

Freqüência e uso de estrangeirismos ingleses no português brasileiro: Um estudo baseado em corpus

Tony Berber Sardinha*

Leila Barbara**

PUC SP

Neste trabalho, apresentamos uma investigação baseada em corpora acerca da extensão do uso de estrangeirismos de origem inglesa no português do Brasil. Para tanto, utilizamos o corpus Banco de Português, do projeto DIRECT (LAEL, PUCSP), com mais de 230 milhões de palavras, de fontes orais e escritas, e o British National Corpus, com cerca de 100 milhões de palavras, também provenientes da escrita e fala. A investigação motiva-se (1) pela necessidade do envolvimento dos lingüistas e lingüistas aplicados em debates acerca de políticas públicas, (2) por ser esse um tema caro ao prof. John Robert Schmitz, e (3) pela escassez de trabalhos sobre estrangeirismos que utilizem corpora eletrônicos. Os resultados indicam que (1) há muitos anglicismos em vigor no português brasileiro, (2) a freqüência deles é baixa e (3) eles parecem adquirir sentidos e usos diferentes da língua de origem.

In this paper, we report a corpus-based investigation on the spread of use of English words in Brazilian Portuguese. For this purpose, we employed the Bank of Portuguese, a corpus created by the DIRECT Project (LAEL, PUC/SP), with over 230 million words, from spoken and written sources, and the British National Corpus, with 100 million words, from writing and speech as well. Our research is motivated by the following points: (1) the need for the engagement of linguists and applied linguists in debates over public policies, (2) the fact that this is a theme pursued by John Robert Schmitz, and (3) the lack of studies on foreign word borrowings based on electronic corpora. Results suggest that (1) a large number of English words are in use in Brazilian Portuguese, (2) their frequency is low, and (3) these words take on new different meanings from English.

* O autor agradece ao CNPq pelo auxílio 350455/2003-1.

** A autora agradece ao CNPq pelo auxílio 500945/2003-9.

Introdução

Recentemente, muitos lingüistas e lingüistas aplicados sentiram-se impelidos a tomar parte em um debate que abarcou a sociedade: a questão dos estrangeirismos. Esses são entendidos como “palavras e expressões de outras línguas, usadas correntemente em algumas áreas do nosso cotidiano” (FARACO, 2001, p. 9). No âmbito desse debate, assistimos a vários profissionais e cientistas de linguagem conclamando seus pares a se pronunciarem diante do projeto de lei nº 1676/99 do deputado Aldo Rebelo, que pretende “proteger” a língua portuguesa de uma suposta “ameaça” de “invasão” de palavras de origem estrangeira (mais especificamente, oriundas do inglês).

Neste trabalho, deixaremos de lado a definição própria de “palavra portuguesa”, tendo em vista as dificuldades em se delimitar claramente o que vem a ser uma palavra genuinamente portuguesa. Note que até mesmo a instância indicada pelo projeto de lei como sendo a autoridade máxima da língua, a Academia Brasileira de Letras, inclui estrangeirismos no Vocabulário Ortográfico da língua portuguesa. Um desses é “site”, que lá consta como “site (saite) s.m. ing.”. Vale notar que “site” é visto assim por alguns: “Da avalanche de expressões inglesas que deformam a língua portuguesa, uma, em particular, é humilhante: site.” (<http://www.igutenberg.org/jj29idioma.html>). Por isso, partiremos para uma abordagem baseada em corpus, isto é, em instâncias de uso da língua (vide abaixo). A própria Academia admite que

O certo é que a língua portuguesa cresceu, até mesmo em virtude da introjeção de termos ligados ao desenvolvimento científico e tecnológico ou de muitos estrangeirismos. É o caso de palavras como teleducação (educação à distância), acessar (entrar), deletar (apagar, anular), decassegui (trabalhador brasileiro no Japão), teleconferência (conferência à distância), lincar (ligar), internet, infovia, intranet, etc. (<http://www.academia.org.br/ortogra.htm>).

No nosso caso, quando nos referirmos a estrangeirismos, estaremos fazendo alusão a palavras existentes na língua inglesa. Tampouco nos aventuraremos no campo teórico, tentando definir o que é ser uma palavra inglesa. Basta lembrar que o inglês contemporâneo possui uma presença massiva de palavras de origem latina para dar-nos conta das dificuldades inerentes de uma tentativa nesse sentido.

A posição do lingüista aplicado na sociedade é a de estabelecer um debate corrente, com muitas e sérias implicações. Tomar posição nele é, portanto, uma atitude política (RAJAGOPALAN, 2003). A sociedade conclama os “cientistas da linguagem” a se pronunciarem a respeito desse tema, que pode afetar a vida das pessoas de várias maneiras, além de sinalizar um entendimento legal a respeito de como deveria ser a língua dos brasileiros.

