao estrato hereditário do léxico. Conseqüentemente, nelas, os processos etimológicos objeto deste estudo aparecem de forma muito menos nítida. Não se exclui, no entanto, a possibilidade de um levantamento da etimologia das palavras gramaticais em futuras pesquisas.

As palavras coletadas foram inseridas em bancos de dados organizados por língua e gênero discursivo, nos quais foram ordenadas alfabeticamente e por classe gramatical. Foram indexadas segundo sua etimologia a partir de pesquisas em dicionários etimológicos e outras fontes bibliográficas pertinentes. Uma pequena parcela do *corpus* teve sua etimologia listada como “insolúvel” ou “problemática” (palavras de étimo desconhecido ou com étimos divergentes de um dicionário para outro). Esse dado (porcentagem de itens problemáticos) também foi computado estatisticamente.

As palavras foram coletadas e inseridas no *corpus* tal qual ocorreram nos textos (*tokens*) e também agrupadas segundo as formas flexionais em que ocorreram (*types*). Isso se deveu ao fato de que algumas formas flexionais têm etimologia diferente de outras (por exemplo, *caráter* é metamorfismo, mas seu plural *caracteres* é cultismo; *transgredir* é cultismo, mas *transgride* é metamorfismo, e assim por diante). Com isso, obteve-se um inventário de *tokens* e outro de *types*, o que permitiu a comparação do comportamento etimológico da língua em termos de sistema e de uso. Em face do neologismo semântico (que em muitos casos é também um decalque semântico de palavra estrangeira), a acepção da palavra em cada uma de suas ocorrências textuais teve de ser levada em conta, e acepções diferentes foram computadas como vocábulos diferentes de um mesmo lema.

Cada item do *corpus* foi indexado por meio de um rótulo formado por um algarismo e uma letra, sendo que o algarismo indica a categoria etimológica (1: vulgarismo; 2: semicultismo; 3: cultismo; 4: palavra inclassificável; 5: palavra de etimologia problemática) e a letra, cada um dos subtipos dessas categorias. Os rótulos utilizados na indexação do *corpus* foram os seguintes:

1 – Vulgarismos:

- 1a – Herança;
- 1b – Estrangeirismo;
- 1c – Tradução de estrangeirismo;
- 1d – Composto ou derivado vulgar;

1e – Reanálise por etimologia popular;

1f – Retroversão de vulgarismo.

2 – Semicultismos:

2a – Tradução de cultismo;

2b – Tradução de semicultismo estrangeiro;

2c – Metamorfismo;

2d – Restauração parcial;

2e – Restituição parcial;

2f – Composto semiculto ou derivado semiculto;

2g – Composto ou derivado “anômalo” de elementos cultos;

2h – Reanálise por etimologia popular;

2i – Empréstimo de semicultismo estrangeiro;

2j – Retroversão de semicultismo;

2k – Empréstimo de cultismo estrangeiro com reprodução gráfica da pronúncia original;

2l – Empréstimo de cultismo estrangeiro acompanhado de desinências.

3 – Cultismos:

3a – Empréstimo direto ou indireto do grego ou latim;

3b – Restauração total de vulgarismo ou semicultismo;

3c – Restituição;

3d – Composto ou derivado culto;

3e – Reanálise por etimologia popular;

3f – Tradução de um cultismo por outro;

3g – Empréstimo de cultismo estrangeiro sem adaptação grafo-fonética.

4 – Inclassificáveis:

4a – Composto ou derivado de nome próprio;

4b – Composto ou derivado de sigla;

4c – Composto ou derivado de fractomorfemas, incluindo palavras-valise;

4d – Neologismo fonológico;

4e – Corruptela;

4f – Composto ou derivado de neologismo fonológico ou corruptela;

4g – Empréstimo de não classificável;

4h – Composto ou derivado de não classificável importado.

5 – Problemáticos.

Para a organização e processamento do banco de dados, bem como para os procedimentos de estatística lexical, foram utilizadas técnicas correntes da lexicometria e da linguística de *corpus*, com emprego de *software* específicos.

