



## HISTÓRIA DAS PALAVRAS: ETIMOLOGIA

Mário Eduardo Viaro (USP)

**Será que a palavra ‘pássaro’ vem de ‘passar’? É melhor pensar duas vezes antes de responder, pois nem sempre a origem das palavras é assim tão fácil. Leia o texto de Mário Eduardo Viaro e aprenda a reconhecer uma “etimologia fantasiosa”.**

As palavras são a manifestação mais visível de qualquer língua.

Aqui se mostra como a palavra é estruturada, como seus sons e sentidos mudam, e como se pode estudar suas origens, ou seja, sua etimologia. E como as palavras guardam nas línguas todo o seu passado, mostra-se com que povos os romanos inicialmente, e os portugueses depois, contactaram e que palavras aprenderam deles.

### Índice

1. [Como reconhecer a raiz de uma palavra?](#)
2. [Por que alguns sons se modificam?](#)
3. [Por que algumas palavras mudam de sentido?](#)
4. [Uma palavra latina pode gerar mais de uma palavra portuguesa?](#)
5. [Como reconhecer uma etimologia fantasiosa?](#)
6. [Algumas palavras são imitações de outras?](#)
7. [Tudo pode ser explicado pelas transformações fonéticas?](#)
8. [De onde vieram as línguas?](#)
9. [Quais as línguas faladas antes do latim em Portugal?](#)
10. [Outras línguas foram faladas em Portugal depois do latim?](#)
11. [Outras palavras influenciaram o vocabulário do português?](#)
12. [Podemos saber de onde vêm todas as palavras?](#)
13. [Novas perguntas](#)
14. [Bibliografia para aprofundamento](#)
15. [Glossário](#)

## 1. Como reconhecer a raiz de uma palavra?

Estudar etimologia requer conhecimentos de muitas línguas e etapas de línguas. O português, por exemplo, tem palavras de origem latina, grega, árabe, tupi, iorubá, entre outras. Além disso, o português medieval não é o mesmo que o do Renascimento ou do Romantismo. Isso promove uma especialização muito grande, uma vez que é possível estudar apenas as etimologias das palavras de origem árabe do português, ou apenas as de origem africana, por exemplo. É possível, ainda, diante de um vocabulário de português medieval, tentar estabelecer etimologias a partir do latim ou fazer reconstruções de suas prováveis etimologias. Poucas coisas trazem mais satisfação e motivação quanto o momento em que entendemos que, por um único radical latino ou grego, dezenas ou até centenas de palavras são explicáveis. Para que esse *insight* ocorra, é preciso que apresentemos inicialmente quatro conceitos básicos: afixação\*, apofonia\*, assimilação\* e a noção de partícipio\*, cujos sentidos nos apressamos a explicar.

Toda palavra tem um núcleo etimológico que é sua raiz\*. Esse termo é sem dúvida alguma uma metáfora botânica. Talvez fosse mais adequado vê-la como um caroço em volta do qual está a fruta. Pequena em relação à palavra, a raiz consiste de uma única sílaba na maior parte das vezes, chegando a ser, não raro, um único som. Há inúmeras discussões a respeito do que vem a ser raiz. Algumas desaparecem ao longo do tempo. Um exemplo disso é a palavra *comer*, cuja raiz latina *ed-* desapareceu completamente, quando se pensa no termo latino que a gerou (ou seja, *comedere*). Um *radical\**, por outro lado, nada mais é do que uma raiz expandida, por meio de pequenas sílabas significativas, ou seja, os afixos\*. Esse processo se chama afixação\*.

Os afixos podem vir antes da raiz e aí serão prefixos\*, ou depois da raiz, ou seja, os sufixos\*. Há inúmeros outros procedimentos (*infixos*, *redobros* etc.), mas eles são de menor importância para estudos de etimologia do português.

Muitas vezes em latim, quando um prefixo se juntava a uma raiz ou a um radical, havia uma troca previsível da vogal da raiz. Esse fenômeno é a apofonia\*. Vital para o latim, a apofonia simplifica centenas de explicações etimológicas em português. Assim, a palavra *sapere* significava “ter sabor” em latim e era formada de uma raiz *sap-* e de uma desinência verbal *-ere*. Com a junção do sufixo *-idus* à mesma raiz desse verbo surge a palavra *sapidus* “que tem sabor”, palavra que também existe em português, embora rara: *sápido*. Mas, ao se juntar o prefixo negativo *in-*, não surge uma palavra *\*insapidus*, mas sim *insipidus* “que não tem sabor, insípido” (Observe que marcamos com um asterisco (\*) anteposto toda palavra cuja existência não é documentada). Ora, o que houve foi a apofonia da vogal *a* do raiz *sap-*, que se tornou *sip-*. A apofonia é sempre constante: *a* se torna *i* em raízes que estiverem em sílabas abertas, isto é, terminando em vogal, mas por outro lado, *a* se torna *e* quando a raiz termina em sílaba fechada, ou seja, terminando em consoante. A raiz *sap-* termina com *-p*, mas essa consoante se junta à sílaba seguinte, reduzindo a raiz, para efeitos de apofonia, somente a *sa-*, que se torna *si-*, ou seja, *sa-pi-dus*, *in-si-pi-dus*. Como se pode ver, a separação silábica não coincide da separação dos componentes da palavra.

Observe que o conhecimento das raízes restabelece a transparência na formação de palavras. Explica, portanto, vocábulos que, muitas vezes, não pareciam interrelacionados ou que, no mínimo, pareciam distantes. Além disso, uma palavra como *sápido* não parece, ao que conhece a raiz *sap-*, uma palavra com um sentido aleatório, como freqüentemente ocorre com as palavras que não nos são conhecidas. Dizendo de outra forma: as técnicas de etimologia ampliam o vocabulário passivo de quem as domina, sem falar que esse conhecimento auxilia problemas de ortografia: sabendo disso, jamais se pensará em escrever *insípido* com *c*.

Vamos para o terceiro termo técnico: assimilação\*. Simplificando bastante os estudos fonéticos, podemos afirmar, por ora, que alguns sons influenciam outros que estão de seu

lado. A vizinhança de um som acaba provocando a descaracterização de um som vizinho, tornando-o idêntico ou parecido a ele. Isso também ocorre na prefixação. Um exemplo disso ocorre com os sons *d* e *g*. O som *g*, quando está ao lado de *d*, por exemplo, num encontro consonantal *dg*, transforma-se em *gg*. A raiz *greg-* significa “rebanho”. Quando se acrescenta o prefixo *ad-*, que significa “aproximação”, forma-se o verbo *ad+greg-are*, ou seja, *aggregare* “juntar todo o rebanho”, “agregar”.

Pode-se resumir essa transformação da seguinte forma: *\*adgregare > aggregare*, usando o tradicional símbolo *>*, que significa “origina”. O símbolo inverso *<* significa “provém de”. Portanto, a mesma afirmação pode ser dita *aggregare < \*adgregare*.

Muito freqüentemente, apofonia e assimilação ocorrem juntas. Uma seqüência como *sf* se assimila em *ff*. Tomemos por exemplo uma palavra como *facilis* “fácil”, derivada da raiz *fac-* “fazer” e do sufixo *-ilis* (ou seja “que se deixa fazer”, “que é feito com facilidade”). Quando se acrescenta o sufixo negativo *dis-* à palavra *facilis*, a raiz *fac-* se transforma em *fic-* (apofonia), além disso, quando o *-s* de *dis-* fica ao lado do *f-* de *fic-*, o som *f-* transforma-o também em *f-* (assimilação). O resultado da junção de *dis+fac-ilis* acaba sendo *difficilis*. Ora, é evidente que as palavras latinas *facilis* e *difficilis* geraram as palavras “fácil” e “difícil” em português, mas sua correlação, para além do sentido de opostos, não estava transparente na forma sem os três conceitos aqui expostos. Além disso, está explicado por que palavras como o inglês *difficult* ou francês *difficile* têm dois *ff* na grafia.

O quarto termo técnico é o particípio\*, corrente quando se aprende conjugação verbal dos verbos em português. Trata-se de uma forma nominal, isto é, um verbo que tem comportamento de adjetivo ou substantivo, por exemplo: *querer* é verbo, mas *querido* pode ser um adjetivo (*querida amiga*) ou substantivo (*meu querido*). O particípio do português termina normalmente em *-ado* ou *-ido*, mas há formas irregulares como *visto*, *posto*, *vindo*

etc. O particípio português corresponde ao particípio passado latino, que era apenas um dos vários tipos que ocorriam em latim.

Voltando ao verbo *sapere*, sabemos que ele passou a significar em latim também “saber”, por um curioso caminho: “ter sabor” é o mesmo que “ter gosto”, daí, figurativamente, “ter bom gosto”, donde “saber das coisas boas” e simplesmente “saber”. De qualquer forma, esse verbo, no particípio ativo presente, gerou *sapiens* “que sabe” (donde o nome científico do ser humano, dado por Lineu em 1758, *Homo sapiens*). Com a regra da apofonia e com o mesmo prefixo negativo *in-*, gera-se *insipiens* “que não sabe”, ou seja, “insipiente”, palavra muito destacada pelos gramáticos de cursos pré-vestibular, que alertam para sua grafia com *s*, distinta da de seu parônimo “incipiente”, que significa “que está começando”, que tem outra raiz, como se verá adiante.

O particípio passado também é responsável por um sem-número de radicais, muitas vezes originados de formas irregulares latinas. Talvez o capítulo dos particípios devesse ser o primeiro a ser tratado pelos professores de latim, dada a quantidade imensa de palavras que podem gerar em português. Para formar particípios passados em latim, normalmente se acrescentava *-atus*, *-etus*, *-itus*, *-tus* ou *-sus*. Assim, o verbo *nare* “nadar” tinha raiz *n-* e formava o particípio passado *natus* “nadado”, a partir da raiz *n-*. Sobre a raiz *n-* surgiu a palavra *navis*, que gerou “nave, nau” e, depois, *navio* em português, pois são todas coisas que nadam. Essa raiz foi expandida para um radical que se compunha da raiz e da desinência de particípio, ou seja, *nat-*, gerando o verbo *natare*, que inicialmente significava “ficar nadando”, mas que, no final, passou a também significar “nadar”, desbancando *nare*. A palavra *natare* gerou um novo particípio, *natatus* “nadado”, cujo radical *natat-* criou um abstrato com um sufixo *-io*: *natatio* “o ato de nadar”, isto é, “natação”. Como veremos, a terminação *-tio* produz sempre *-ção* em português, que é freqüentíssima. Sempre se escreve com *ç* quando a palavra deriva de um *ti* latino.

Outro exemplo: *capere* significa “pegar” e tem raiz *cap-*. O particípio passado desse verbo era *captus* “pegado”. A partir do particípio surge um novo radical, *capt-*, donde se formou o verbo *captare*, inicialmente “ir pegando”, mas depois também “pegar”. Ora, da raiz *cap-* e do radical *capt-* surgem muitíssimas palavras, como veremos mais adiante. Apenas uma será mencionada agora para ilustrar.

O prefixo *ex-*, que significava “para fora”, foi associado com o radical *capt-* causando-lhe apofonia (como é sílaba fechada, o *a* do radical passa para *e*, como previsto acima), donde *exceptus* “pegado (de dentro) e tirado (para fora)”. Em latim é muitíssimo comum encontrar uma imagem visual presente nas palavras em que claramente se descrevem noções como “para cima”, “para baixo”, “junto”, “para perto”, “para longe”. O mesmo sufixo formador de abstratos *-io* (presente em *natatio*) pode ser associado a esse novo radical *except* gerando *exceptio* “o ato de pegar (de dentro) e tirar”. Como se pode ver, é uma definição visual para aquilo que chamamos de “exceção”, afinal, o que é uma exceção senão um fato isolado que se retira de uma regra geral e que não se encontra, portanto, integrado a ela? Como dissemos, formas em *-tio* geram palavras em *-ção* em português (em *-ción* no espanhol, em *-zione* no italiano, em *-tion* no inglês e no francês), o que justifica sua grafia com *ç*, todavia a junção do prefixo com o radical explica também por que essa palavra se escreve com *xc* e não com *ss*, *sc*, *c* etc.