Um desses cientistas de linguagem que jamais se furtou a participar dos debates públicos envolvendo questões lingüísticas nacionais é John Robert Schmitz. No referido debate em torno dos estrangeirismos, lá vemos John falando e escrevendo publicamente a partir do ponto de vista da lingüística aplicada e da lexicografia (SCHMITZ, 1999, 2000a, b, 2001a, b, 2002). Em outro, envolvendo ataques à Lingüística Aplicada por parte de um grande nome de um ramo da lingüística formal, também lá estava o Prof. John (SCHMITZ, 2004). Aliás, cabe lembrar que muito antes do referido projeto de lei, John já escrevia sobre o assunto (SCHMITZ, 1988). Por conta desse envolvimento, sofreu várias críticas (JORGE, 2002), acendendo outro debate (CUNHA, 2002).

Neste trabalho, desejamos fazer uma contribuição ao debate sobre os estrangeirismos na língua portuguesa a partir da Lingüística de Corpus, que é um ramo da lingüística voltado ao estudo da linguagem por meio de corpora eletrônicos, isto é, coletâneas de amostras de língua falada e escrita armazenadas e manuseadas por computador (BERBER SARDINHA, 2004; BIBER, CONRAD & REPPEN, 1998). Mais especificamente, trazemos evidências a respeito da quantidade de palavras inglesas na língua portuguesa e de seus usos em comparação com o emprego das mesmas palavras na língua inglesa. Investigações sobre a presença e freqüência de estrangeirismos no português brasileiro ainda são escassas (ALVES & MARONEZE, 2003; BERBER SARDINHA, 2004, p. 177-184). Entretanto, elas podem identificar com mais clareza e exatidão a verdadeira extensão da chamada “ameaça” dos estrangeirismos.

Um dos lingüistas aplicados que nos insta a buscar respostas em corpora é o próprio John Schmitz. Ele lança questões que somente podem ser respondidas por meio de freqüências extraídas de corpus eletrônico. Por exemplo, em relação à palavra “printar”, ele indaga: “será que ocorre livremente em textos?” (SCHMITZ, 2001b, p. 92). Ele resume, no mesmo trabalho, uma posição chave que pretendemos avançar neste trabalho: “Seria de grande utilidade, sem dúvida, identificar os estrangeirismos e sua freqüência” (ibid, p. 99). Outros lingüistas também colocam

questões ou posições acerca da freqüência de estrangeirismos que poderiam ser iluminadas com o uso de corpora. Por exemplo, Bagno (2001, p. 81) pergunta se “haveríamos de temer que um punhado de palavras de origem inglesa poderia ser “lesivo” ou “descaracterizar” a língua portuguesa falada no Brasil”. Com contagens baseadas em corpora extensos, poderíamos saber várias coisas acerca desse punhado, como seu tamanho, composição, e até mesmo se é um punhado ou uma “avalancha”.

Além disso, com apoio em corpora, poderemos evitar recorrer à intuição sobre como seria a língua portuguesa para vê-la como seus falantes realmente a empregam, pois “uma abordagem intuitiva dos fatos lingüísticos [é] quase sempre baseada em mitos, superstições e aparências enganadoras” (BAGNO, 2001, p. 73). Precisamos salientar a importância de que o corpus seja eletrônico. Conforme já enfatizado pelos lingüistas de corpus (MCENERY & WILSON, 1996; SINCLAIR, 1991), um corpus eletrônico torna a pesquisa não somente mais confiável, como também mais apta a revelar aspectos que desconhecíamos a respeito da língua. Assim, por ser eletrônico, usaremos programas de computador especializados para buscar, contar e comparar ocorrências de palavras. Esses procedimentos se distinguem do tipo de análise com corpus que preconiza uma leitura feita “com lápis na mão para ir assinalando cada ocorrência” (BAGNO, 2001, p. 77).

Desejamos rever, à luz da evidência advinda de corpora, dois pressupostos básicos dos simpatizantes do projeto de lei 1676/99, quais sejam:

- (a) Haveria uma invasão de palavras estrangeiras na língua portuguesa, particularmente de origem anglo-saxã.
- (b) As palavras estrangeiras não se adaptariam ao português.

Corpora empregados na pesquisa

Os corpora empregados na pesquisa foram os seguintes:

Variedade	Corpus	Palavras (tokens)
Inglês britânico	BNC (British National Corpus)	100.602.129
Português brasileiro	Banco de Português	231.960.832

Tabela 1: Descrição dos corpora

O British National Corpus é uma referência entre os corpora de língua inglesa. Ele é fechado, isto é, sua composição é imutável, e é formado por textos orais e escritos de procedência britânica (em sua quase totalidade). Esse corpus é comercializado por um consórcio de editoras e universidades.