A realização da estatística lexical incluiu procedimentos matemáticos (cálculo de médias, desvio-padrão, teste de Pearson etc.), que forneceram as tabelas 1 (*tokens*) e 2 (*types*), com o perfil etimológico do léxico das línguas estudadas, divididos por gêneros e subgêneros discursivos, o que permitiu tirar algumas conclusões preliminares a respeito das escolhas realizadas por cada língua, variedade e gênero no tocante à criação vocabular e suas implicações no reconhecimento de padrões neológicos e na inferência de comportamentos lexicais ideologicamente orientados.

Resultados e discussão

A partir do levantamento estatístico sobre o *corpus*, obtiveram-se duas tabelas de frequência média percentual (μ) de processos etimológicos: por *tokens* (tabela 1) e por *types* (tabela 2). As disparidades de comportamento entre a frequência de *tokens* e *types* se deve ao fato de que, na análise por *tokens*, as palavras-tema e as palavras-chave do texto, que por definição ocorrem com frequência superior à média da língua, geram distorções significativas, supervalorizando certos processos em detrimento de outros. Por exemplo, num texto de informática em português, a palavra *software* costuma ocorrer com mais frequência do que em textos de outros gêneros. Isso eleva sobremaneira o índice percentual de estrangeirismos na amostra, o que não corresponde à realidade da língua. Por isso, procedeu-se também à análise por *types*, na qual cada forma (*type*) é computada uma única vez, independentemente de sua frequência absoluta na amostra. Esse procedimento neutraliza possíveis distorções devidas ao gênero ou ao tema do texto coletado e retrata melhor o comportamento do léxico como um todo. A análise dos resultados tomará como base, portanto, a tabela 2.

O teste de Pearson (χ^2) produziu valores dentro do limite de aceitabilidade para uma amostra normal, e os desvios-padrão (σ^2) raramente superaram 3,0%, o que indica um razoável grau de precisão em face do tamanho da amostra. No entanto, a despeito da dimensão do desvio, optamos por exibir os valores com duas casas decimais, pois alguns valores afiguraram-se muito baixos e, num arredondamento, teriam sido reduzidos a zero.

Examinando primeiramente os totais de cada macrocategoria, podemos observar que, no português, espanhol e italiano, os vulgarismos giram em torno de 40%, em francês são aproximadamente 47%, em inglês 56%, e no alemão chegam a 68%. Os semicultismos em português, francês, italiano e alemão correspondem a cerca de 20% do léxico; em espanhol e inglês, esses índices são mais baixos (14% e 8%, respectivamente). Os cultismos situaram-se entre 32% e 36% em todas as línguas estudadas, com exceção do espanhol (44%) e do alemão (10%). O espanhol apresentou uma frequência de cultismos superior à das demais línguas, ao passo que o alemão teve frequência nitidamente inferior. Finalmente, as palavras não classificáveis em nenhuma das três categorias anteriores ficaram abaixo de 1% em praticamente todos os idiomas. Esses totais são convergentes com os resultados obtidos em pesquisa anterior (BIZZOCCHI, 1998), o que indica a consistência do levantamento.

8.1. *Análise comparativa de vulgarismos*

Dentre os vulgarismos, a primeira coisa a ressaltar a partir do exame dos dados é que a frequência de palavras herdadas ficou em torno de 30% em todas as línguas. Essa igualdade de comportamento era esperável em relação a esse estrato lexical, já que o léxico de base tende a ser o mesmo em qualquer língua europeia.

Os estrangeirismos ficaram abaixo dos 5%, exceto em inglês, em que a forte presença de galicismos desde a invasão normanda do século XI é característica marcante do idioma.

A tradução de estrangeirismo é pouco frequente em todas as línguas analisadas (menos de 3%), revelando-se, pois, um processo pouco produtivo.

A síntese (composição ou derivação) de vulgarismo ficou abaixo dos 5% em português, espanhol e italiano; girou em torno de 10% em francês e inglês, ao passo que, no alemão, atingiu os 25%. Isso revela a alta produtividade desse processo em alemão, o que já havia sido detectado em nossa prospecção anterior (1998).

