

# Linguística de *Corpus* e ensino: a compilação de um *corpus* de especialidade para preparação e implementação de um curso preparatório rápido para exame de proficiência

Corpus Linguistics and teaching: the compilation of a specialized corpus for the preparation and implementation of a crash prep course for a proficiency examination

Stella E. O. Tagnin\*  
Danilo S. Murakami\*\*

---

**RESUMO:** Este artigo apresentará o processo de compilação de um *corpus* de especialidade na área de Relações Exteriores e seu uso para definir o conteúdo programático e a preparação de material didático para candidatos a um exame de proficiência em inglês para preenchimento de um cargo público no âmbito do governo federal.

**ABSTRACT:** This article will present the compilation of a specialized corpus in Foreign Affairs and how it was used to define the syllabus and support the preparation of teaching materials for candidates that would take a proficiency exam as one of the requirements to fill a federal government position.

**PALAVRAS-CHAVE:** Linguística de *Corpus*. Ensino de inglês. Conteúdo programático. Material didático.

**KEYWORDS:** Corpus Linguistics. English teaching. Syllabus. Teaching materials.

---

## 1. Introdução

Neste artigo apresentaremos o processo de compilação de um *corpus* de especialidade para informar um curso preparatório rápido para candidatos a um cargo público federal na área de Relações Exteriores. A investigação do *corpus* norteou o conteúdo programático assim como a preparação do material para o curso.

O curso destinava-se a candidatos que deveriam se submeter a um exame de proficiência em língua inglesa como um dos requisitos para o preenchimento das vagas. A duração do curso era de 12 horas, divididas em quatro dias, ou seja, aulas de três horas por dia. Os candidatos já haviam passado por uma prova que envolvia uma redação em inglês. Essas redações foram corrigidas manualmente e avaliadas sob três quesitos: conteúdo, estrutura e expressão. Na parte do ‘conteúdo’, que valia 20 pontos, era julgada a “perspectiva adotada no tratamento do tema; [a] capacidade de análise e senso crítico em relação ao tema proposto [e a] consistência dos argumentos, clareza e coerência no seu encadeamento”. No

---

\* Livre docente, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

\*\* Mestrando, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

questo ‘estrutura’, que valia 30 pontos, foram avaliados “o respeito ao gênero solicitado; [a] progressão textual e encadeamento de ideias; [e a] articulação de frases e parágrafos (coesão textual)”. Finalmente, no quesito ‘expressão’, responsável pelos 50 pontos restantes da prova, foram levados em conta o domínio, a proficiência e o emprego correto das estruturas próprias da língua inglesa escrita. Como a parte relativa à ‘expressão’ era responsável por metade do valor da prova, a entidade organizadora da prova de proficiência optou por privilegiar esse aspecto num curso rápido preparatório a ser oferecido aos candidatos qualificados, isto é, aos que haviam atingido ou ultrapassado a nota mínima exigida na redação em língua inglesa. Colocava-se o problema de ‘o quê’ abordar nesse curso, uma vez que o cargo a que os candidatos almejavam previa a redação de textos em língua inglesa. O primeiro passo foi analisar os erros mais recorrentes detectados durante a correção da redação. A fim de abordar esses problemas num contexto real de uso, compilou-se um *corpus* com textos relacionados à área de Relações Exteriores. Na seção seguinte (seção 2) discutiremos os objetivos que pretendíamos alcançar com o curso. Em seguida (seção 3), apresentaremos os princípios da abordagem denominada *Data-Driven Learning* (DDL). O *corpus* compilado e as ferramentas utilizadas para definir o conteúdo a ser abordado no curso e para preparar os respectivos exercícios são discutidos na seção 4. Na seção 5 analisaremos os resultados alcançados para concluir, na seção 6, com algumas considerações sobre as possibilidades de apropriação da metodologia por parte dos candidatos.

## 2. Um curso de revisão para exame de proficiência em língua inglesa – o que abordar?

Dado o curto tempo disponível para a pretendida revisão gramatical, optamos por, inicialmente, conscientizar os candidatos sobre os aspectos convencionais da língua (TAGNIN, 2013), ou seja, embora gramaticalmente seja por vezes possível expressar-se de certa forma, o convencional, ou seja, o usual, não é empregar essa forma. Por exemplo, apesar de ‘gatos e cães’ ser uma forma gramaticalmente correta, não é a forma consagrada de nos referirmos a ‘cães e gatos’. Da mesma forma, ‘severamente ferido’, apesar de gramatical, não é a forma usual de denominarmos alguém que foi ‘gravemente ferido’. Em outras palavras, a língua é um sistema probabilístico (HALLIDAY, 1961), em que, embora certas formas sejam *possíveis*, elas não são *prováveis* de ocorrer numa produção fluente da língua. Nesse sentido, dominar as estruturas convencionais contribui de forma significativa para se alcançar a fluência nessa língua. A Linguística de *Corpus* tem provado ser extremamente eficaz em

identificar essas formas convencionais, uma vez que são padrões que se repetem. Familiarizar os candidatos com os princípios dessa abordagem foi outro objetivo, pois que lhes daria autonomia para fazerem suas próprias investigações. Observe-se, no entanto, que eram candidatos com várias formações (Administração de Empresas, Direito, Economia etc.), mas praticamente sem qualquer formação linguística. Por essa razão, pretendíamos apresentar-lhes *corpora on-line* a que pudessem recorrer sempre que tivessem alguma dúvida de redação. Para tanto, foram preparados vários exercícios baseados em *corpora*, numa abordagem denominada *data-driven learning* (DDL) (JOHNS, 1991), que preconiza o aprendizado pela observação de várias instâncias de um fenômeno linguístico.