Mas há palavras que são escritas com *-são*. Por quê? O motivo é o mesmo: o radical formado a partir do particípio latino. Alguns particípios são em *-tus* outros em *-sus*. Quando a raiz termina em consoante, às vezes é possível determinar como é o particípio. Se termina em *-p* ou *-c*, normalmente temos *-tus*. Idem se termina *b* ou *-g*, com a diferença que estas mudam para *-p* e *-c*. Se a raiz termina em *-l*, *-t*, *-d* e, às vezes, *-r*, temos *-sus* e a raiz perde as consoantes finais (por isso em português ainda hoje, grande partes dos verbos com raiz terminada em *-d* formam derivados em *-s* e não em *-ç*, por exemplo: *comprender*: *compreensão*, *ofender*: *ofensa*, *ascender*: *ascensão*, *escandir*: *escansão*, *pretender*:

*pretensioso* e não \**compreensão*. \**ofença*, \**ascenção*, \**escanção*, \**pretencioso*). Raízes terminadas em *-n* podem ter participios tanto em *-tus*, quanto em *-sus*. Mas há exceções: raízes terminadas em *-l* com participios em *-tus* e palavras terminadas em *-c* ou *-g* com participio em *-sus* ou *-xus*, sem falar de outros, terminados em *-r* em *-ssus* ou *-stus*.

O verbo *pellere* em latim significava “lançar, atirar”. O participio passado era *pulsus*. Temos novamente a raiz *pel-* e o radical *puls-* com o mesmo sentido. Com o prefixo *ex-*, que, como vimos, significava “para fora”, forma-se o verbo tardio *expulsare* “lançar para fora”, ou seja, “expulsar”. O ato de lançar para fora era formado pelo acréscimo do mesmo sufixo *-io*, donde *expulsio*, que gerou a palavra “expulsão”, com *s*, porque vem de *-sio*, e não com *ç*, porque não vem de *-tio*. Observe que com a raiz *pel-* também há o verbo *expellere* “lançar para fora”, ou seja, “expelir”. Da raiz *pel-*, associada ao prefixo *re-* “para trás”, forma-se a palavra *repellere* “jogar para trás”, ou seja, “repelir”, que também poderia ter um participio presente *repellens* “que repele”, “repelente”. Com o mesmo prefixo, com o radical expandido e com um outro sufixo surgem palavras como *repulsivo* e assim por diante. Parece estranho que antes dessas noções não fosse óbvio que *expulsão*, *expelir* e *repelente* tivessem algo em comum, mas o *insight* do conhecimento etimológico se constrói a partir de dados como esses.

A partir daí, o leitor facilmente percebe famílias de palavras e, tendo dominado a real extensão de aplicabilidade desses quatro fatores e tendo adquirido um vocabulário composto de raízes e afixos (que, curiosamente, não são muitos), teremos uma chave que abre o significado de milhares de palavras em português e de outras línguas, inclusive palavras desconhecidas que se fechavam numa aparente opacidade. Muitas exceções complicadas e aparentes arbitrariedades ortográficas passam a ser meras decorrências da aplicação dessas regras. Quando dominarmos isso tudo, passaremos a ver com mais facilidade o que se passa por trás das palavras.

## 2. Por que alguns sons se modificam?

Muitas vezes percebemos que existem algumas alterações entre a palavra latina e a portuguesa. Por exemplo, *gradus* > *grau*. O desaparecimento de qualquer som no meio das palavras (no caso, um *-d-*, mas poderia ser um *-l-* ou um *-n-*, bem como uma vogal qualquer) é chamado de síncope\*. O mesmo não ocorre com o *-d-* do verbo *aggredi* > *agredir*. Há explicação para isso: todas as palavras que estamos estudando vêm do latim, mas algumas eram populares, enquanto outras não. Quando o Império Romano se fragmentou em reinos, a partir da invasão de outros povos, sobretudo germânicos, no séc. V, apenas as palavras populares permaneceram durante algum tempo, a ponto de se formarem as chamadas línguas românicas. A força que tentava evitar a fragmentação total sempre foi o latim, contudo, agora na sua vertente medieval. O Latim foi única língua escrita da Idade Média até séc. IX na França e até séc. XII em Portugal. O latim medieval, era dominado apenas por alguns, que se baseavam em algum conhecimento dos autores clássicos e tardios e se utilizavam sobretudo do estilo da tradução latina da Bíblia (Vulgata), feita no séc. IV, por São Jerônimo. Mesmo muito tempo depois do surgimento das línguas românicas escritas, o uso do latim continuou sendo usado como língua escrita

Voltaram, portanto, constantemente, para o léxico vernáculo, palavras latinas abandonadas, como se fossem verdadeiros empréstimos. É por isso que algumas palavras, por exemplo, perdem o *-d-* entre vogais como *gradus* > *grau* ou *pedem* > *pee* > *pé* ou *manus* > *mão* ou *color* > *coor* > *cor* são chamadas *palavras populares* (ou *palavras vulgares\**), enquanto as que não o perdem, como *aggredi* > *agredir* são chamadas *palavras eruditas\** (ou *palavras cultas*). Essas modificações (acréscimos, supressões e transformações de sons) são chamadas de *metaplasmos\**. É importante entender o seguinte: o português é uma língua românica (como o espanhol, o francês, o italiano, o romeno etc.), isto é, vem do latim, mas as palavras não vieram todas ao mesmo tempo. Se nos é permitida uma comparação, as palavras populares se assemelham a um sapato velho. Embora gasto, se adapta mais às



peculiaridades dos pés que servem. Um sapato novo pode estar intacto, mas não encaixa direito num primeiro momento. As palavras usadas pelo povo têm longa tradição e são as que mais se modificaram, tanto na forma quanto no conteúdo, no entanto estão perfeitamente adaptadas ao sistema da língua que formaram. As palavras que não modificaram ou sofreram ligeiras alterações são tardias: apareceram a partir do séc. IX, em momentos de revitalização da língua: sobretudo nos séc. XI e séc. XIV, mas principalmente no séc. XVI. Nelas não raro encontramos grupos consonantais incomuns e elementos fonéticos outros que fogem ao sistema da língua em que estão. No séc. XV encontramos a figura do rei de Portugal D. Duarte preocupado em enriquecer o vocabulário do português. No séc. XVI, o número de palavras proparoxítonas, raríssimo até então, aumenta enormemente.

Mesmo tendo perdido para o vernáculo, o latim permaneceu como língua usada no culto católico até meados do séc. XX, bem como língua utilizada na zoologia e botânica, tanto na descrição quanto na nomeação. Existem neologismos calcados no latim ainda hoje, quando se diz sobre fertilização *in vitro*, em *curriculum vitae*, em pós-graduação *lato sensu*, em *campus* universitário e mesmo o *fax*, invenção moderna, que provém da expressão *fac simile*.

Além disso, muitas palavras portuguesas ainda provém de imitações de criações italianas (sobretudo no séc. XVI) e francesas (séc. XIX). Esse é um movimento que podemos chamar de relatinização moderna. Ainda hoje isso ocorre – paradoxalmente – com o inglês, língua que não provém do latim, mas tem um número imenso de palavras latinas herdadas do francês (por causa da invasão normanda, do séc. XI ao séc. XIV). Assim, o inglês tem, para “entre” palavras germânicas (isto é, anglo-saxãs) como *between* ou *among*, mas na confecção de neologismos usa o *inter* latino: *international*, *interplanetary*, *internet*, como qualquer língua românica. Não é de se admirar que a Informática, ciência desenvolvida nos Estados Unidos, ironicamente nos relatinize por meio de expressões como *deletar* (*to*

*delete*, vindo do latim *delere* “destruir”, participio *deletus*), *acessar* (*to access*), *inicializar* (*to initialize*), entre tantas.

Os puristas torcem o nariz para essas palavras e, da mesma forma que eram contra os galicismos até meados do séc. XX, são hoje contrários aos anglicismos. Não fomos nós, brasileiros, que inventamos a Informática, então por que rejeitar os termos do país em que se originou essa Ciência? Também não inventamos o *violino* e o *violoncelo* e usamos esses termos italianos desde o séc. XVI. Não inventamos a *pizza*, a *lasanha*, o *estrogonofe*, o *hambúrguer* ou o *sushi* e não me consta que precisemos de nomes alternativos para eles. Longe de macular uma pretensa pureza lingüística (que obviamente não existe em nenhuma língua do Planeta Terra, nem no chinês, nem no islandês nem em qualquer língua indígena), os estrangeirismos enriquecem o vocabulário das línguas e se hoje o inglês (e não mais o latim) é a língua internacional, isso se deve especialmente ao seu caráter misto e não ao elemento anglo-saxão, que é mínimo.

Houve momentos em que a língua portuguesa exportava palavras para o mundo, como nas grandes navegações: a palavra *pão* ficou no japonês *pan*; a roupa usada pelos navegantes, o *gibão*, também ficou: *zubon*. A fauna e flora da Ásia e da América também foram divulgados por intermédio da língua portuguesa: o mamífero *tatu* aparece no italiano e no francês (respectivamente grafados *tatù* e *tatou*), o peixe *piranha* aparece nos dicionários de inglês, francês, alemão e italiano. A palavra *urubu* ocorre em francês e em italiano. O mesmo se pode observar do *jacarandá*, que aparece nos dicionários franceses, italianos, alemães e ingleses. A fruta indiana que deu origem ao nosso *jambo* também chegou ao italiano *giambo* pelo português. Também em turco, *palavra* (e em alemão *Palaver*, romeno *palavră*, albanês *palavi*) significa “tagarelice, mentira”. Tomando por base o dicionário Zingarelli da língua italiana, encontraremos outros termos da língua portuguesa em processo de internacionalização: *autodafê*, *bossa nova*, *favela*, *samba*, *viado*. Por isso, antes

de lançarmos batalhas quixotescas contra invasões de termos estrangeiros, é preciso observar que também exportamos palavras para as demais línguas do Globo.

### **3. Por que algumas palavras mudam de sentido?**

Dissemos, inicialmente, que as palavras se compõem de raízes, que podem ser seguidas de sufixos e/ou antecedidas de prefixos. Cada componente tem sentidos básicos e sentidos derivados. Além disso, a forma das palavras segue padrões, que são os chamados metaplasmos.

Quem imaginaria que, na linguagem falada, *cara* pudesse um dia vir a significar “pessoa” (e ainda por cima no gênero masculino: *um cara*)? Todavia, a palavra *pessoa*, por sua vez, sofreu um desenvolvimento semântico parecido, pois veio de *persona*, que significava “máscara”. Quem poderia imaginar que o milenar *tu* (que remonta ao indo-europeu) seria substituído completamente em alguns Estados brasileiros, a partir do séc. XVI, por um modismo italiano, a saber, *Vossa Mercê*, origem de nosso *você*? Quem imaginaria que a palavra latina *testa*, que significava “vaso, telha, caco”, passaria a significar “cabeça” em francês e italiano (respectivamente *tête* e *testa*) ou a fronte em português? Por outro lado quando se diz que fulano “faz o que lhe dá na telha” é exatamente o mesmo que se diz.

O estudo das mudanças semânticas é fascinante, mas bastante complicado e, por isso, protelado ou abandonado por muitos que se dedicam ao estudo da linguagem.

Já com relação à forma, pode-se fazer muitas generalizações. Por exemplo, as *consoantes duplas* latinas (as chamadas *geminadas*), com exceção de *ss* e *rr*, simplificaram no português: *aggregare* > *agregar*. As vogais têm transformações bem específicas: *a* não se modifica, nem *e* ou *o*, mas *i* se transforma em *e*: *recipere* > *receber*, e *u* se transforma em *o*: *conceptus* > *conceito*. Também as consoantes entre vogais (tecnicamente chamadas de

*intervocálicas*) costumam mudar: *p* se torna *b*, ao passo que *t* se torna *d*. Essas duas transformações se chamam sonorizações\*. Chama-se sonorização a transformação de um som surdo num sonoro. É relativamente fácil saber quais sons são surdos e quais são sonoros, pois colocando os dedos no pomo-de-adão, estes vão vibrar quando pronunciamos os sons sonoros, mas não vibrarão quando pronunciados os sons surdos.

Foi visto já que a terminação *-io* acaba se transformando em *-ão* e se antes houver um *-t*, o conjunto *-tio* se torna *-ção*: *occupatio* > *ocupação*. Dessa forma, o *t* se torna *ç*. Esse fenômeno se chama lenização\*. Lenização seria a transformação de um som plosivo para outro fricativo. Plosivas são todas as consoantes que, na sua produção, impedem totalmente a passagem do ar vindo dos pulmões, como o *p*, *t*, *b*, *d*, *c* antes de *a*, *o*, *u* e o *g* antes das mesmas vogais, bem como *qu* ou *gu* antes de *e*, *i*. As demais consoantes, exceto *m*, *n*, *nh*, *l*, *lh* são fricativas, pois deixam o ar passar à medida que são pronunciadas. Outro exemplo de lenização ocorre quando o *b* se transforma em *v*. A letra *c* latina representava sempre o mesmo som, a saber, *k*, de modo que *ca*, *ce*, *ci*, *co*, *cu* se lia *ka*, *ke*, *ki*, *ko*, *ku*. Essas sílabas, quando iniciais, passaram a ser pronunciadas, em português, *ka*, *se*, *si*, *ko*, *ku*: houve permanência do som antes de *a*, *o*, *u*, mas houve lenização antes de *e*, *i*. Entre vogais, essas mesmas sílabas se transformam em *ga*, *ze*, *zi*, *go*, *gu* ou seja, houve sonorização e no caso de *ze*, *zi*, houve, ainda por cima, lenização (por exemplo *dicere* se transforma em *dizer*). O mesmo se pode afirmar da letra *g*: a seqüência *ga*, *ge*, *gi*, *go*, *gu* se lia *ga*, *gue*, *gui*, *go*, *gu* em latim. Em posição inicial, o português lenizou as sílabas com *e* e com *i* e hoje essa seqüência se lê: *ga*, *je*, *ji*, *go*, *gu*. Em posição intervocálica, o que ocorre, porém, é a síncope (*ga*, *e*, *i*, *go*, *gu*), por exemplo, *regem* se tornou *\*ree* e depois *rei*.