A reanálise por etimologia popular é rara em todas as línguas (menos de 0,5%), não tendo ocorrido na amostra em português, francês e italiano. A retroversão de vulgarismo só ocorreu

no francês, mesmo assim com índice baixíssimo (0,1%), e sempre em relação à língua inglesa, o que faz sentido: no passado, o francês influenciou grandemente o inglês; hoje se dá o inverso. Por isso, o inglês exporta para o francês palavras que ele próprio havia importado daquela língua anteriormente.

8.2. *Análise comparativa de semicultismos*

A tradução de cultismo situou-se em torno dos 10%, exceto em espanhol (5%) e em inglês (1%). Como, à exceção do alemão, a maioria das traduções é parcial, constatou-se no *corpus* que o espanhol e o inglês traduzem menos os afixos latinos do que as outras línguas.

A tradução de semicultismo ficou abaixo de 1% em todas as línguas (em francês nem chegou a ocorrer), sendo, a exemplo da tradução de cultismo, mais rara ainda em inglês, pela razão já exposta.

O metamorfismo é bem mais frequente nas línguas românicas do que nas germânicas (média de 7% naquelas e de 0,5% a 1% nestas), o que parece lógico, tendo em vista que esse processo se dá em geral por analogia com palavras herdadas do latim, as quais praticamente só existem nos idiomas neolatinos.

Restaurações parciais foram pouco frequentes nas línguas românicas (menos de 0,5%) e inexistentes nas germânicas, o que também faz sentido, já que a restauração é o restabelecimento da forma latina de uma palavra herdada do latim que sofreu evolução fonética.

A restituição parcial mostrou-se rara em todas as línguas, não tendo ocorrido em francês e alemão. Em espanhol e inglês, ficou abaixo de 0,1%; no português, apresentou 0,4%, índice pequeno, mas significativamente superior aos demais.

A síntese de semicultismo (composição ou derivação) situou-se ao redor de 1%, exceto em francês (3%) e, sobretudo, em alemão (7%). Esse fato já havia chamado nossa atenção na pesquisa anterior. Isso se explica porque o francês foi durante muito tempo um polo gerador de neologismos, especialmente sintagmáticos, ao passo que as outras línguas importavam ou traduziam essas criações. Quanto ao alemão, é também um idioma em que tal tipo de neologismo é bastante produtivo, dada a facilidade dessa língua em justapor palavras para formar compostos, embora não seja uma língua eminentemente exportadora de palavras.

Os compostos ou derivados anômalos só foram registrados em espanhol, mesmo assim com baixa frequência (0,2%), o que indica que se trata de processo típico dessa língua.

A geração de semicultismo por etimologia popular só se deu em espanhol e francês, com frequências bem baixas (0,06% e 0,03%, respectivamente). Trata-se de processo bastante raro.

O empréstimo de semicultismo oscilou de 0,2% a 0,6%, com exceção do inglês, que teve 4%. Isso se deve à tendência dessa língua de importar termos franceses, dentre os quais muitos semicultismos.

A retroversão de semicultismo, prevista teoricamente, não ocorreu em nenhuma língua. No entanto, não podemos saber, até que levantamentos mais exaustivos sejam levados a cabo, se tal processo não existe ou se simplesmente é muito raro.

A importação de cultismo com reprodução gráfica da pronúncia vulgar só apareceu em italiano e inglês (por volta de 1%). Já a importação de cultismo com incorporação de desinências vulgares só se deu em inglês (0,07%) e em alemão (0,08%). De fato, praticamente só nas línguas germânicas esse processo faz sentido, já que é nelas que as desinências românicas não são sentidas como tal e, por isso, se agregam ao radical da palavra.

8.3. *Análise comparativa de cultismos*

A importação de palavras gregas ou latinas corresponde a aproximadamente 30% do léxico das línguas analisadas (37% em espanhol), à exceção do alemão, que apresentou apenas 7%. Esse resultado é coerente com medições anteriores e revela a tendência alemã de substituir palavras greco-latinas por criações com material autóctone.

Em todos os idiomas, a restauração total ficou abaixo de 1% (zero em alemão), sendo um processo bastante raro. O mesmo pode-se dizer da restituição total (índices abaixo de 1%, com zero para o francês).

A síntese de cultismo girou em torno dos 4% em todas as línguas (um pouco menos – 2,3% – em alemão).