O Banco de Português é um corpus aberto, cujo conteúdo é constantemente atualizado. Ele contém textos orais e escritos (em sua maioria jornalísticos), quase inteiramente de origem brasileira. Todos os textos e transcrições contidos nele são de circulação brasileira, mesmo que não tenham sido escritos ou falados por autores falantes nativos do português brasileiro. O Banco de Português é armazenado na Pontifícia Universidade Católica de São Paulo, sendo fruto do projeto DIRECT. Para mais informações sobre o BP, favor consultar o website do corpus, em <http://lael.pucsp.br/corpora/bp>, onde estão disponíveis ferramentas para exploração do corpus, além de acesso aos textos integrais.

Os instrumentos usados na análise dos corpora foram:

- WordSmith Tools (SCOTT, 1998): Listador de frequências, concordanciador.
- Programas específicos: Rotinas criadas especialmente para esta pesquisa por um dos autores, usando *Text Utilities* e outros utilitários do sistema operacional Unix.
- Calculador de Mutual Information e T-Score (BERBER SARDINHA, 2003): Programa disponível online, em <http://lael.pucsp.br/corpora>, que efetua o cálculo de estatísticas de associação (mais detalhes abaixo).

Quantidade de palavras inglesas no português brasileiro

Nesta parte do artigo, faremos uma incursão nos corpora de estudo com vistas a revisar o primeiro dos pressupostos elencados acima, qual seja, o de que há uma invasão de palavras estrangeiras (de origem inglesa) no português do Brasil.

A tabela 2 traz o resultado da comparação dos dois corpora (Banco de Português e BNC), feita item a item, para saber quantas palavras¹ (*types*) as duas línguas possuem em comum.

¹ É importante lembrar que foram comparadas as formas ortográficas das palavras e não seus sentidos. Assim, uma palavra como 'a' é comum ao português e ao inglês, mas possui significados diferentes em cada uma.

	Palavras diferentes (<i>types</i>)
Português total	618.868
Inglês total	385.750
Palavras em comum	100.172

Tabela 2: Porcentagem de palavras comuns aos dois corpora

Como se percebe, há cerca de 100 mil palavras em comum entre os dois corpora. Em termos do corpus de português, isso corresponde a 16,10%, ou aproximadamente uma palavra a cada seis.² A tabela 3 a seguir traz os números relativos a cada corpus.

Língua	Itens não compartilhados	% do total
Português	518.696	84%
Inglês	285.578	74%

Tabela 3: Número e porcentagem de itens diferentes em cada língua

Quando comparamos os dois corpora não por completo, mas em faixas de frequência, descobrimos que o grau de compartilhamento decresce rapidamente.

Palavras em comum entre as ...	Palavras (tokens)	% (do total do português)
50 mil mais freqüentes	6.998	1,13%
10 mil mais freqüentes	672	0,11%

Tabela 4: Porcentagem de palavras em comum em dois tipos de frequência

Entre as 50 mil palavras mais freqüentes (o que significa palavras com frequência em torno de 120, isto é, de 1 vez a cada 2 milhões de palavras), as palavras em comum são apenas cerca de 1% do total. Indo para frequências mais altas, na faixa das 10 mil palavras mais freqüentes, que ocorrem cerca de 1700 vezes no Banco de Português (ou 7 vezes por milhão), as palavras em comum correspondem a apenas 1 décimo de 1 por cento. Isso significa que os estrangeirismos, ao contrário do que se diz, não são freqüentes na língua portuguesa.

² Em relação ao total de itens (*types*).

Conforme mostrado acima, 74% das palavras em uso no inglês não estão presentes no português. Dito de outro modo, 26% do vocabulário que está em uso no inglês também está na língua portuguesa. Isso pode parecer bastante alarmante, mas é preciso lembrar que essa é uma via de mão dupla: estamos contando as palavras em comum, sem distinção de sua origem. Vale lembrar que boa parte do vocabulário inglês é de origem latina, assim como no português. Visto por esse lado, não é de se espantar esse grau de compartilhamento.

Determinar a etimologia das palavras compartilhadas foge ao propósito deste trabalho. No entanto, uma indicação inicial da origem dessas palavras pode ser obtida por meio da presença de letras banidas do nosso alfabeto oficial: k, w e y, além de seqüências de caracteres típicas do inglês, como “tion”, “ing” e “s”. Contando as palavras que possuem tais letras e seqüências entre as palavras em comum, obtivemos um total de 6311, isto é, cerca de 94% do total de palavras compartilhadas. Esses números sugerem que entre as palavras compartilhadas figuram, maciçamente, palavras com grafia inglesa. Em conclusão, isso indica que as palavras compartilhadas entre as duas línguas se parecem mais com o inglês do que com o português. Sendo assim, a origem imediata mais provável do vocabulário em comum é o inglês.