Percebemos ainda que outras transformações não são sistemáticas: o grupo *-ptio*, que aparece em várias palavras, às vezes mantém o *p*: *perceptio* > *percepção*, às vezes o elimina por síncope: *exceptio* > *exceção*. Outras vezes, ainda, ele se transforma em *i*: *conceptio* > *conceição*. Não se trata de transformações caóticas. Na verdade, a manutenção

do *p* é marca da inquestionável origem erudita da palavra. Em Portugal, escreve-se *excepção*, embora não se pronuncie esse *p*, mas tão pouco se pronuncia o *p* de *percepção* ou de *decepção*.

A escrita muitas vezes interfere na pronúncia: isso faz com que pessoas pronunciem a palavra *adquirir* como “adqüirir”. Um caso curioso é o da palavra *sob*, que vem do latim *sub*. Com o tempo, devido à dificuldade de se pronunciar um *-b* final, *sub* > *so* e essa forma (*so*) seria a única até século XVI (aparece inclusive como prefixo em *soterrar*, *socapa* etc.), mas com preocupações etimológicas (as mesmas que faziam escrever palavras como *pharmácia*, *asthma*, *hypochondríaco*, *hippódromo*, entre outras), resolveu-se restaurar o *-b* final de *sob*, que não era pronunciado. Passado o tempo, a palavra *sob* foi substituída na língua falada (e, depois na escrita) por outras expressões mais encorpadas, de modo que se deixou de falar *sob a mesa*, para se falar *debaixo da mesa*. Por causa da queda da frequência de uso, a palavra ficou rara e quando aparecia em textos, as pessoas passaram a lê-la pronunciando o *b*. A conclusão que se tira daí é que a ausência ou presença de *p* nesses casos é problema histórico: depende de quando a palavra entrou no português, pois as palavras latinas não entraram no português todas na mesma época. Já a passagem de *p* para *i*, mostra outra coisa: que se trata de palavra erudita antiga, que se popularizou e tem uma forma mista, que não é nem erudita, nem popular. De fato, a religião católica teve, ao longo de séculos, cunho muito popular e, mesmo que uma palavra tivesse entrado no repertório da linguagem cotidiana por meio da erudição dos clérigos que sabiam latim, logo se popularizava. Essas palavras de forma mista são chamadas semi-eruditas\* e há várias delas, sobretudo no âmbito eclesiástico.

É importante ressaltar que a distinção entre palavras eruditas e populares não segue a frequência de uso atual. Há palavras eruditas que continuam com aspecto culto, sendo desconhecidas da população sem estudo, já outras se popularizaram de tal forma impressiona saber que tiveram origem erudita. Por outro lado, palavras de origem popular

podem deixar de ser usadas e se tornar arcaicas. Nesse caso, passam a ser desconhecidas da grande maioria das pessoas, tornando-se muitas vezes palavra de jargão ou regionalismo. Como exemplo podemos citar uma palavra como *último*. Como praticamente todas as proparoxítonas, essa palavra não veio por meio da linguagem popular. De fato, não existe nos textos muito antigos, sendo suas primeiras atestações no séc. XIV, quando foi introduzida, para imitar o latim *ultimus*. Em seu lugar, os textos do séc. XII e XIII dizem *derradeiro*, essa sim, palavra popular. O uso dessas duas palavras é bastante variável nas regiões do Brasil, de modo que há lugares que só usam *último*, sendo para esses falantes, a palavra *derradeiro* um regionalismo ou uma palavra arcaica; em outros, porém, *derradeiro* é mais freqüente, embora ninguém desconheça o sentido de *último*. Seguindo a linha das proparoxítonas, também as formas superlativas com *-íssimo* ou *-érrimo* eram praticamente desconhecidas até século XVI e entraram no português, por imitação do italiano, para revitalizar um uso que era do latim *-issimus* ou *-errimus*. Hoje em dia, essas formas muito cultas se tornaram de uso comum do povo, de modo que se criam palavras bastante esdrúxulas como *grandessíssimo*, *coisíssima nenhuma*, *bacanérrimo* e na Folha Ilustrada, p. E7, de 10/01/03 encontramos “modernerrérrimos e montaderrérrimos”. O inverso ocorre com a palavra *grei*, vista no capítulo anterior, que apesar de originalmente popular, se tornou arcaica e presente apenas em textos cultos, como o da Bíblia.

Dessa forma concluímos que há palavras de origem culta de uso culto e palavras de origem culta de uso popular, bem como palavras de origem popular de uso culto e palavras de origem popular de uso popular. Nosso trabalho é sobre a origem, não sobre o uso, então, quando falarmos que uma determinada palavra é culta ou popular, é sobre sua etimologia que falamos. E a base para sabermos distinguir é, muitas vezes, os metaplasmos. Como já dissemos, as palavras cultas são mais conservadoras quanto à forma: *p* entre vogais continua como *p*, não passa para *b*; da mesma forma, *t* entre vogais continua como *t*, não passa para *d*, por fim *d* entre vogais continua como *d*, não cai por meio da síncope.

Nas palavras cultas, apenas algumas adaptações ocorrem: palavras terminadas em *-us* passam a *-o* no português: *egregius* > *egrégio*, palavras em *-io* passam a *-ão*, mas no português antigo já foi *-om*: *deceptio* > *decepção*. Os infinitivos em *-re* passam a *-r*, eliminando a última vogal, metaplasmo conhecido como *apócope*: *aggregare* > *agregar*. A terminação *-bilis*, que já foi *-bel* ou *-ble* no português arcaico, hoje em dia é *-vel*: *perceptibilis* > *perceptível*. A terminação *-tas* se torna *-dade*. *capacitas* > *capacidade*. A terminação *-ns* se torna *-nte*: *ingrediens* > *ingrediente*.

O leitor começa a perceber que a maior parte das palavras que apresentamos até aqui são formas eruditas. Isso quer dizer que, não só é verdade que nossa língua vem do latim, mas também que ela praticamente é latim, muitas vezes via francês, via italiano ou via inglês. O intercâmbio lingüístico que se iniciou na Idade Média unificou o vocabulário dessas línguas, aproximando-as, muito mais do que o simples fato de serem línguas vindas do latim. O romeno também provém do latim e não participou desse intercâmbio até século XIX, daí seu aspecto peculiar, importante para estudos de reconstrução do latim popular (MAURER Jr. 1951). O inglês não provém do latim, mas participou desse intercâmbio, de modo que seu vocabulário é extensamente românico. O alemão, o sueco, o dinamarquês, o russo participaram em parte e as influências estão muito escondidas, como veremos. O que promoveu esse intenso intercâmbio foi, sem dúvida, no primeiro momento, o Cristianismo e, depois, a Revolução Industrial.

#### **4. Uma palavra latina pode gerar mais de uma palavra portuguesa?**

Tanto a palavra *sino* quanto a palavra *signo* têm o mesmo étimo: vêm de *signum*. Da mesma forma, tanto *sina* quanto *senha* vêm de *signa*. O mesmo se pode dizer de *desenhar* e de *designar*, que vêm de *designare* bem como de *resenhar* e *resignar*, que vêm de *resignare*. Esses pares são chamados de formas divergentes\* ou *doublets*, isto é, uma única palavra latina pode originar duas ou mais palavras na mesma língua. Novamente, isso não

seria possível de acontecer na mesma época, a menos que essas duas formas fossem variações regionais. Não é esse o caso. Uma das formas é mais antiga que a outra e é possível medir sua idade pelos metaplasmos. Assim, o encontro consonantal *-gn-* pode se transformar em *-nh-*, *-in-* ou *-n-* no português ou continuar como *-gn-*. As três primeiras formas são mais antigas, pois estão ligadas ao latim popular, ou seja, a variante do latim que continuou sendo falado ininterruptamente, de geração em geração, até se transformar ao longo dos séculos nas línguas românicas. A conservação da forma aponta para o latim clássico, que deixou de ser falado espontaneamente quando da caiu o Império Romano, no século V, sendo continuamente retomado durante o período medieval e pós-medieval, sob a forma de palavras cultas. Da mesma forma, o *i* se transformou em *e* no latim popular, mas nas palavras cultas continua como *i*. As transformações das vogais atendem pelo nome de *vocalismo*, ao passo que as das consoantes formam o *consonantismo*. É fácil agora perceber que *senha*, *desenhar* e *resenhar* são palavras populares, enquanto *signo*, *designar* e *resignar* são palavras cultas. Sobram *sino* e *sina*, que têm vocalismo culto, mas consonantismo popular: são as chamadas palavras semi-eruditas. Não pertenciam ao latim popular, pois ressurgiram numa época posterior à queda do Império, porém se popularizaram rapidamente. Novamente, são formas ligadas à Igreja.

As formas divergentes são numerosas e algumas bastante surpreendentes. Há um metaplasmo curioso em que o encontro *pl-* inicial se torna *ch-* em português, *ll-* no castelhano, *pi-* no italiano, mas permanece *pl-* no francês, de modo que a palavra *pluvia* deu origem ao português *chuva*, ao castelhano *lluvia*, ao francês *pluie*, ao italiano *pioggia*. A palavra *planus* latina, portanto, estava submetida a essa transformação e, de fato, temos, como derivada dela, a palavra *chão*, do português, que corresponde a *llano* no castelhano, a *plain* no francês e a *piano* no italiano. Assim, temos, de *planus* latino, cinco formas divergentes em português: *plano* (forma erudita, reconstruída durante a Idade Média, com ligeira adaptação da terminação e manutenção do significado), a *chão* (forma popular, em que se seguem todos os metaplasmos e com mudança semântica, uma vez que o chão



normalmente é plano), *lhano* (empréstimo do castelhano, com mudança semântica para “sincero”, isto é, sem torneios, sem voltas, metáfora de algo plano), *piano* (empréstimo do italiano, como redução da palavra *pianoforte*, com menção às duas dinâmicas musicais) e *porão*, antigo *prão* (forma semi-erudita, com mudança semântica, referindo-se ao chão nivelado desse aposento).

Inversamente, há formas convergentes\*, ou seja, duas ou mais formas latinas podem adquirir uma única forma nas línguas românicas. Assim a terceira pessoa do plural do verbo “ser”, *sunt*, gerou no português antigo a forma *som*. Da mesma forma *sanctus* “santo” gerou, paralelamente a *santo*, a forma apocopada *sam*. Por fim *sanus* “saudável” gerou a palavra *são*. Por volta do séc. XVI, as terminações *-om*, *-am* e *-ão* se convergiram numa única forma: *-ão*, de modo que as três palavras hoje em dia são pronunciadas *são*. Isso é flagrante nas formas do plural. O latim *panis* gerou, por meio da forma *panem* (tecnicamente chamada *acusativo*), a forma *pan* (com apócope do *-m* e depois do *-e*). A palavra *visio*, no acusativo *visionem*, gerou a palavra *vison* (síncope do segundo *-i-*, apócope do *-m* e depois do *-e*). A palavra *manus*, no acusativo *manum*, gerou *mão* (síncope do *-n-* intervocálico, com nasalização da vogal, apócope do *-m* e transformação do *u* em *o*). Tudo isso parece extremamente regular segundo as regras dos metaplasmos. No entanto, no séc. XVI, talvez segundo a influência da pronúncia lisboeta, *pan* se tornou *pão*, *vison* se tornou *visão* e *mão* continuou como era. Em castelhano, as terminações continuam como no português antigo: *pan*, *visión*, *mano*. O plural, no entanto, criou uma irregularidade, de modo que o acusativo dessas três palavras era respectivamente *panes*, *visiones*, *manus* (bem próximo do castelhano: *panes*, *visiones*, *manos*). O português elimina, por meio de um metaplasmo chamado *apócope*, todas as consoantes finais, com exceção do *-s*. Não caindo o *-s*, da mesma forma que caiu o *-m*, vemos que não cairá também o *-e* final, de modo que o *-n-* continua na posição intervocálica, o que o faz cair e nasalizar a vogal anterior (como em *manum* > *mão*). Portanto: *panes* > *pães*, *visiones* > *visões*, *manus* > *mãos*.