A etimologia popular é raríssima como geradora de cultismo, afinal dificilmente as camadas menos escolarizadas da população, onde costuma nascer esse tipo de neologismo, utilizam material greco-latino em suas criações. Por isso, somente o espanhol registrou o processo, com 0,2% de frequência.

A tradução de um cultismo por outro só ocorre em francês e italiano e, mesmo assim, com baixa frequência (0,04% e 0,14%, respectivamente). O empréstimo de cultismo sem adaptação grafo-fonética só apareceu em italiano e alemão (cerca de 0,05%).

8.4. *Análise comparativa de inclassificáveis*

Compostos ou derivados de nomes próprios são bem pouco frequentes (em torno de 0,1%), não tendo ocorrido nem em espanhol nem em francês. Compostos ou derivados de siglas ou de fractomorfemas, bem como os neologismos fonológicos, também se mostraram raríssimos. As corruptelas, mais comuns no registro popular e na gíria, não apareceram no *corpus*, formado de textos em língua padrão na sua variedade escrita. Já os compostos ou derivados de neologismo fonológico ocorreram em francês, italiano e inglês, embora com baixa frequência (aproximadamente 0,04%).

Empréstimos de palavras não classificáveis foram registrados em todas as línguas (entre 0,1% e 0,4% nas românicas, e 0,04% nas germânicas). Já os compostos ou derivados formados a partir dessas palavras só ocorreram em português (0,12%) e alemão (0,11%).

Restou ainda uma pequena margem (cerca de 0,1%) de palavras que não puderam ser classificadas satisfatoriamente, mas trata-se de um índice pouco expressivo, que não chega a comprometer a confiabilidade da estatística.

Conclusão

O presente estudo pretendeu, a partir de pesquisa empírica, revelar novos processos de criação lexical nas línguas europeias, os quais não costumam ser devidamente reconhecidos em descrições etimológicas desses idiomas. Por conseguinte, muitas palavras têm sido catalogadas e classificadas erroneamente, com graves repercussões tanto na produção científica quanto no ensino. Nesse sentido, pretendeu-se contribuir para uma melhor compreensão da real dinâmica da criação e renovação do léxico das línguas.

Em relação ao perfil etimológico, depreendido da análise estatística comparativa, pudemos perceber que o inglês e, sobretudo, o alemão são mais vulgarizantes do que as línguas românicas, o que parece revelar um apego maior destas pelas formas cultas e semicultas. Em relação aos semicultismos, merece destaque sua fraca presença no inglês, compensada pela alta frequência de vulgarismos e cultismos. Estes últimos situam-se acima da média em espanhol e bem abaixo em alemão.

Em relação ao perfil etimológico, depreendido da análise do *corpus*, alguns processos são típicos de uma única língua (por exemplo, a retroversão de vulgarismo em francês e a síntese anômala em espanhol), enquanto outros predominam em certos idiomas, demonstrando, assim, as preferências etimológicas de cada língua e, conseqüentemente, seu perfil “ideológico”, que,

embora se tenha definido historicamente, repercute até hoje nas escolhas lexicais que tais línguas operam.

Os resultados aqui obtidos são consistentes com os anteriormente aferidos (BIZZOCCHI, 1998), embora novas categorias etimológicas tenham sido postuladas neste trabalho, o que reforça a validade do método para o estabelecimento do espectro etimológico de uma língua e seu uso na análise contrastiva. Esse método, aliás, pode ser estendido às demais línguas europeias e mesmo às não europeias, com a ressalva de que, ao ampliar-se o *corpus* da pesquisa, novos processos etimológicos poderão vir a ser identificados, razão pela qual o modelo teórico aqui apresentado, como ocorre em qualquer pesquisa científica, não tem caráter definitivo, servindo apenas de baliza para futuros estudos.

Referências bibliográficas

ALVES, I. M. **Neologismo**: criação lexical. 3. ed. São Paulo: Ática, 2007. 93 p.

BARBOSA, M. A. Acrograma e sigla: estatuto semântico-sintático e tratamento na obra lexicográfica. In: REUNIÃO ANUAL DA SBPC, 45, 1993, Recife. **Anais**. Recife: Sociedade Brasileira para o Progresso da Ciência, 1993, p. 477.