Entre as seis mil palavras que não possuem tais traços característicos da grafia inglesa, estão, quase sempre, nomes próprios e siglas, como por exemplo:

ABDUL	BRUM	FERRARA
ABDULAH	BUA	JARRELL
ABE	BUARQUE	JAVA
ABEF	CPD	JAZZ
ABEL	CPH	RUDER
BRUDER	CPL	RUEFF
BRUEL	CPM	ZUMA
BRUER	FERBER	ZURCHER
BRUHL	FERMER	ZVEZDA
BRULÉE	FERMÉ	

Notamos que entre essas estão palavras com grafia alemã, francesa e italiana. Dado que são nomes próprios, uma das coisas que revelam, em certa medida, é o grau de “miscigenação” dos povos, provavelmente ocasionado pela imigração, tanto no Brasil quanto nos EUA.

Uma palavra estrangeira em uso no português: *marketing*

Dentre as várias palavras comuns ao português e ao inglês que possuem grafia inglesa está *marketing*. Essa é uma palavra freqüente tanto no português quanto no inglês, conforme mostra a tabela a seguir:

Língua	Freqüência	% do total	Posição
Português	17.113	0,007	1461
Inglês	5.021	0,005	1986

Tabela 5: Freqüência da palavra *marketing* nos dois corpora

Como se percebe, *marketing* é mais comum no português do que no inglês. Por sua alta freqüência e por sua grafia, essa se torna uma palavra indicada para se verificar até que ponto uma palavra estrangeira incorpora-se ao português.

A fim de conhecermos o uso dessa palavra no corpus, aplicamos o conceito de padrão (*pattern*), conforme usado na Lingüística de Corpus. Padrão é a fraseologia mais típica de uma palavra, representada por suas colocações, coligações e prosódia semântica mais freqüentes (BERBER SARDINHA, 2004, p. 31). Para detectar a padronização das palavras, extraímos seu conjunto de colocados (palavras que aparecem próximas) e fizemos uma estimativa do grau de associação entre essa palavra e seus colocados. As concordâncias e os colocados foram obtidos com o programa WordSmith Tools (SCOTT, 1998).

O grau de associação foi medido por duas estatísticas: Mutual Information e T-Score (BERBER SARDINHA, 2004, p. 200-205). Conforme dito acima, o cálculo foi realizado com a ferramenta “Analisador de Tabela de Colocados”, disponibilizada *online* na página de ferramentas de análise de corpus do CEPRIIL/LAEL (vide anexos para reproduções parciais dos resultados fornecidos pelo programa).

Os principais colocados de *marketing* no corpus de português, ordenados pelo valor de Mutual Information, são os seguintes:

	Colocado	fc	fnc	MI	T-Score
1	QUINZENAL	305	91	10.396	9.532
2	PALESTRANTE	527	137	10.202	11.695
3	ARENA	3971	535	9.250	23.092
4	SPORTS	1158	151	9.202	12.267
5	TRAFFIC	1038	74	8.331	8.576
6	GERENTE	18743	1262	8.247	35.408
7	CICLO	9196	436	7.742	20.783
8	ESTRATÉGIAS	4182	163	7.460	12.694
9	JOGADA	5123	163	7.167	12.678
10	RELACIONADAS	2710	75	6.966	8.591
11	PROPAGANDA	15915	401	6.829	19.850
12	ISL	2132	52	6.783	7.145
13	PUBLICIDADE	12612	305	6.770	17.304
14	ESTRATÉGIA	19383	468	6.769	21.435
15	DIRETORA	11765	223	6.419	14.759
16	PALESTRA	7884	147	6.395	11.980
17	DIRETOR	103362	1508	6.042	38.244
18	CONSULTOR	7154	103	6.023	9.993
19	PELÉ	12892	148	5.695	11.931
20	VENDAS	37570	428	5.685	20.286
21	COMUNICAÇÃO	25069	273	5.619	16.187
22	ASSISTENTE	5862	61	5.554	7.644
23	CIRCULAÇÃO	8279	85	5.535	9.021
24	GETÚLIO	6757	67	5.485	8.003
25	COLA	7243	69	5.427	8.114

Tabela 6: Principais colocados de *marketing* no corpus de português, ordenados pelo valor de Informação Mútua

Os colocados indicam *marketing* sendo usado como:

- Cargos dentro de uma área de atuação profissional: gerente (~ de marketing), diretor, diretora, consultor, assistente.
- Tema de evento especializado: ciclo, quinzenal (“ciclo quinzenal Arena do Marketing”), palestrante, palestra.

Há um colocado que destoa desse conjunto: “jogada”, que forma “jogada de marketing”. A concordância abaixo ilustra essa colocação:

Estado, isso não passa de uma simples jogada de marketing. Como todos nós sabemos que o pa
cno, sem que isso sinalize apenas uma jogada de marketing suja, é sinal de que o mundo pop
da Villeneuve à F-1 foi uma grande jogada de marketing. Campeão da Indy e com um sobren
e que o petista seja envolvido numa “jogada de marketing” do governo. Debate eleito
o criada pelas TVs pagas não passa de jogada de marketing para atrair publicidade. O Brasi

Essa colocação parece carregar um prosódia semântica negativa, indicada pelo léxico:

- não passa de (uma simples) jogada de marketing
- apenas uma jogada de marketing
- envolvido numa jogada de marketing.