Por causa desse acidente fonético da história da língua, a convergência dos sufixos no singular não ocorreu no plural, de modo que podemos afirmar que os metaplasmos de uma palavra no singular não são necessariamente idênticos aos dessa mesma palavra, no plural. Irregularidades na língua atual foram, com frequência, regularidades num estágio mais antigo: atualmente se diz que uma palavra portuguesa terminada em *-ão* pode ter plural em *-ãos*, *-ões* ou *-ães* e é uma dificuldade que a Gramática Normativa oferece aos estudantes. Nem sempre essa situação foi tão complicada: no século XIII se poderia afirmar tranquilamente que as palavras em *-ão* tinham plural em *-ãos*, as em *-an* faziam plural em *-ães* e o plural das em *-on* era *-ões*.

## **5. Como reconhecer uma etimologia fantasiosa?**

Porém, nem sempre os estudos etimológicos foram assim tão científicos. Na busca do étimo\* (termo que vem do grego *étymos* “verdadeiro”), muitos autores não se preocupam com as irregularidades das mudanças formais e muito menos com os percursos sinuosos das explicações para a mudança do significado. No diálogo intitulado *Crátilo*, de Platão (V a.C.) e nos textos de Isidoro de Sevilha (VII d.C.) são comuns soluções muito fantasiosas para explicar a mudança do significado. Ainda hoje abundam nas bancas livros repletos de explicações mirabolantes para o significado dos nomes próprios (referimo-nos a publicações do tipo *Mil nomes para seu bebê*). No séc. XVI, Fernão de Oliveira, primeiro gramático da língua portuguesa, desconfia de toda e qualquer etimologia e apresenta seu ceticismo com relação à palavra *pássaro*, que, segundo alguns, viria de *passar*. De fato, Oliveira estava certo: *pássaro* não vem do verbo *passar*, mas do acusativo latim *passerum* “pardal”. O verbo *passar*, por sua vez, vem de um suposto *\*passare*, derivado de *passum*, que quer dizer “passo”. Há, portanto, etimologias fantasiosas e etimologias embasadas em dados. Se *pássaro* viesse de *passar*, por que o castelhano diria *pájaro*, se também tem um verbo *pasar*?

Normalmente quem pratica etimologias fantasiosas imagina que todas as palavras usadas na língua são derivadas de outras que também usadas atualmente. Não se concebem os séculos necessários para a formação de um vocábulo, nem a situação discursiva pela qual a palavra teria entrado no vocabulário. Não imagina que as mudanças fonéticas tenham de ser mais ou menos sistemáticas para não se caracterizar uma explicação *ad hoc*. Não se pensa, enfim, que a palavra vem de uma etapa anterior. Segundo os etimólogos fantasistas, as palavras do português moderno viriam do próprio português moderno e não do português antigo. É antológica a explicação para a palavra *news* do inglês. A etimologia é límpida: trata-se de um decalque da palavra *novas* do latim (ainda hoje se diz em português: *boas-novas*) e uso de uma terminação de plural num adjetivo, sempre invariável no inglês, torna essa explicação quase evidente. Mas houve quem propusesse que a palavra *news* fosse uma espécie de sigla (ou acróstico) de *North, East, West, South*, algo como “as mensagens vindas dos quatro cantos do mundo”. Hoje em dia é comum formarem-se palavras com siglas, mas, definitivamente, isso não é comum em palavras antigas. A área da etimologia é uma das mais invadidas por opiniões pessoais. Explica-se (e muitas vezes divulga-se) que o bairro de *Campeche* em Florianópolis viria de “com peixe”; explica-se que o nome próprio *Elvira* é uma corruptela de “alvura”, mas faltam os documentos que o provem. Pior que isso: falta método, pois fica difícil saber por que o *o* se transformou em *a* e por que o ditongo *ei* (ou no máximo a vogal fechada *e*) se tornou a vogal aberta *e* no primeiro caso e, quanto ao segundo, por que as vogais mudaram tanto. Também é difícil de imaginar a situação discursiva em que tais palavras se formaram.

No caso de *passer*, observamos que a palavra significava “pardal” e passou a significar simplesmente “pássaro”, de modo que, quando dizemos hoje que *o sabiá é um pássaro* não se afirma que o sabiá seja um pardal. Tal procedimento se chama generalização\*. Há muitos exemplos disso. A generalização vem de uma aplicação inicialmente indevida que se populariza. Quando os portugueses vieram para o Brasil, isso aconteceu com muita frequência com os animais do mundo novo que se descortinava. Vendo o *tapir*, chamaram-

no de *anta*, espécie de cervo sem chifre que conheceram na África e que julgavam ser o mesmo ou parecido. Ainda hoje há quem chame o *gambá* brasileiro de *raposa* (que é européia) ou o *urubu* de *corvo*. Animais bem distintos que se mesclaram no imaginário. Fernão Cardim, tendo contato pela primeira vez com um *caju*, descreve-o como uma espécie de maçã. De fato, a extensão dos nomes aumenta, de modo que vale aqui dizer que a metáfora atuou usando um nome velho para uma coisa nova.

Ocorre também o contrário: muitas palavras que eram gerais se tornam específicas e esse procedimento se chama especificação\*. O caso mais freqüente é o de nomes próprios que se tornam genéricos. Assim diz-se que *Carrasco* era o sobrenome de um sanguinário algoz na Idade Média (a saber, Belchior Nunes Carrasco); a *guilhotina* deve seu nome a um certo médico Guillotin, do séc. XVIII. É bastante conhecido o fenômeno de marcas registradas que se generalizaram para o produto que vendiam: *danone* para “iogurte”, *bombril* para “palha de aço”, *gilete* para “lâmina de barbear”.

É comum ainda que uma das acepções se torne mais usual que outras, suplantando o sentido básico. Assim, o particípio *complicatus* vem do verbo latino *complicare*, que significava “fazer muitas pregas”, de modo que isso se tornou, por metáfora, “fazer algo difícil”. Por oposição *simplex* significava “com uma só prega”, ou seja, “algo fácil, algo simples”. Nada sobrou do sentido básico, apenas do metafórico, de modo que vemos não só no português o par *complicado*: *simples*, mas também no francês *compliqué*: *simple*, no inglês *complicated*: *simple* e até o alemão tem *kompliziert*. Muitas palavras mudam de sentido, num processo de *pejoração*: em Portugal, *esquentar* aos poucos se torna palavra-tabu. Aqui no Brasil, isso ocorre com freqüência: a palavra *zona*, absolutamente neutra tornou-se restrita à acepção *zona de meretrício*. Palavras como *sujeito* e *indivíduo* começam a ter valores negativos, a ponto de serem substituídos por sinônimos em alguns textos. Nem sempre o sentido básico, presente como primeira acepção dos dicionários, é o mais usual, pois o mais corriqueiro é o que sobreviverá.

Estudar etimologia é vasculhar esses meandros tortuosos da mudança do significado, mas isso deve ser feito com critério e, principalmente, com abonações que atestem as suas afirmações. Isso é difícil de fazer, portanto raros são os trabalhos confiáveis nessa área do conhecimento. Normalmente a explicação mais popular se confunde com a mais convincente.

Também é delicada uma outra questão: a da reconstrução. O verbo *\*passare* acima não se encontra nos textos latinos. Como saber se ela realmente existiu? Ora, a forma *passar* do português não está isolada, pois há, como dissemos, *pasar* em castelhano, *passer* em francês. Se conseguirmos provar que essas formas são independentes, ou seja, que a forma portuguesa não vem da francesa, por exemplo, é possível criar uma forma hipotética *\*passare* por meio dos metaplasmos, fazendo uma operação inversa. A forma reconstruída, portanto, precisa ter abrangência\*, ou seja, valer para mais de uma língua e para seus metaplasmos particulares. Também precisa ter sistematicidade\*, isso é, deve ser um fenômeno comum à língua e, de fato, *\*passare* é um verbo da primeira conjugação formado sobre o radical do participio, fenômeno bastante comum no latim tardio (assim se explica também *\*visare*, de *avisar*). Por fim, é preciso haver coerência\* na transmissão dos metaplasmos e, de fato, a passagem de *a* para *e* nos verbos da primeira conjugação do francês é regra e não exceção. Se satisfizer as três condições, a etimologia reconstruída tem quase a mesma força de um dado, mas não é incomum encontrar nos dicionários etimologias consagradas que não obedecem a uma dessas condições ou mais. Assim, a forma *\*ad tenus* para justificar a preposição *atens* do português antigo (atual *até*) não satisfaz duas delas: *tenus* só ocorre no português e em nenhuma outra língua vinda do latim (é, pois, um étimo não-abrangente), uma forma adverbial (no caso, *tenus*) antecedida de *ad* é muito recorrente nas etimologias (satisfaz a condição da sistematicidade) e, por fim, as transformações por metaplasmo gerariam *\*ateios* (étimo não-coerente). Apesar de *\*passare*

e *\*ad tenus* serem marcadas como formas hipotéticas, não há o mesmo grau de certeza na reconstrução de ambos os étimos.

Outro grave problema que percorre a etimologia é a datação e a abrangência dessas formas. Hoje em dia, por exemplo, ninguém negaria que o romeno seja uma língua românica assim como o português, castelhano, catalão, francês, italiano. Mesmo para quem não entenda essa língua, bastará correr os olhos sobre um texto para verificar isso. No entanto, muitíssimas palavras parecidas com o português são na verdade inovações do séc. XIX, inspiradas no italiano e, sobretudo, no francês. Da mesma forma, o inglês tem muitíssimas palavras herdadas do francês. O que nos faz afirmar que o inglês seja uma língua germânica (derivada do anglo-saxão) e o romeno não seja uma língua eslava, a despeito do grande número de palavras derivadas do tronco eslavo? Conta aqui mais a estrutura interna do que o vocabulário, ou seja, os artigos, os pronomes, os numerais, a conjugação verbal, as flexões em geral, as preposições e as conjunções são claramente germânicas e não latinas no caso do inglês e obviamente latinas e não eslavas no caso do romeno. Assim também, apesar de grande número de palavras árabes no persa, a sua estrutura é indo-européia e não semítica. A despeito do grande número de palavras árabes, tupis e africanas no português, sua estrutura interna é basicamente latina. Há, portanto, que se considerar a diferença entre herança e empréstimo. As palavras herdadas do português seriam todas do latim, já os empréstimos são de qualquer língua, inclusive das línguas vindas do latim. Os termos eruditos que foram estudados acima são verdadeiros empréstimos do latim. O número de empréstimos numa língua pode ser muito maior do que o das palavras herdadas. Em que século, porém, os empréstimos entram numa língua é um problema. O fato de entrarem em mais de uma língua é outro.

O estudo das etimologias sempre gerou forte interesse e fascínio da parte de todos, tanto de estudiosos de Letras quanto de pessoas de outras áreas. Por isso, surgem amiúde algumas obras sobre o assunto, de variada qualidade: desde as que têm certo fundamento, mas são

confusas e crípticas, até as que são totalmente não-fundamentadas, pois seguem a linha das etimologias fantasiosas.

No entanto, é nossa postura constante preferir não indicar uma etimologia a apresentar algo duvidosamente reconstruído. O etimólogo que tem respostas para tudo se aproxima do charlatão, uma vez que estamos lidando muitas vezes com línguas ágrafas, das quais não temos documentação suficiente. Erros grosseiros em reconstruções, infelizmente não incomuns, acabam por dismantelar todo cuidadoso edifício dos estudos históricos, construído por muitos pesquisadores durante mais de dois séculos. Dúvidas e desconfianças sobre o rigor do trabalho dos etimólogos surgem, contudo, desde há muito tempo. Fernão de Oliveira (1536), por exemplo, brilhante precursor da perspectiva sincrônica da Lingüística Moderna, em sua *Grammatica da lingoagem portuguesa* chama o estudo etimológico de *patranha sobeja*, isto é “mentira desnecessária”:

*Ora pois, se como adivinhando dixéremos que **homem** se chama porque é o meio de todas as cousas ou porque está no meio do mal e do bem; e se dixéremos que **molher** se chama porque é molle e **velho** porque vio muito; e **antigo** porque foi antes d'agora, e **tempo** porque tempera as cousas; (...) e também **escrever** quasi discretamente ver; e **alfaiate** porque faz alfaias, e **passaro** porque passa voando (...) e assi com' estas, podemos também cuidar outras dozentas patranhas, as quaes sempre são sobejas e muitas vezes falsas, e pouco recebidas antre homens sabedores, que do pouco que com muito lendo e trabalhando aqueriram se prezam e não de imaginação aldeãs sem juizo. (cap. XXXI)*

Ao se fazer etimologia, de fato, hoje em dia, alguns ainda buscam explicar tudo, sem se valerem de abonações em textos confiáveis. É fácil imaginar explicações para expressões como *estar na pindaíba*, *onde Judas perdeu as botas*, *o diabo a quatro*, *fazer nas coxas* e tantas outras. É divertido, mas não deve ser o único trabalho de um etimólogo.