_____. **Léxico, produção e criatividade**: processos do neologismo. 3. ed. São Paulo: Plêiade, 1996. 323 p.

BIZZOCCHI, A. **Léxico e ideologia na Europa ocidental**. São Paulo: Annablume/FAPESP/UNIP, 1998. 271 p.

_____. Por uma revisão da historiografia tradicional das línguas germânicas: o conceito de *germance*. In: CONGRESSO NACIONAL DE LINGUÍSTICA E FILOLOGIA, 7, 2003, Rio de Janeiro. **Cadernos do CNLF**, ano VII, n.º 4. Rio de Janeiro: Círculo Fluminense de Estudos Filológicos e Linguísticos, 2003, p. 140-146.

_____. A ideologia das raízes. **Língua Portuguesa**, São Paulo, n.º 40, p. 60-63, fev. 2009.

COSERIU, E. **Teoria da linguagem e lingüística geral**: cinco estudos. Rio de Janeiro: Presença, 1979. 239 p.

GUILBERT, L. **La créativité lexicale**. Paris: Larousse, 1975. 285 p.

LINO, M. T. R. da F. Observatório do português contemporâneo. **Actas do Colóquio de Lexicologia e Lexicografia**. Lisboa: Universidade Nova, 1990, p. 28-33.

MAURER JR., T. H. **A unidade da România ocidental**. São Paulo: FFCL-USP, 1951. 227 p.

ANEXOS

Tabela 1: frequência de *tokens* (%)

	PORTUGUÊS		ESPAÑHOL		FRANCÊS		ITALIANO		INGLÊS		ALEMÃO	
	μ	σ^2	μ	σ^2	M	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
1a	33,53	3,72	31,41	5,97	34,74	3,19	31,07	3,84	29,25	4,82	37,67	5,99
1b	3,76	2,21	5,30	3,32	3,08	2,43	4,15	3,14	19,02	2,74	6,27	3,06
1c	2,87	1,15	1,85	0,92	0,38	0,44	2,20	0,69	1,00	0,84	1,86	1,44
1d	2,07	2,24	3,81	2,49	9,16	2,97	5,37	3,80	11,85	4,35	22,84	3,93
1e			0,05	0,11					0,21	0,33	0,03	0,08
1f					0,08	0,17						
TOTAL 1	42,23		42,42		47,43		42,79		61,33		68,67	
2a	11,42	3,02	4,80	2,39	8,15	3,01	8,84	2,96	1,30	0,77	8,83	4,10
2b	0,62	0,25	0,81	0,80	0,41	1,23	0,53	0,57	0,07	0,12	0,47	0,38
2c	7,97	3,39	5,26	1,97	6,30	4,56	10,23	4,02	0,58	0,67	0,98	0,55
2d	0,56	0,98	0,42	0,59	0,02	0,07	0,09	0,18				
2e	0,39	0,48	0,10	0,31			0,08	0,17	0,02	0,05		
2f	1,05	0,98	1,12	1,03	3,45	1,15	1,70	1,08	1,15	0,56	7,72	5,35
2g			0,18	0,36								
2h			0,05	0,15	0,07	0,20						
2i	0,19	0,27	0,40	0,61	0,19	0,45	0,25	0,30	4,22	2,11	0,61	0,77
2j												
2k							0,05	0,11	0,12	0,22		
2l									0,04	0,09	0,06	0,19
TOTAL 2	22,19		13,14		18,59		21,79		7,51		18,67	
3a	29,76	4,24	37,14	3,05	29,12	5,88	30,40	3,47	27,81	5,06	9,09	4,86
3b	0,69	0,64	0,32	0,27	0,63	1,55	0,21	0,38	0,71	1,97		
3c	0,32	0,30	0,81	0,99			0,50	0,59	0,59	0,88	0,18	0,28
3d	3,72	1,54	5,17	2,63	3,00	1,35	3,29	1,66	3,24	2,53	2,21	1,38
3e			0,15	0,45								
3f					0,03	0,09	0,11	0,16				
3g							0,05	0,14			0,22	0,66
TOTAL 3	34,49		43,59		32,77		34,54		32,35		11,69	
4a	0,13	0,25					0,02	0,06	0,07	0,22	0,22	0,45
4b	0,06	0,16			0,05	0,14						
4c	0,62	1,09	0,03	0,08	0,12	0,20	0,11	0,23	0,13	0,21	0,04	0,12
4d			0,02	0,05								
4e												
4f					0,04	0,13	0,02	0,06	0,04	0,11		
4g	0,36	0,85	0,20	0,42	0,39	1,04	0,38	0,47	0,23	0,69	0,03	0,08
4h	0,22	0,61									0,09	0,20
TOTAL 4	1,38		0,24		0,60		0,53		0,46		0,38	
5	0,04	0,11	0,14	0,36	0,21	0,64	0,08	0,16			0,11	0,23