Outras colocações semelhantes de “marketing” são as seguintes:

- lance de marketing (34 ocorrências)
- golpe de marketing (29 ocorrências).

Os principais colocados do inglês aparecem a seguir.

	Colocado	fc	fnc	MI	T-Score
1	SALES	10596	303	7.539	17.314
2	ADVERTISING	4495	89	7.010	9.361
3	MILK	5003	83	6.754	9.026
4	DISTRIBUTION	6351	94	6.590	9.595
5	AGGRESSIVE	1959	27	6.487	5.138
6	PRODUCT	11504	121	6.096	10.839
7	JOINT	7180	75	6.086	8.533
8	PERSONNEL	3393	33	5.984	5.654
9	PROMOTION	3393	32	5.939	5.565
10	MANUFACTURING	4451	37	5.757	5.971
11	DIRECT	11957	95	5.692	9.559
12	STRATEGY	6185	49	5.688	6.864
13	DIRECTOR	12435	98	5.681	9.707
14	PRODUCTION	15943	115	5.552	10.496
15	COMPANY'S	3886	28	5.551	5.179

16	GLOBAL	3633	26	5.541	4.990
17	VICE	4333	29	5.444	5.262
18	PRODUCTS	10759	62	5.229	7.664
19	FINANCE	7560	43	5.211	6.380
20	MANAGER	13661	73	5.120	8.298
21	PRESIDENT	17122	84	4.997	8.878
22	INTERNATIONAL	22155	105	4.947	9.915
23	SELLING	5897	25	4.786	4.819
24	MANAGEMENT	21982	91	4.752	9.185
25	DEPARTMENT	17299	57	4.423	7.198

Tabela 7: Os principais colocados de marketing em inglês, ordenados pelo valor de Informação Mútua

Os padrões de *marketing* no inglês, a partir da tabela acima, podem ser resumidos nas seguintes categorias:

- Partes ou cargos de uma empresa: (*marketing* ~) *personnel, director, vice-president, manager, department*
- Entidades que promovem *marketing*: *Milk* (~ *Marketing Board*)
- Âmbito do *marketing*: *International* (~ *marketing*), *Global*
- Atividades de uma empresa: (*marketing* ~) *sales and distribution, (marketing and ~) advertising.*

Também aqui há um colocado que destoa desse conjunto: *aggressive* (~ *marketing*), ilustrado na concordância a seguir:

same time. But it suffered from aggressive marketing by established players such as
dor has been characterized by very aggressive marketing and it is more than apparent th
he majority of the population. The aggressive marketing of seats to foreign cosmonauts
about glamour. Does the somewhat aggressive marketing of the Labeque lustre — tous
a record breaking year by using an aggressive marketing campaign. Flying high: Thre

A concordância parece indicar a não existência de uma prosódia semântica negativa para *aggressive marketing*. Parece se tratar de uma atividade de *marketing* intensiva, feita com o intuito de iniciar, aumentar ou recuperar vendas.

Uma colocação equivalente também existe no português: “*marketing* agressivo”, com usos semelhantes.

Comparando os sentidos das duas palavras, podemos sugerir que *marketing* possui sentidos semelhantes nas duas línguas, com usos voltados à área empresarial. Seria esse, então, um caso de estrangeirismo “invasivo”, ou seja, que entrou na língua portuguesa e pouco se adaptou? A resposta é não. A evidência disso vem de colocações que fogem da área empresarial. Uma delas é ‘jogada de marketing’, com sua prosódia semântica negativa. Não há algo diretamente semelhante em inglês com ‘marketing’ (talvez *marketing scheme* ou algum dos sentidos de *hype* seja aplicável nesse caso).

O termo equivalente português para *marketing* é “mercadologia”, segundo o dicionário Aurélio. Essa palavra possui apenas 15 ocorrências no Banco de Português. Um terço dos usos da palavra mercadologia refere-se à esfera acadêmica, de cursos, conforme mostra a concordância abaixo:

instrução com concentração curricular em mercadologia, 40 para administração com propaganda e administração com ênfase em mercadologia Exame: 23 e 24 de janeiro (300 vagas), administração com ênfase em mercadologia (60) e administração com ha mercadorias e imagens ideológicas, entre mercadologia e produção cultural, vão se ós vemos... marketing... Mercadologia Um Mercadologia Dois Administração EconoMIA

Aproximadamente 25% dos usos indicam semelhança com *marketing* no sentido não empresarial:

disputas eleitorais em um torneio de pura mercadologia. Duda, com Nizan Guanaes, esmo, um efeito especial. Trata-se de uma mercadologia de alto nível. Nada do que uma surra de Maradona no departamento da mercadologia. Impaciente com as idas e v balho da banda, é inevitável pensar que a mercadologia da revolta está na ronda. T

Assim, parece que “mercadologia” não foi abraçada pela língua em geral e tem preferência por contextos acadêmicos e como substituto de “propaganda”.