Ao se fazer etimologia é sempre preciso diferenciar a língua estudada da língua-fonte. Os passos de uma etimologia confiável podem ser resumidos como:

- 1) conhecimento da estrutura e do léxico das línguas-fonte;
- 2) conhecimento do momento histórico do contato da língua-fonte com a língua estudada;
- 3) datação dos textos que comprovem os étimos ou que justifiquem a sua reconstrução;
- 4) conhecimento dos metaplasmos regulares da passagem da língua-fonte para a língua estudada;
- 5) conhecimento da freqüência de uso das palavras nas línguas em questão.

## 6. Algumas palavras são imitações de outras?

Decalque seria uma espécie de tradução literal para o vernáculo dos elementos que compõem um determinado vocábulo, como encontramos na palavra *cachorro quente*, composição que não foi criada no português, mas montada por decalque a partir do inglês *hot dog*. A força do latim foi tão grande na Idade Média, que, muitas vezes, línguas que não são românicas escondem a sua influência. Existem verdadeiras traduções ao pé da letra de alguns vocábulos. Assim, *incluir* e *excluir* remontam ao latim *includĕre* e *excludĕre* e estão presentes em muitas línguas: inglês *to include: to exclude*, no francês *inclure: exclure* etc. A raiz dessas palavras é *clud-*, forma apofônica de *cludĕre* “fechar”, que está na base da palavra *clavis* “chave”. Literalmente, *includĕre* significa “fechar por dentro” e *excludĕre* “fechar (deixando) para fora”. Essa imagem se encontra no alemão: *einschließen: ausschließen* (derivados de *schließen* “fechar”, mesma raiz de *Schlüssel* “chave”, com *ein-:aus* equivalentes a *in-:ex-*), no russo: *vključit’: izključit’* (derivados de *ključ* “chave”, com *v-: iz-* também equivalentes a *in-:ex-*), no dinamarquês *indelukke: udelukke* “excluir” (com *inde-* equivalente a *in-*, *ude-* equivalente a *ex-* e *lukke* “trancar”) e mesmo no islandês, cioso de não possuir estrangeirismos, encontram-se *loka inni*



“incluir” e *loka úti* “excluir” (literalmente “trancar para dentro” e “trancar para fora”). Até uma língua não-indo-européia, como o húngaro teve influência das mesmas imagens: *kizár* “excluir” (com raiz *zár* “fechadura, fecho” e *ki-* equivalente a *ex-*) e *bezárólag* “inclusive” (*be-* equivalente a *in-*). Por sua vez, a forma latina *excludere* seria uma imitação do grego *ekkleío* (com *ek-* equivalendo a *ex-* e *kleís* “ferrolho, chave”). Já havíamos visto que existem palavras internacionais. Agora vemos que há imagens internacionais, que, por contato cultural, uma língua faz empréstimos de outra. O alemão e o húngaro podem não ser línguas vindas do latim, mas depois que esses povos se cristianizaram havia bastantes pontos em comum no pensamento, que justificassem vocábulos desse tipo, a ponto de não reconhecermos, à primeira vista, numa palavra como *Barmherzigkeit* do alemão um decalque de nossa conhecida palavra latina *miser cordia*, ou seja, *(er)barm(en) = miser(i)-, Herz = cor(d)-, -ia = -keit*.

Alguns decalques são realmente difíceis de serem percebidos. Quando o *trem* foi inventado, no séc. XIX, a palavra inglesa *train* acabou vindo ao português do Brasil na forma de empréstimo. Em alemão, a palavra é *Zug*, aparentemente, não tem nada a ver com esse étimo. Porém, da mesma forma que *train* remonta ao verbo *traîner* do francês, que significa “puxar”, *Zug* é decalcado sobre o verbo *ziehen* (no pretérito: *zog*), da mesma etimologia. No búlgaro, *vlak* “trem” é derivado do verbo *vleká* “puxar”. No húngaro, também *vonat* “trem” pertence à mesma raiz do verbo *von* “trazer”. As nações são orgulhosas de suas línguas, mas o patrimônio comum é maior do que o particular.

## **7. Tudo pode ser explicado pelas transformações fonéticas?**

A analogia\* é um fenômeno complexo que contraria a ordem normal dos metaplasmos. Se as línguas se alterassem só por meio dos metaplasmos, haveria um alto grau de previsibilidade, mas o que ocorre é que há modelos que caracterizam subconjuntos numa

língua, de modo que as coisas novas vão se encaixando neles, a fim de se evitarem as irregularidades.

Um exemplo: *desejo* vem, por meio dos metaplasmos, de uma forma \**desidium* (como castelhano *deseo*), que por sua vez mescla o elemento semântico de *desiderium* com a forma da palavra *desidia*. A palavra *desiderium*, que gerou o francês *désir*, italiano *desiderio*, não tem a raiz *sid-*, mas a raiz *sider-* que significa “estrela” (por ex. *sideral*). Na língua dos áugures, *considerare* significava “observar os astros” e posteriormente “cogitar, pensar atentamente”, donde a palavra *considerar* em português. Já *desiderare* significava “deixar de ver”, donde “sentir falta, sentir saudades”. Dessa última acepção é que nasce *desiderium* “desejo”. Ora, *desiderium* teria dado algo como \**deseeiro* no português e, portanto, não é a base da etimologia da palavra *desejo*. Houve um cruzamento, como dissemos, com *desidia* “ociosidade, preguiça”, gerado por analogia.

Há inúmeros exemplos de analogia no português. Certa vez, ouvi alguém comentar que ia a um *croquetel*, caso evidente de misto mental entre duas palavras: *coquetel* e *croquete*. Eu não soube avaliar se esse lapso fora momentâneo ou costumeiro naquele falante. A mistura formal tem, porém, como base uma evidente relação metonímica: em coquetéis há, comumente, croquetes.

Mas nem todo cruzamento fica circunscrito num discurso: alguns se popularizam e se tornam, aos poucos, palavras oficiais. A palavra *veruculum* em latim, sofreu síncope, \**veruchu*, o que teria gerado a palavra \**verolho* em português, mas como esse objeto era normalmente feito de *ferro*, a palavra que se originou foi *ferrolho*. Também *cheminée* em francês se transformou em *chaminé* no português por analogia com a palavra *chama*. A expressão *silvam forestem* “bosque do lado de fora (da muralha medieval)”, abreviada simplesmente para *forestem*, gerou a palavra *forêt* no francês, mas se transformou em *floresta* no português, por causa de uma etimologia com *flor*. Também a palavra

*delphinum*, que deveria ter gerado \**delfinho*, mudou para *golfinho*, por associação com a palavra *golfo*, onde eles costumavam aparecer. Nos empréstimos, é comum atuar com a analogia para se resolverem casos de irregularidade sonora. Há casos muito curiosos, como o do inglês *craw-fish* “tipo de caranguejo”, antigo *cray-fish*, que vem na verdade do francês *crevice*, com analogia da forma *fish*. Quem disse que, para o povo, caranguejo não é peixe?

Um caso curioso do português é a palavra *rimiriangombe*, vinda do quimbundo, ao pé da letra “língua de vaca”, para designar o nome de uma determinada planta. Como a sonoridade é estranha, essa forma foi se tornando aos poucos *mariangombe* e, posteriormente, *maria-gomes* (há muitas plantas com o nome próprio Maria: *maria-preta*, *maria-sem-vergonha*, *maria-pereira* etc.).

Interessante também foi o que aconteceu com o jogo de crianças chamado *marelle* do francês. Muito possivelmente essa palavra tem alguma relação semântica com uma língua pré-romana desconhecida, em que \**marr-* significava “pedra”. De qualquer forma, essa palavra veio para o Brasil e passou a ser chamado *maré* ou *marela*. Por causa de alguma associação sonora, passou a *amarela* e, posteriormente, a *amarelinha*. Hoje é comum encontrar uma amarelinha pintada (de amarelo!) em muitas escolas, no entanto, a sua etimologia não tem nada a ver com a cor amarelo (em francês “amarelo” se diz *jaune*).

Outro caso curioso é a palavra *contradança*, que parece dispor do prefixo *contra-*, mas, na verdade, provém do francês *contredanse*, por sua vez, do inglês *country-dance*, “dança da roça”.

Dessa forma, as etimologias populares não só adaptam o sentido das palavras, mas interferem até mesmo no comportamento das pessoas. No âmbito religioso, em que abundam santos como *São Tomás*, *Santiago* entre outros, a palavra *sacristão* se torna, em Portugal, *são-cristão* e *Satanás* se torna *São-Tanás*!

São conhecidas ainda as brincadeiras que se popularizam e que transformaram expressões como *mal e parcamente* em *mal e porcamente*, ou *esculpido e encarnado* que se tornou *cuspidado e escarrado*. Na Literatura, Guimarães Rosa no conto “Nós os temulentos” diz: “desistindo do elevador, *embriagatinhava* escada acima”, fundindo dois verbos: *embriagar* e *engatinhar*.

Quer por lapso momentâneo ou popularizado, quer intencionalmente, a atuação da analogia tira totalmente a mecanicidade e a previsibilidade dos metaplasmos.

## **8. De onde vieram as línguas?**

No século XIX, a organização dos metaplasmos permitiu que se explicassem várias semelhanças existentes entre muitas línguas européias e asiáticas. Semelhanças entre o latim, o grego e o sânscrito já tinham sido apontadas desde o séc. XVI. Mas foi por meio de pesquisadores como Gyármathi, Rask e Bopp que se estabeleceram regras para a reconstrução lingüística de formas não documentadas. Esforços para se reconstruir uma língua-mãe, que desse conta de todas as línguas do globo, foram todos em vão (apesar de estudiosos como Trombetti dedicarem muito da sua vida nesse intento).

Na falta dessa “língua de Adão e Eva”, as propostas mais razoáveis e científicas conseguiram estabelecer várias famílias lingüísticas, a partir da noção de afiliação genética das línguas. Assim, o português é uma língua que provém do latim, que, por sua vez, vem de um grupo hipotético chamado itálico (juntamente com línguas como o osco e o umbro). O itálico, por outro lado, provém de um grupo mais antigo, o indo-europeu.

Demorou muito, porém, para se entender que reconstruções lingüísticas não têm nada a ver com reconstruções culturais ou reconstruções étnicas. Já por volta de 3000 a.C., quando o indo-europeu era falado (em dialetos), a mistura racial e os contatos culturais eram

intensos. Como hoje, o hábito de vestir-se com calças não é privilégio de povo algum e se fala português por descendentes de italianos, japoneses e libaneses, aqui e na África, não se deve imaginar que nessa época fosse muito diferente, salvas as devidas proporções (menor população mundial e maior isolamento).

Uma pretensa pureza racial e lingüística é uma visão tardia, nacionalista, pós-bonapartista, pós-Revolução Industrial, enfim, um modo romântico de entender os fatos. Estudar indo-europeu não tem nada a ver com a teoria que surgiu posteriormente que os considerava guerreiros loiros que invadiam áreas por terem dominado as técnicas de montaria. Chamados de “arianos” por leituras errôneas que mesclavam língua, raça e cultura, a teoria indo-européia conduziu o homem a uma visão racista, bem viva no início do séc. XX e em parte responsável pelas catástrofes da Segunda Guerra. Isso, porém, não invalida a teoria indo-européia, que é bastante sólida, quando nos atemos *apenas* ao aspecto lingüístico.

Os pesquisadores encontraram semelhanças impressionantes e metaplasmos muito precisos na formação dos diversos troncos do indo-europeu, a saber: o grego, o armênio, o albanês, o eslavo (que gerou o russo, o ucraniano, o bielo-russo, o polonês, o tcheco, o eslovaco, o servo-croata, o esloveno, o búlgaro), o báltico (que gerou o extinto prussiano, o letão e o lituano), o germânico (que incluem os extintos gótico e o anglo-saxão - donde sai o inglês-, além do alemão, do holandês, do sueco, do dinamarquês, do norueguês, do islandês), o indo-iraniano (donde provém o persa, o extinto sânscrito e muitas das línguas da Índia), o extinto tocário, o extinto hitita, o celta (cujas línguas vivas atuais são o gaélico irlandês, o gaélico escocês, o galês e o bretão) e o ramo itálico, donde sai o latim (e, conseqüentemente, as línguas românicas: o português, o galego, o castelhano, o asturiano, o aragonês, o catalão, o francês, o occitano, o italiano, o grupo reto-românico, o extinto dálmata e o romeno). Essas línguas são tão importantes quanto os dialetos que não têm o prestígio social de “língua de cultura”. O alemão e o italiano oficiais, por exemplo, são

construtos um tanto quanto artificiais e os dialetos do alemão e do italiano são tão antigos quanto muitas dessas línguas e, dispõem, muitas vezes, de estruturas bem diferentes.