Tabela 2: frequência de *types* (%)

	PORTUGUÊS		ESPAÑHOL		FRANCÊS		ITALIANO		INGLÊS		ALEMÃO	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
1a	30,92	3,05	30,25	4,89	34,45	2,94	30,44	3,11	25,93	3,17	36,11	4,05
1b	3,47	1,63	4,03	2,10	2,61	1,88	3,93	2,93	17,85	4,07	5,43	2,13
1c	2,99	1,26	1,67	0,72	0,41	0,43	2,17	0,68	0,91	0,69	2,02	1,58
1d	2,23	2,06	4,29	2,67	9,70	3,41	5,20	2,52	10,69	3,21	25,14	3,60
1e			0,07	0,15					0,31	0,50	0,04	0,11
1f					0,11	0,21						
TOTAL 1	39,61		40,31		47,28		41,74		55,69		68,74	
2a	11,35	3,59	5,59	3,09	8,91	3,01	9,11	3,13	1,31	0,81	9,82	4,47
2b	0,84	0,33	0,88	0,86			0,54	0,53	0,09	0,18	0,52	0,44
2c	7,59	3,26	5,40	1,98	6,09	3,61	9,97	3,23	0,42	0,61	1,13	0,60
2d	0,28	0,31	0,40	0,45	0,03	0,10	0,10	0,21				
2e	0,42	0,50	0,05	0,14			0,11	0,21	0,02	0,06		
2f	1,17	0,94	1,33	1,09	3,43	1,48	1,89	0,88	1,69	0,45	7,78	2,89
2g			0,20	0,39								
2h			0,06	0,19	0,03	0,10						
2i	0,26	0,37	0,33	0,39	0,35	0,62	0,23	0,31	4,47	1,93	0,69	0,81
2j												
2k							0,08	0,15	0,13	0,20		
2l									0,07	0,14	0,08	0,24
TOTAL 2	21,90		14,25		18,86		22,03		8,20		20,04	
3a	30,87	3,39	37,91	3,75	28,40	2,90	30,33	3,01	29,78	3,70	7,53	3,65
3b	0,89	0,83	0,45	0,39	0,13	0,20	0,64	1,17	0,07	0,16		
3c	0,45	0,42	0,82	0,71			0,61	0,77	0,42	0,43	0,25	0,41
3d	4,50	1,87	5,08	1,85	4,15	1,04	3,21	2,32	4,56	2,48	2,33	1,42
3e			0,19	0,57								
3f					0,04	0,12	0,14	0,21				
3g							0,05	0,16			0,04	0,11
TOTAL 3	36,72		44,45		32,72		34,98		34,83		10,15	
4a	0,18	0,37					0,03	0,10	0,10	0,29	0,26	0,53
4b	0,09	0,24			0,06	0,17						
4c	0,39	0,49	0,04	0,11	0,15	0,24	0,10	0,19	0,18	0,30	0,05	0,14
4d			0,02	0,06								
4e												
4f					0,05	0,14	0,03	0,10	0,05	0,15		
4g	0,39	0,88	0,11	0,17	0,19	0,43	0,45	0,54	0,04	0,11	0,04	0,11
4h	0,12	0,35									0,11	0,25
TOTAL 4	1,17		0,17		0,45		0,92		0,36		0,46	
5	0,06	0,16	0,17	0,40	0,24	0,72	0,10	0,19			0,14	0,28

Artigo recebido em: 03.10.2012

Artigo aprovado em: 08.02.2013