Em resumo, esta análise de *marketing* indica que mesmo uma palavra como ela, que mantém uma semelhança grande em relação ao seu uso na língua de partida, ainda assim adquire sentidos novos na nova língua. Isso sugere que o processo de incorporação de palavras estrangeiras estampa nessas palavras sentidos que são peculiares aos falantes da língua de chegada.

Conclusão

Nossas questões de partida diziam respeito a duas alegações dos defensores de legislação anti-estrangeirismos, que eram a de que haveria uma invasão de palavras estrangeiras na língua portuguesa e a de que as palavras estrangeiras não se adaptariam ao português.

Em relação ao primeiro aspecto, descobrimos que as palavras em uso no inglês correspondem a cerca de 16% do vocabulário em uso do português brasileiro. Além disso, a pesquisa revelou que as palavras advindas do inglês são de baixa freqüência. Cremos que esses achados não autorizam a aplicação do rótulo de invasão. Alguns poderiam argumentar que tal taxa de uso do vocabulário inglês é abusiva. Contudo, usando de elementos da retórica daqueles que poderiam assim pensar, cremos que dada a globalização, a influência da mídia, o poderio econômico dos EUA, e a própria “submissão” das classes dominantes do Brasil, seria de se supor que qualquer palavra do inglês entraria no português. Entretanto, mesmo depois de anos desse “ataque” interminável, o vocabulário do português do Brasil permanece independente do inglês, tanto em unidades (*types*) quanto em ocorrências (*tokens*). É neste último ponto que aparece o maior grau de independência do português, pois as palavras não inglesas são, de longe, muito mais freqüentes do que as inglesas no português.

Em relação ao segundo aspecto, observamos que palavras inglesas, mesmo quando de alta freqüência e de grafia marcadamente inglesa, adquirem usos e sentidos peculiares ao português. Ou seja, mesmo aquelas palavras que poderiam ser mais facilmente “denunciadas” como corrompendo o idioma nacional sofrem aportuguesamento.

Ambos os achados salientam a necessidade de trazer para o debate a noção de freqüência. Sem ela, a discussão se dá em termos absolutos (“a palavra *x* do inglês está/não está no português”). Assim como em outros assuntos relativos à língua, o componente freqüencial, probabilístico é de extrema importância para entender como a língua funciona (BOD, HAY & JANNEDY, 2003; HALLIDAY, 1991).

Os achados desta pesquisa corroboram outros estudos baseados em corpora sobre a mesma temática. Por exemplo, Berber Sardinha (BERBER SARDINHA, 2004, p. 177-184) notou que a freqüência de palavras com grafia estrangeira no português brasileiro é baixa (0,9% do total de ocorrências) e que palavras colocadas “na lista negra”, como

delivery, *light*, *playoffs*, entre outras, possuem uma freqüência total baixa na língua. Em relação a *marketing*, nosso estudo vem ao encontro de Serralvo (2004), que enfocou a tradução de textos acadêmicos da própria área de *marketing*. Segundo esse estudo, *marketing* possui usos específicos no português que não são traduzíveis diretamente, mesmo em textos especializados de dentro da própria área.

Quando deixamos de lado os aspectos quantitativos da presença de unidades ortográficas nas duas línguas e nos detivemos em aspectos mais qualitativos de uso das palavras, notamos que mesmo uma palavra como *marketing*, com características ortográficas inglesas (“k” e “ing”) e de alta freqüência, adquire sentidos freqüentes peculiares ao português, ou seja, que inexistem ou que são pouco comuns no inglês. Sendo assim, parece haver uma clara adaptação à nova língua, às preferências de seus falantes e de seu discurso. A adaptação de palavras inglesas à língua portuguesa já havia sido apontada por outros autores. Por exemplo, Garcez e Zilles (2001, p. 32) notam que *outdoor*, palavra de origem inglesa, passou a significar no Brasil o que nos Estados Unidos se chama de *billboard*.

Acreditamos ter trazido subsídios empíricos para auxiliar a discussão sobre a defesa da língua portuguesa frente ao suposto ataque do vocabulário inglês.

Referências

ALVES, I. M.; MARONEZE, B. O. Neologismos por empréstimo na imprensa brasileira contemporânea. Trabalho apresentado no 13º InPLA. LAEL, PUCSP, São Paulo, 30 de abril a 3 de maio de 2003.