Estudar indo-europeu é, portanto uma tarefa muito complexa.

Falamos de um fenômeno que ocorre na passagem do latim arcaico para o clássico: o rotacismo do *s*. O som *s* transforma-se em *-r-* na posição intervocálica. Esse *-s-*, porém, era uma herança do indo-europeu *\*s*. Alguns documentos marcam IOVVESAT, que equivale a IURAT “jura” (em latim arcaico, outra mudança é bem significativa: uma série de ditongos se simplificaram: *ou > u*, *oi > u*, *ei > i* ). Essa mudança, portanto, é uma inovação\* do latim.

Podemos afirmar com segurança que toda problemática da Lingüística Histórica se resume no binômio conservação: inovação\*. E da mesma forma que não é possível inovar tudo, também não é possível que uma língua seja tão conservadora a ponto de nunca ter mudado, como muitas vezes pensam alguns a respeito das línguas indígenas, do basco, do islandês ou do lituano.

Nessa mesma posição intervocálica, o grego transformava o *s* em *h*, mas o sânscrito o conservava. Não se deve pensar por isso que o sânscrito seja mais conservador do que o grego e o latim ou que ele seja mais puro do que o grego ou o latim. Esse raciocínio deve ser abolido nos primeiros passos de quem queira entender o indo-europeu. O sânscrito conserva umas tantas coisas que são inovadas pelo grego e pelo latim, mas às vezes é o contrário que ocorre. O sânscrito, por exemplo, transforma as vogais *\*e* e *\*o* do indo-europeu em *a*, ao passo que o grego e o latim os mantêm. Em germânico, um *\*a* equivale a um *\*a* ou a um *\*o* do indo-europeu. O *\*i* e o *\*u* indo-europeus se mantêm, exceto no germânico que tende a transformá-los em *\*e* ou em *\*o*.

Tenhamos em mão as formas da primeira pessoa do singular do presente do indicativo do verbo “ser” nessas três línguas: em latim, *eram*, em grego, *ên*, em sânscrito *asam*. Aparentemente não há nada em comum, mas se sabemos que em grego há uma forma mais antiga, *hean*, e que o metaplasmo *ea* > *ê* é freqüente no dialeto ático, pode-se chegar com facilidade à forma indo-européia *\*esam*.

indo-europeu <i>*esam</i> > latim <i>eram</i> (transformação <i>s</i> > <i>r</i> ) indo-europeu <i>*esam</i> > grego <i>*ehan</i> > <i>ên</i> (síncope) indo-europeu <i>*esam</i> > sânscrito <i>asam</i> ( <i>e</i> > <i>a</i> ).
--

Descobertas como essa empolgavam os estudiosos do séc. XIX a ponto de afirmarem que o único estudo lingüístico científico era o histórico. Surgem leis para o estabelecimento dos metaplasmos a partir do indo-europeu, como a famosa lei de Grimm.

A descoberta das regularidades fonéticas entusiasmou sobremaneira os pesquisadores do séc. XIX, que se dedicaram quase exclusivamente a isso. Mas um bom etimólogo deve saber onde parar. De fato, se possível recuperar parte do indo-europeu pela técnica das reconstruções, muita coisa fica sem explicação. Além disso, nem todas as línguas são indo-européias. O húngaro, o finlandês e o estoniano, apesar de européias, não pertencem à família indo-européia, mas sim à família fino-ugriana. O basco, falado na Espanha, não tem afiliação. O árabe, o hebraico e línguas extintas como o fenício e o caldeu pertencem à família das línguas semíticas. Na Índia falam-se muitas línguas indo-européias atualmente: o hindi, o gujarate, o bengali são algumas delas. Mas na mesma Índia, há milhões de falantes de línguas não-indo-européias: o tâmil e o malaiala, por exemplo, são da família dravídica. No Cáucaso, há centenas de línguas, algumas, como o georgiano, pertencem à família kartveliana. Há tentativas recentes de reconstrução de uma língua da qual teria vindo o indo-europeu, o semítico, o kartveliano, o dravídico etc. que formaria a

superfamília nostrática\*. Extremamente polêmica, a reconstrução do nostrático, apesar de bem feita, não nos dá a mesma segurança que o indo-europeu. De qualquer forma, qualquer estudioso sério está de acordo que a variação lingüística no mundo é tão grande que não é possível reconstruir uma língua-mãe. Fora da Europa encontram-se muitas famílias: o que chamamos de línguas africanas\* são centenas de línguas agrupadas em mais de uma dezena de famílias. O mesmo se pode dizer das línguas ameríndias e asiáticas. Quando recuamos no passado, é bem certo que já havia grande variedade de línguas, uma vez que se pode chegar cerca de 80 famílias lingüísticas que agrupam as mais de 5000 línguas do Globo.

### **9. Quais as línguas faladas antes do latim em Portugal?**

Na Península Itálica, ao lado do latim, havia línguas indo-européias muito aparentadas, das quais sobreviveram algumas inscrições em osco e em umbro. No sul, havia gregos (*Nápolis*, por exemplo vem de *Neá pólis* “cidade nova”). Ao norte de Roma, os etruscos dominavam com uma língua até hoje sem decifração satisfatória. Além disso, havia os vênéticos, os réticos e os lígures. No norte havia ainda a presença de povos celtas, que falavam línguas indo-européias sem muito registro escrito. Por onde quer que o Império Romano ia-se expandindo, encontrava povos com línguas distintas: nos Bálcãs havia ilírios, trácios e dácios. Na Península Ibérica, havia gregos, fenícios cartagineses, celtas e iberos. Há quem aposte numa antiqüíssima presença ambroilíria. A menção dos autores antigos a povos nem sempre coincide com as línguas que eles falavam, mas apenas na região de Portugal são citados turdetanos, túrdulos, igeditanos, presuros, gróvios, brácaros, zelas. Os lusitanos formavam um povo provavelmente indo-europeu que não se aparentava com os celtas, que teriam entrado na Península por volta do séc. VIII a.C. Mesmo os celtas eram referidos por vários nomes de tribos ao norte do Douro e ao sul do Alentejo: ártabros, cônios, seursos.



Como temos apenas os nomes desses povos e pouca coisa se manteve escrita, em alfabetos ibéricos de difícil decifração, sobram apenas palavras cuja atribuição a esta ou àquela língua é impossível. A língua basca, única sobrevivente, é tida como uma língua ibérica. A maior parte desse elemento chamado substrato\* sobrevive nos nomes das cidades e na recorrência de determinados elementos de composição. O sufixo *-asco*, tido por ambroilírio, aparece em numerosos topônimos: *Velasco*, *Vaasco*, *Panasco*, *Rabasco*, *Fontascos*, bem como uma raiz \*ambr em *Hambrón*, *Ambroa*, *Ambrães*, *Ambrões*. A terminação celta *-briga* para “cidade” aparece em *Ara.briga*, *Tala.briga*, *Conim.briga* (de onde vem o nome *Coimbra*) e por aí vai.

Ainda se podem citar muitas palavras características do português ou das línguas da Península Ibérica que remetem a essas extintas línguas do substrato. Outras são formações do substrato francês que entraram no léxico.

Além desses radicais, altamente controversos, outros costumam ser acrescentados. Por exemplo, para a palavra *caspa* há soluções pré-romanas e arábicas. Para a raiz de *entupir* há soluções pré-romanas e onomatopaicas (interjeição *tup!*). Para *baía* se atribui etimologia pré-romana ou germânica.

Para *garra* mistura-se celta com árabe. A multiplicidade de explicações apontam para um aumento da incerteza, uma vez que pouco sobreviveu do substrato e deve ter havido línguas que desapareceram sem o menor traço. Outras palavras atribuídas ao substrato são documentadas muito tardiamente, de modo que podem não ter a antigüidade que aparentemente lhe atribuem. O latim tinha contato com celtas em várias áreas do Império Romano e divulgavam suas palavras por meio de formas já alatinadas, como *beccus* > *bico*, *carrus* > *carro*, *braca* > *braga*, *gurdus* > *gordo*, *lancea* > *lança*. Outras, mais tardias também têm certa divulgação, como *capanna* > *cabana*, *camisia* > *camisa*, *cerevisia* > *cerveja*, *bardus* > *bardo* (via francês).

Palavras das mais diversas origens entraram no latim no período do Império Romano e se espalharam por muitas línguas inclusive o português: talvez seja fenícia a palavra *mappa* > *mapa*.

De origem grega há muitas palavras eruditas, mas no litoral mediterrâneo da Península Ibérica, a língua grega gerou palavras pela via popular, como: *kára* > *cara*, bem como *býrsa* > *bursa* > *bolsa*, *kybernân* > *gubernare* > *governar*, *platýs* > *\*plattus* > *chato*, *khorde* > *chorda* > *corda*, *kaûma* > *cauma* > *calma*, *spátthe* > *spatha* > *espada*. Palavras de origem etrusca já estavam muito entranhadas no latim, antes da sua divulgação pelo mundo. Da raiz céltica de *carpentum* “tipo de carro” nasce a palavra *carpinteiro*. Outras palavras de línguas célticas modernas provêm do francês ou do inglês, como *dólmen* (bretão *taol men* “mesa de pedra”, sendo que *taol* < latim *tabula*), *menir* (bretão *men hir* “pedra longa”), *drúida* (irlandês *druí* “feiticeiro”, via inglês). A palavra *duna* vem do francês, que por sua vez, vem do holandês, derivada de um latim *dunum* que remonta ao celta. Muito há para ser feito nessa área ainda, sobretudo no tocante à cronologia dessas palavras.

## **10. Outras línguas foram faladas em Portugal depois do latim?**

A partir do séc. V aumenta drasticamente a quantidade de palavras germânicas nas línguas românicas que formam o chamado superstrato\* germânico. Muitas palavras, como *sapo* > *sabão*, já existiam antes da chamada Queda do Império Romano, uma vez que havia contatos com as tribos germânicas, que integravam parte dos exércitos de Roma. Também palavras latinas entraram em período muito antigo nos dialetos germânicos (*caupo* “vendedor ambulante” > *kaufen* “vender”; *cellarium* “celeiro” > *Keller* “porão”). O reconhecimento das palavras germânicas no português é relativamente fácil, uma vez que muitas línguas germânicas existem ainda hoje, no entanto, a determinação de qual língua

provém, sobretudo nas palavras antigas, é extremamente complicada. Sabe-se que há três grandes grupos de línguas:

- a) o grupo oriental, ao qual pertencia o gótico, que deixou poucos mas importantes documentos (o maior deles é a tradução da Bíblia, atribuída a Vúlfilas, no séc. IV);
- b) o grupo ocidental, donde provêm os atuais dialetos do alto-alemão e do baixo-alemão, sem falar do franco e do anglo-saxão;
- c) o grupo setentrional, que deu origem ao Nórdico Antigo.

As línguas oficiais faladas hoje em dia formam uma certa continuidade de dialetos ou de complexos dialetais específicos. Como visto, o germânico é um dos troncos da família indo-européia. Na Península Ibérica, faz-se menção aos suevos, aos vândalos da tribo dos silingos (ano de 409 d.C.) e, posteriormente, aos visigodos (418). Os alanos, povo incluído entre os “bárbaros”, que teriam entrado com os suevos e vândalos, não falavam uma língua germânica, mas indo-iraniana, da qual não se estudaram seus traços no português. Por causa da influência do francês, muitas palavras de origem franca entraram indiretamente no português.

Línguas germânicas modernas também contribuíram para o vocabulário português: inicialmente palavras holandesas nos séc. XVI-XVII, depois palavras alemãs, no séc. XIX e palavras inglesas sobretudo no século XX. Uma palavra como *ombudsman*, de origem sueca (via inglês), originalmente “representante do povo”, é um raro exemplo de palavra de uma língua nórdica moderna.

Com relação às palavras germânicas antigas, também é possível haver mais incerteza em algumas palavras que em outras. Etimologias para a palavra *talco* se dividem entre a explicação germânica e a árabe. A palavra *brasa* tem explicações no germânico e no pré-romano. Outras, como *trepas*, entre a explicação onomatopaica e a germânica.

Muitas outras palavras mais recentes têm origem germânicas. O mineral *cobalto*, por exemplo, deriva seu nome do alemão *Kobold*, nome de um duende que os mineiros acreditavam que substituía a prata por um minério sem valor. Outro tipo de duende era o *Nickel*, donde sai também a palavra *níquel* e provavelmente também *Quarz*, donde sai *quartz*. O holandês, concorrente das navegações dos portugueses sobretudo nos séc. XVI e XVII, exportou algumas palavras do âmbito da navegação para o português, como a palavra *iate* < *jacht*, *bombordo* < *bakboord* “do lado de trás” (via francês *bâbord*), *estibordo* < *stierboord* “do lado do timão”.

Desnecessária é a exemplificação de palavras inglesas que se internacionalizam (lembrem-se apenas *golf*, *football*, *tennis*, *volleyball*, *basketball*, *spray*, *aspirin*, *jeans*, *AIDS*, *game*, *internet*, *site*, *mouse*, *driver*, *CD*, *rock'n'roll*, *design*), mas há raros exemplos de palavras que entraram no português de um contato direto com os falantes. No Brasil, grande parte da tecnologia ferroviária importada no começo do séc. XX era inglesa, momento em que alguns engenheiros vieram para cá, também algumas palavras aparecem desse contato direto. É o caso de *chulipa* “peça de madeira sobre as quais se colocam os trilhos”, que vem de *sleeper*, decalcado como “dormente”. De um contato direto também veio, provavelmente, *panqueca* < *pancake* “bolo de frigideira”.

No séc. VIII entra na Península Ibérica um novo elemento lingüístico: o árabe. Língua da família semítica, assim como o hebraico, o assírio, o aramaico, permanece em contato direto até séc. XIII em Portugal e até séc. XV na Espanha. O grande número de palavras proveniente desse contato pode ter influência direta ou vir de outras línguas, como o francês ou o italiano. Muitas dessas possuem um artigo *al-* prefixado, que muitas vezes é assimilado com as chamadas ‘consoantes solares’ em *as-*, *at-*, *ar-*, *az-*, *aç-*.

## 11. Outras palavras influenciaram o vocabulário do português?

A partir do final séc. XV, um novo grupo de palavras entrará nas línguas européias. Trata-se termos provenientes de centenas de línguas do Novo Mundo. Línguas americanas, africanas e asiáticas contribuem com um grande número de palavras. Inicialmente as línguas da América Central e do México, via castelhano, depois as da América do Sul (do taino, língua do Haiti: *cacique*, *savana*, *tabaco*, *batata*, *furacão*, *maíz* < *mahís* (donde *mais.ena*); das Antilhas: *canibal* < derivado de *caribe*; do asteca/ náuatl, língua do México: *chocolate*, *cacau* < *kakáwatl*, *sapoti* < *zapotl*, *coiote* < *koyotl*, *abacate* < *awakátl*, *tomate* < *tómatl*, *xícara* < *xikálli*; do aruak: *canoa*, *iguana* < *iwana*, talvez *goiaba*; do quíchua: *pampa*, *puma*, *chácara* < *chakra*, *mate* < *mati*, *condor* < *kuntur*, *lhama* < *llama*, *alpaca* < *pako*, *vicunha* < *huik'uña*). Da América do Norte também concorrem palavras de línguas variadas, via inglês ou francês (de alguma língua da família algonquina: *mocassin* < *mokkasín*).

No caso do Brasil, das mais de cem línguas ainda vivas, nenhuma outra língua influenciou tanto o português quanto o tupi. O tupi no entanto, faz parte da família tupi, que costuma se dividir em tronco tupi-guarani (que comporta várias línguas, desde os *mbiá* no Rio Grande do Sul até os *wayãpi* do Amapá) e os grupos tupi-não-guarani, a maioria do Mato Grosso e região Norte: *karitiana*, *juruna*, *mondé*, *mundurucu*, *ramarama*, *tupari*, *aweté*, *puruborá* e *sateré-mawé*. Além da família tupi há outras, cuja influência no léxico português ainda não foi suficientemente estudada, são elas: *karib*, *aruak*, *tukano*, *maku*, *yanomami*, *guaikuru*, *txapakura*, *pano*, *mura*, *katukina* e *macro-jê*, famílias que muitas vezes comportam línguas bem diferentes. Há ainda as chamadas línguas isoladas, sem afiliação: *aikaná*, *arikapu*, *awaké*, *irantxe*, *jabuti*, *kanoê*, *koaiá*, *maku*, *trumai* e *tukuna* (Rodrigues 1986).

O tupi, por meio do português e do francês, entrou em muitas línguas, como se vê no inglês: *toucan* “tucano”, *jaguar*, *tapioca*, *tapir*, *ipecacuanha*, *jacaranda*, *piranha*. Na

transcrição abaixo, *y* representa um som fechado, que não é nem *i*, nem *u*, antes a posição da língua é a mesma do *a*, mas a abertura da boca é a mesma do *i*. Baseou-se sobretudo em Cunha (1978) nas palavras citadas abaixo.

Apesar de muito mais recentes que os termos do substrato, nem sempre as etimologias tupis são pacíficas. Deriva-se *pipoca* do tupi, mas pode ser vocábulo expressivo (cf. inglês *popcorn*, do verbo *to pop*, de origem onomatopaica). A palavra *biboca*, por exemplo, é interpretada como tupi *ymbymbóka* “fenda no chão”, já outros vêm aí uma palavra africana (quicongo *bibóka* “lugar, casa suja”). Da mesma forma, a etimologia mais conhecida para *coroca* é a tupi *kurúka*, “resmungão”, no entanto há quem afirme vir do quimbudo *nkuluka*.

Tão importantes quanto as palavras tupi são as de origem africana que entraram para o léxico do português, principalmente na variante brasileira. Na África, em muitos locais, convivem ao lado do português e das línguas africanas nativas, vários tipos de falares crioulos, verdadeiras línguas derivadas do português ou com grande quantidade de léxico português (muitas vezes equivalente ao português do séc. XVI). Ainda há muito que pesquisar sobre o étimo de muitas palavras, uma vez que se desconhece até mesmo de quais línguas africanas provieram. Dois grupos de línguas muito distintos são citados nas bibliografias. A um deles pertencem línguas da África Ocidental, entre as quais pertencem línguas muito distintas como o *haussá* (da família lingüística afro-asiática, ramo tchádico), o *iorubá* e o *fon* (ambos da família congo-cordofaniana, tronco kwa), línguas tonais de característica isolante. Ao segundo grupo pertence a denominação genérica de línguas *bantu* (derivado de *bantu*, plural de *mntu* “gente”), da África Centro-Meridional, na verdade, uma família de línguas altamente aglutinantes e prefixais, às quais pertence, por exemplo, o *quicongo* e o *quimbundo*. Convém que se distingam os dois grupos nitidamente. Muitas etimologias nessa área são polêmicas e não é incomum pensar que se tratam de formações expressivas, sem uma origem definida. As seguintes listas de palavras foram retiradas, em grande parte, dos levantamentos de Castro (2001).

Palavras do grupo kwa são características pela presença de redobros silábicos. São faladas nos atuais países de Gana, Togo, Benim e Nigéria. As línguas bantu (ao todo mais de 500) têm palavras mais longas e normalmente com encontros *ng*, *nd*, *mb* que as caracterizam. Faladas em quase toda África Centro-Meridional, no caso das palavras portuguesas são de especial importância os grupos do Gabão, da República do Congo (capital Brazzaville), da República Democrática do Congo (ex-Zaire, capital Kinshasa), de Angola e de Moçambique. Dentre as línguas bantu, o quimbundo (falada sobretudo em Angola) é a mais representativa nas etimologias do português, embora, com certeza, os dialetos que teriam originado de fato sejam de difícil determinação.

O efeito da analogia e a imensa variação dialetal dessas línguas poderiam justificar por que a palavra original não apresenta exatamente as mesmas feições da derivada. Em nenhuma área, porém, há maior incerteza do que nas palavras de origem africana. Palavras como *cachaça* são tidas como ou derivadas do vernáculo (de um termo *cacho*, com explicações sinuosas) ou do quicongo *kisasa*. A palavra *banana* oscila entre uma etimologia africana, uma ameríndia e uma árabe. O mesmo ocorre com *balangandã*, ora associado a um quimbundo *mbalanganga*, ora ao verbo *balangar* (forma alternativa de *balançar*). Às vezes há etimologias em iorubá e em quicongo, como ocorre com a palavra *gogó* (respectivamente *gògòngò* e *ngongoló*), sem que a distância entre os falantes na África e a total incompreensibilidade entre seus falantes intimidem quem formule tais soluções, o que, *mutatis mutandis*, equivaleria a não saber se uma palavra é espanhola ou finlandesa. Na verdade faltam pessoas que conheçam bem as línguas em questão e, ao mesmo tempo, os métodos que devem ser seguidos para o estabelecimento de uma boa etimologia. Faltam, ainda, estudos mais profundos para entender as circunstâncias de sua formação e do contato entre falantes. Há palavras africanas que se tornaram internacionais, como *chimpanzé*, vinda do francês *chimpanzé*, vinda, por sua vez, de alguma língua banta (quicongo *kimpeensi*).

É comum dizer que o português tem palavras de origens muito variadas. No entanto, grande parcela dessas palavras é, na verdade, vinda pelo inglês ou pelo francês. O contato direto do português com a Finlândia foi praticamente nulo, mas isso não impediu que uma palavra dessa língua não-indo-européia, a saber, *sauna*, entrasse no português. Do dinamarquês vem *edredom* < *ederdun* “acolchoado com penas de êider (tipo de pato)”. Há portanto muitas palavras internacionais. Também as palavras do leste europeu não têm origem direta: *coche* < castelhano *coche* < francês *coche* < alemão *Kutsche* < tcheco ou eslovaco *koč* < húngaro *kocsi*. A partir do húngaro *paprika* chega-se ao português *páprica*, provavelmente via italiano. Do irlandês, língua céltica moderna, vem a expressão *sluagh gairm* “grito de guerra”, que o inglês transformou em *slogan*, forma com a qual chegou no português. Outra palavra irlandesa é *uisce* “água”, que passou para o inglês *whisky*, donde o português *uísque*, da mesma forma que o russo *vodka* “agüinha” não chegou ao português diretamente. Também do russo (língua indo-européia do tronco eslavo) vêm *samovar* “que se auto-aquece”, *perestróica* “reestruturação”, *glasnost* “transparência”, *sputnik* “companheiro de viagem”.

Assim, da Oceania vêm *canguru* < *kangaroo* (de uma língua indeterminada, com o significado de “não entendi você”), *bumerangue* < *boomerang*. Palavras de línguas do Oceano Pacífico também vêm ao português pelo inglês (como o tonga *tabu*, o havaiano *luau* ou o taitiano *tatau*, que gerou o termo *tatuagem*, via inglês *to tattoo*).

Palavras de origem persa (língua do tronco indo-iraniano, da família indo-européia) sempre acabaram se internacionalizando e chegavam até Portugal em vários momentos, por exemplo, passavam ao turco (língua da família altaica, completamente diferente do árabe e do persa), que, por sua vez, iam aos comerciantes de Veneza e de lá para o francês. Isso ocorreu com o persa *dulbānd* > turco *tülbent* > italiano *turbante* e *tulipa*; persa > turco *havyar* > italiano *caviale* > francês *caviar*, assim também indiretas são as palavras vindas do persa *bazar* < persa *bâzâr*, *divã* < *dîuân* “conselho”, *xale* < *shâl*, *gaze* < *gazî*, *paraíso* <



persa antigo *paridaeza* “recinto circular” (via grego *paradeisos*, latim *paradisus*), *quiosque* < *kûshk* “palácio”, *tafetá* < *tâftah*, *carmesim* < *kirm* (pelo turco *kırmızı*) “vermelho”.

A maioria das palavras hebraica (língua da família semítica, como o árabe) que conhecemos chegou ao português via latim ou grego: *sábado* < hebraico *shabbât* ∞ (via grego *sábbaton*, latim *sabbatum*), *páscoa* < hebraico *pâssah* (pelo grego *Paskha*, latim *Pascha*), *fariseu* < perusim “os separados” (pelo grego *Pharisaîos*, latim *pharisaeus*), *maná* < *mân* (aramaico *mannâ*’, pelo latim *manna* ou grego *mánna*), *aleluia* < *hallelûyâh* “louvai ao Senhor” (pelo latim *halleluia*), *amém* < *amén* “assim seja” (pelo latim *amen*), *messias* < *mâshîah* “ungido” (pelo latim *Messias*, decalcado em grego como *Khristós* - derivado do verbo *khriein* “passar óleo, untar, ungir”, latim *Christus*), *Satã* < *sât*∞*ân* “inimigo, adversário” (pelo latim *satan*, *satanas*, como no grego).