BAGNO, M. Cassandra, Fênix e outros mitos. In: FARACO, C. A. (Org.). *Estrangeirismos – Guerras em torno da língua*. São Paulo: Parábola Editorial, 2001. p. 49-84.

BERBER SARDINHA, A. P. Calculator for WordSmith Tools Collocate Tables. Online software. <http://lael.pucsp.br/corpora>, 2003.

BERBER SARDINHA, A. P. *Lingüística de Corpus*. São Paulo: Editoria Manole, 2004.

BIBER, D.; CONRAD, S.; REPPEN, R. (Org.). *Corpus Linguistics – Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.

BOD, R.; HAY, J.; JANNEDY, S. (Org.). *Probabilistic Linguistics*. Cambridge, Mass: MIT Press, 2003.

CUNHA, R. B. O chute e o escorregão. Observatório da Imprensa, 27/3/2002. Disponível na Internet em <http://observatorio.ultimosegundo.ig.com.br/artigos/fd270320024.htm>.

FARACO, C. A. Apresentação. In: FARACO, C. A. (Org.). *Estrangeirismos - Guerras em torno da língua*. São Paulo: Parábola Editorial, 2001. p. 9-14.

GARCEZ, P. M.; ZILLES, A. M. S. Estrangeirismos - Desejos e Ameaças. In: FARACO, C. A. (Org.). *Estrangeirismos - Guerras em torno da língua*. São Paulo: Parábola Editorial, 2001. p. 15-36.

HALLIDAY, M. A. K. Corpus studies and probabilistic grammar. In: AIJMER, K.; ALTENBERG, B. (Org.). *English corpus linguistics: Studies in honour of Jan Svartvik* London: Longman, 1991. p. 30-43.

JORGE, F. Os deslizos e os ilogismos do Prof. John. *Imprensa*, v. 16, p. 59-61, 2002.

MCENERY, T.; WILSON, A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.

RAJAGOPALAN, K. *Por uma Lingüística crítica: linguagem, identidade e a questão ética*. São Paulo: Parábola, 2003.

SCHMITZ, J. R. A Língua Portuguesa e os estrangeirismos. *Diário Oficial de Leitura*, v. 8, p. 4-5, 1988.

SCHMITZ, J. R. A língua portuguesa e o polêmico projeto de lei n. 1676 de 1999. *Todas as Letras*, v. 2, p. 64-70, 2000a.

SCHMITZ, J. R. Palavras estrangeiras e a língua portuguesa: invasão cultural ou desenvolvimento técnico-científico. *Revista Caliban*, v. 3, p. 42-46, 2000b.

SCHMITZ, J. R. Língua Pausterizada. *Folha de S.Paulo*, 6/1/1999, p. 3. Tendências/Debates.

SCHMITZ, J. R. O projeto de Lei n.1676/99 na imprensa de São Paulo. In: FARACO, C. A. (Org.). *Estrangeirismos - Guerras em torno da língua*. São Paulo: Parábola Editorial, 2001b. p. 85-106.

SCHMITZ, J. R. Para que servem os dicionários? *Imprensa*, v. 15, p. 66-67, 2002.

SCHMITZ, J. R. Taking linguistics seriously: on the varied dimensions of applied linguistics. *Lingua*, v. 114, p. 95-99, 2004.

SCOTT, M. *WordSmith Tools Version 3*. Oxford: Oxford University Press, 1998.

SERRALVO, G. *Versão de textos acadêmicos de marketing: uma abordagem baseada em corpus*. 2004. MA Thesis, LAEL, PUCSP, São Paulo.

SINCLAIR, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

Anexos: Amostras dos resultados fornecidos pelo calculador *online* de estatísticas de associação

Anexo 1: Palavra “marketing” no corpus Banco de Português

results_55148862_marketing_bnc.txt

file:///Volumes/CASA%3ETONY/MyDocuments/Articles/In%20preparation/schmitz/r Google

WebMail OS X Amazon Amazon.co.uk Google Images Google Groups News Apple Mac eBay Yahoo!

*** Analisador de tabela de colocados do WordSmith Tools v.3 ***
 PUC/SP, LAEL, CEPRI
 Thu Apr 8 11:10:20 EST 2004
 Informações retiradas da lista de palavras:
 Enviada em Thu Apr 8 11:47:20 EST 2004 por 385750
 Corpus: 020202 -- Nome: Tony Berber Sardinha
 N: 100602128 {tokens}
 {Types: 385750}
 Nódulo: marketing {fn: 5150}
 Horizonte: 3
 Ordenação dos resultados: Mutual Information, decedente
 Legenda:
 fn: freqüência do nódulo
 fc: freqüência do colocado {se igual a 0, significa que não estava presente no corpus de estudo}
 fnc: freqüência do nódulo e colocado juntos
 OE: Razão Observado / Esperado [fórmula: {fnc/N}/({fn/N}*{fc/N}*horizonte)]
 MI: Informação Mútua [fórmula: log2{OE}] {se não tiver sido calculado, significa que fc era igual a 0}
 T: T-Score [fórmula: {fnc-({fn*fc)/N}*horizonte)/raiz_quadrada{fnc}] {se não tiver sido calculada, significa que fc era igual a 0}
 Valores mínimos recomendados:
 MI > +3
 T > +2
 Valores abaixo desses {inclusive negativos}: palavras não têm atração entre si

#	Colocado	fo	fnc	OE	MI	T
0001	MARKETING	5275	5249	6479,860158966716	12,661	72,439
0002	AMP	698	31	289,337089201877	8,175	5,549
0003	SALES	10596	303	186,227910715389	7,539	17,314
0004	ADVERTISING	4495	89	128,942282466112	7,010	9,361
0005	MILK	5003	83	108,059201047806	6,754	9,026
0006	DISTRIBUTION	6351	94	96,396678015062	6,590	9,595
0007	AGGRESSIVE	1959	27	89,820281124497	6,487	5,138
0008	PRODUCT	11504	121	68,498035195626	6,096	10,839
0009	JOINT	7180	75	68,027283511269	6,086	8,533
0010	PERSONNEL	3393	33	63,349555813055	5,984	5,654
0011	PROMOTION	3393	32	61,429895712630	5,939	5,565
0012	MANUFACTURING	4451	37	54,149734982332	5,757	5,971
0013	DIRECT	11957	95	51,737563006793	5,692	9,559
0014	STRATEGY	6185	49	51,590615400910	5,688	6,864
0015	DIRECTOR	12435	98	51,321532058374	5,681	9,707
0016	PRODUCTION	15943	115	46,9722222222222	5,552	10,496

Anexo 2: Palavra “marketing” no British National Corpus

results_35258955_marketing_bp.txt

file:///Volumes/CASA%3BTONY/MyDocuments/Articles/In%20preparation/schmitz/r... Google

Webmail OS X Amazon Amazon.co.uk Google Images Google Groups News+ Apple Mac eBay Yahoo!

*** Analizador de tabela de colocados do WordSmith Tools v.3 ***
 PUC/SF, LAEL, CEPRIIL
 Thu Apr 8 10:38:02 EST 2004
 Informações retiradas da lista de palavras:
 Enviada em Thu Apr 8 11:47:20 EST 2004 por 618868
 Corpus: 010101 -- Nome: Tony Berber Sardinha
 N: 231960832 (tokens)
 (Types: 618868)
 Núdulo: marketing {fn: 17113}
 Horizonte: 3
 Ordenação dos resultados: Mutual Information, decendente
 Legenda:
 fn: freqüência do núdulo
 fc: freqüência do colocado (se igual a 0, significa que não estava presente no corpus de estudo)
 fnc: freqüência do núdulo e colocado juntos
 OE: Razão Observado / Esperado {fórmula: {fnc/N}/({fn/N}*{fc/N}*horizonte)}
 MI: Informação Mútua {fórmula: log2{0E}} (se não tiver sido calculada, significa que fo era igual a 0)
 T: T-Score {fórmula: {fnc-({fn*fc}/N)*horizonte}/raiz_quadrada{fnc}} (se não tiver sido calculada, significa que fo era igual a 0)
 Valores mínimos recomendados:
 MI > +3
 T > +2
 Valores abaixo desses (inclusive negativos): palavras não têm atração entre si

#	Colocado	fc	fnc	OE	MI	T
0001	MARKETING	17113	16189	4274.896177875780	12,060	127,206
0002	QUINZENAL	305	91	1348.134020618556	10,396	9,532
0003	PALESTRANTE	527	137	1178.874251497005	10,202	11,695
0004	ARENA	3971	535	609.197834125726	9,202	23,092
0005	SPORTS	1158	151	589.647644927536	9,202	12,267
0006	TRAFFIC	1038	74	322.241414141414	8,331	8,576
0007	GERENTE	18743	1262	304.231560700106	8,247	35,408
0008	CICLO	9196	436	214.275763793889	7,742	20,783
0009	ESTRATÉGIAS	4182	163	176.116290726817	7,460	12,694
0010	JOGADA	5123	163	143.790464497646	7,167	12,678
0011	RELACIONADAS	2710	75	125.176151761517	6,966	8,591
0012	PROPAGANDA	15915	401	113.860238424553	6,829	19,850
0013	ISL	2132	52	110.213864306784	6,783	7,145
0014	PUBLICIDADE	12612	305	109.272583728081	6,770	17,304
0015	ESTRATÉGIA	19383	468	109.105667315595	6,769	21,435
0016	DIRETORA	11765	223	85.660607680655	6,419	14,759
0017	MARKETING	17113	16189	4274.896177875780	12,060	127,206