Como vimos, palavras persas chegaram ao português e às línguas européias via árabe. Também muitas palavras árabes chegaram indiretamente. A palavra *café* é uma palavra árabe que passou para os turcos e ao português, via italiano, pois eram os italianos que faziam comércio com os turcos no Mediterrâneo, no Império Otomano e não os portugueses. A palavra *sofá* é outra palavra árabe que chegou ao turco e daí ao francês, de onde se deriva a palavra portuguesa. Imaginar que palavras como *café* e *sofá* são empréstimos do período em que os árabes estavam na Península Ibérica é um grande erro, porém bastante comum.

No entanto os portugueses a partir do séc. XVI tiveram contatos diretos com vários povos da Ásia. Esses povos, como se pode esperar, pertencem a muitas famílias distintas.

As inúmeras línguas da Índia são divididas em duas grandes famílias conforme dito atrás: a indo-européia e a dravídica. As línguas indianas com as quais os portugueses tinham mais

contato eram o concani (falada em Goa e arredores), o tâmil (do Ceilão, atual Sri Lanka, a antiga Taprobana de Camões) e o malaiala (na região de Malabar). A primeira é indo-européias, já as duas últimas, dravídicas. O chinês e seus inúmeros dialetos pertencem a uma família distinta, o sino-tibetano. O malaio (língua muito distinta do malaiala acima citado) faz parte da grande família malaio-polinésio. O japonês é uma língua isolada, que já foi associada à família ural-altaica, junto com o turco e o coreano, mas hoje em dia isso não é mais considerado verdadeiro. Os inúmeros dialetos ciganos provêm de línguas indianas também. Subestima-se muito a sua influência e são muito pouco estudadas e, provavelmente, explicariam muitas palavras que se atribuem aos substratos. Normalmente cita-se *calão* < *caló*, nome de uma dessas línguas. A palavra portuguesa *gajo*, redução de *gajão*, provêm de *gachó* “homem adulto, estrangeiro”.

Por outro lado, muitas das palavras que revelavam um contato direto com os asiáticos por meio dos portugueses se perderam completamente: palavras como *bogari*, espécie de arbusto, do concani *mogri* ou ainda a palavra *nele*, tipo de arroz, do malaiala *nel* são hoje totalmente desconhecidas.

Há, ainda, palavras que vieram da Ásia e passaram ao português via castelhano (sobretudo Filipinas), via holandês ou francês (da antiga Indochina) ou inglês (sobretudo Índia). Uma palavra como *xampu*, tem origem no hindustani, língua indo-européia indiana, via inglês (o mesmo se pode dizer de *chutney* < hindi *chatni*); outro exemplo é *pijama*, via francês. No séc. XIX os estudos do sânscrito revelaram muitos aspectos de práticas indianas que chegaram ao português pelo francês (*carma* < *karman*, *avatar* < *avatâra*, *ioga* < *yôga*, *brâmane* < *brâhmana*), mas, na verdade, se tornaram palavras internacionais. Outras palavras em sânscrito tiveram sua divulgação via árabe: é o caso de *açúcar* e *cânfora*. Sabendo-se *mahâ* significa “grande” e que *râjâ* significa “rei”, temos a composição do hindustani *mahârâjâ*, que, via inglês *maharaja*, chega ao português *marajá*. O mesmo *mahâ* está no nome de Mahatma Gandhi, pois *Mahatma* significa “grande alma”.

De línguas indo-européias indianas derivam-se ainda *chita* (sânscrito *citra*), *jambo* (sânscrito *jambû*) e *palanquim* (sânscrito *palyañka*). Do urdu vem a palavra *cáqui* < *kâkî*. Das línguas dravídicas aponta-se como originário do tâmil a palavra *angelim* < *añjili*, bem como *pária* < *pareiyan* “tocador de bombo”.

Influência inversa também ocorreu, uma vez que há muitos falares crioulos espalhados pela Índia, Timor Leste e Macau, quase todos em risco de extinção galopante. Na Malásia ainda hoje se fala *boneka* < português *boneca* (donde se derivou um verbo *bonekanan* “tratar como uma boneca”), bem como *sepatu* < português *sapato*. No Timor Leste o português convive com línguas malaio-polinésias, como o tétum, e, portanto, nesta região, o número de palavras de empréstimo é bem maior.

O japonês, língua isolada, manteve contatos com os portugueses durante o séc. XVI, mas com o fechamento dos portos, só a partir do séc. XX muitas palavras voltaram a entrar, sobretudo no Brasil, onde houve grande imigração. Algumas palavras japonesas vieram diretamente do português do séc. XVI: *pan* < *pão*, *koppu* < *copo*, *zubon* < *gibão* e muitos acham que *arigatō* < *obrigado*, sem muita comprovação.

O coreano, outra língua isolada, possui um número igualmente grande de termos de origem chinesa. Palavras coreanas no português são as internacionais, como *taekwondo* “método do soco” (literalmente “punho fechado, punho de ferro”).

Da família sino-tibetana provêm a língua tibetana e as variantes chinesas. Do tibetano veio para o português a palavra internacional *dalai-lama*, aparentemente com o primeiro elemento de origem mongol (língua da família altaica). Dos muitíssimos falares chineses, destacam-se o mandarim e o cantonês, que deram origem a muitas palavras que também se internacionalizaram:

## 12. Podemos saber de onde vêm todas as palavras?

Quando se conhece o momento do contato que permitiu à língua-fonte a entrada no léxico da língua estudada, não se apresentam etimologias anacrônicas ou misteriosas (aparições de palavras pertencentes a línguas que nunca estiveram realmente em contato). Entre uma língua-fonte e uma língua estudada há, às vezes, línguas-ponte: assim, palavras finlandesas vieram ao português não de forma direta, mas por meio do inglês ou do francês.

Quando se datam as primeiras ocorrências de uma palavra, pode-se finalmente teorizar acerca de qual língua teria originado a palavra em questão. É preciso lembrar que uma palavra pode aparecer escrita somente séculos depois de ser empregada na fala, pois a escrita é bastante conservadora. Isso é um pouco diferente hoje em dia, pois nos jornais há seções que utilizam uma linguagem tremendamente coloquial, refletindo, assim, muitas vezes, a língua tal qual se fala. Na *internet* também é possível, por meio de buscadores de palavras, encontrar abonações para qualquer palavra sem a necessidade de *corpora*.

Conhecendo bem os metaplasmos, não se reconstruirão formas absurdas, nem se derivarão palavras a partir de outras sem muito rigor, como nos exemplos de Fernão de Oliveira. Os casos de irregularidade devem ser minuciosamente estudados, quer por causa da origem culta de alguns vocábulos, quer por causa da ação analógica de outros. É preciso criar hipóteses sobre como essa irregularidade ocorreu e, se possível, explicar não só esta palavra mas outras de comportamento parecido.

Por fim, como ideal de qualquer etimologia, podemos afirmar que, ao conhecermos a frequência de uso das palavras nas línguas-fonte (tarefa bastante difícil, diga-se de passagem), podemos afirmar com mais segurança sobre uma atuação analógica qualquer.

Será que as explicações etimológicas que circulam por aí tiveram todos ou pelo menos alguns desses cuidados? É comum derivarem uma palavra portuguesa de outra portuguesa, como se todas tivessem sido formadas hoje. A sensação que surge é que não se necessita mais do que a intuição para ser etimólogo. Isso tudo, sem falar dos anacronismos flagrantes que surgem dessas explicações.

Concluindo, convém observar que uma boa etimologia está sempre sendo construída e não é algo pronto. Em muitos momentos é preciso humildade para admitirmos, após tentarmos todos os meios acima apresentados, que não temos a menor idéia de onde algumas palavras vêm.

### **13. Novas perguntas**

Procure em dicionários etimológicos a origem das palavras de que tenha curiosidade, consultando dicionários de Latim como o de Gaffiot (1934), e dicionários etimológicos como os de Nascentes (1952), Machado (1956/1977), Corominas (1994), Cunha (1982 e 1989).

Você pode também descobrir o sentido de seu nome, caso ele tenha aparecido no segundo volume do *Dicionário Etimológico* de Antenor Nascentes.

### **14. Bibliografia para aprofundamento.**

Para aprofundar as questões aqui levantadas, leia Viaro (2004), Beekes (1995), Benveniste (1995), Costa (2000), Faria (1958, 1970), Nunes (1945), Romanelli (1964), Said Ali (2001), Silva Neto (1979), Teyssier (1990) e Williams (1973).

## 15. Glossário

*Texto:* Como reconhecer a raiz de uma palavra? ([Link1](#))

- Noções de Particípio - Não consta
- Raiz - Não consta
- Prefixos - Não consta
- Sufixos - Não consta
- Radical - Não consta
- Afixação - Processo, nas línguas flexivas e aglutinantes, em que pequenas sílabas (afixos) se agregam a uma raiz, formando o radical. Os afixos mais importantes são os prefixos, como em *refazer*, que vêm antes da raiz, e os sufixos, que vêm depois, como em *refazendo*.
- Apofonia - Transformação fonética, muito freqüente no latim, que consiste na mudança da vogal quando ocorre a junção de um prefixo a uma raiz ou a um radical. Assim, *barba* sofre apofonia quando essa palavra é prefixada por *in-*, donde *imberbe*, “homem sem barba”.
- Assimilação - Transformação fonética que ocorre quando, por causa da proximidade de dois sons, um deles afeta o outro de tal modo que o torna idêntico ou muito parecido a si. Assim, o segmento *rs* da palavra *persona* transformou-se em *s*, grafado *ss*, em *pessoa*. Dizemos então que o *r* se assimilou (= se assemelhou) ao *s*.

*Texto:* Por que alguns sons se modificam? ([Link2](#))

- Palavras vulgares - Não consta
- Palavras eruditas Não consta
- Síncope - Transformação fonética que consiste no desaparecimento, com o passar do tempo, de um som no meio da palavra.
- Metaplasmos - Qualquer mudança fonética caracterizada por uma adição, subtração, transformação ou transposição de sons ou de acentos de uma fase lingüística mais antiga para uma mais recente.

*Texto:* Por que algumas palavras mudam de sentido? ([Link3](#))

- Semi-eruditas - Não consta
- Sonorizações - Transformação fonética que consiste na mudança de um som surdo (*p, t, k, f, ss, x*) para um som sonoro (*b, d, g, v, z, j*).
- Lenização - Transformação fonética em que as plosivas\* se tornam fricativas\*.

*Texto:* Uma palavra latina pode gerar mais de uma palavra portuguesa? ([Link4](#))

- Divergentes - Diz-se quando um único étimo gera mais de uma palavra numa língua derivada. Assim, a palavra latina *umeru* deu em português as formas divergentes ombro, e o termo técnico *úmeru*.
- Convergentes - Diz-se quando mais de um étimo (= palavra que dá origem a outra) se torna uma palavra homônima numa língua derivada. Assim, latim *sunt* mudou para *são* (em *eles são*), tornando-se homônimo de *sanu* > são (em homem *são*). Dizemos então que *são* (verbo) e *são* (adjetivo) são formas convergentes.

*Texto:* Como reconhecer uma etimologia fantasiosa? ([Link5](#))

- Étimo - Não consta
- Generalização - Não consta
- Especificação - Não consta
- Abrangência - Não consta
- Sistematicidade - Não consta
- Coerência - Entendeu-se inicialmente a coerência como uma propriedade centrada no texto, que faz com que o texto faça sentido para o receptor. Atualmente, concebe-se a coerência como construção de sentidos efetuada no processo de interação, a partir do texto. Isto porque a coerência está diretamente ligada à possibilidade de se estabelecer um sentido para o texto, em uma determinada situação comunicativa, o que depende de vários fatores, dentre eles o conhecimento de mundo e o conhecimento partilhado pelos interlocutores. Se o interlocutor não domina conhecimentos prévios para apreender o sentido do texto e se o locutor não calcula adequadamente o sentido de seu texto, omitindo informações não conhecidas pelo interlocutor, fica prejudicada a construção de sentidos para o texto, ou seja, a construção da coerência.

*Texto:* Tudo pode ser explicado pelas transformações fonéticas? ([Link7](#))

- Analogia - Não consta

*Texto:* De onde vieram as línguas? ([Link8](#))

- Inovação - Não consta
- Nostrática - Não consta
- Línguas africanas - Não consta

*Texto:* Quais as línguas faladas antes do latim em Portugal? ([Link9](#))

•Subtrato - Conjunto de línguas faladas num determinado território antes da implantação de uma língua que a elas se sobrepôs, e que passou a contribuir com materiais léxicos para a língua vencedora, que servem de base para as considerações etimológicas.