### 3. Como “ensinar” o conteúdo programático? A abordagem *data-driven learning* (DDL)

Diferentemente de métodos de ensino mais consagrados em que o professor transmite ao aluno o que sabe acerca da língua, na perspectiva DDL, o aprendiz participa ativamente do processo de construção de conhecimento. Conforme aponta Johns (2002), o instrutor fornece ao aprendiz ocorrências autênticas de uso da língua na forma de linhas de concordância. O aprendiz, por sua vez, examina os dados reais da língua estudada a fim de encontrar padrões linguísticos recorrentes. Tal processo leva o aprendiz a levantar hipóteses sobre a organização linguística das informações que observa nesses dados. A partir de suas observações, o aprendiz é capaz de testar as hipóteses que levantou com o objetivo de fazer generalizações a respeito do funcionamento da língua estudada. Nessa abordagem, portanto, o aprendiz atua como um pesquisador linguístico (JOHNS, 2002, p. 108), uma vez que tem de aprender por meio da análise de dados de que dispõe.

Tendo em vista que cabe ao aluno observar-levantar hipóteses – fazer generalizações para aprender uma língua, pode-se ter a errônea impressão de que o instrutor que adota essa abordagem, não teria o trabalho de “ensinar”. No entanto, Johns (1988, p. 10) chama a atenção para o fato de que o papel do professor está no trabalho prévio de preparação e apresentação das informações que serão examinadas pelo aprendiz. Em vez de preparar um texto, que poderia ser concebido artificialmente, ou seja, escrito pelo próprio professor para a lição, o instrutor deve preparar o material de modo que contemple especificamente o tópico a ser abordado e seja adequado ao nível de conhecimento de seus aprendizes.

Entretanto, a concepção dos estudantes de que, para compreender qualquer elemento da língua, eles devem entender todos os elementos que o cercam, pode, no início, configurar

uma desvantagem da DDL (JOHNS, 1988, p. 10). Quando consideramos a exposição do aprendiz a linhas de concordância, pode-se levantar pelo menos dois pontos que podem causar-lhe certa estranheza. Primeiramente, uma concordância tem o formato peculiar em que o tópico em estudo é apresentado centralizado de maneira que há a mesma quantidade de caracteres em ambos os lados, esquerdo e direito. Tal formatação limita o número de caracteres dispostos a partir do centro, independentemente do limiar entre sentenças ou palavras. Dessa forma, uma linha de concordância pode ou não coincidir com o início de uma sentença ou com uma sentença completa, da mesma forma que a palavra no início e no fim da linha pode ser lida em sua plenitude (*palavra*) ou vermos somente algumas letras que a compõem (*[pala]vra*). Em segundo lugar, dado que as linhas de concordância provêm de usos reais da língua, o aprendiz pode encontrar palavras que desconhece, as quais podem desviar sua atenção do foco de aprendizado (Quadro 1)

Quadro 1: linhas de concordância para *Brazil*.

N	Concordance
1	nded by the Trade Ministers of Brazil, Egypt, Argentina, Ind
2	oviding inputs to negotiators. Brazil 's WTO Ambassador, Clo
3	o shift the blame to India and Brazil for not offering enoug
4	tral to Doha Round," India and Brazil state positions Inglês
5	tral to Doha Round," India and Brazil state positions Hong K

Entre as vantagens de adotar a DDL no ensino, podemos citar a independência da metalinguagem gramatical. Muitas vezes o estudante não domina, especialmente em língua estrangeira, os termos usados pelo professor em sua explicação, porém pode alcançar sucesso nas tarefas práticas de análise da língua que não necessariamente dependem da nomeação de fenômenos da língua. Além disso, o aprendiz pode descobrir algo que o professor não pensou em incluir em sua exposição tradicional, ou algum fato que até mesmo dicionários ou gramáticas não demonstram com clareza.

#### 4. Por trás dos bastidores

Como dissemos, o primeiro passo foi compilar um *corpus* com textos na língua inglesa relacionados às Relações Exteriores. Esses foram coletados da internet a partir de jornais, revistas ou *sites* especializados em relações exteriores, segundo os critérios descritos em Berber Sardinha (2004).

Para a coleta dos textos que compuseram o *corpus*, utilizamos a ferramenta BootCat (BARONI e BERNARDINI, 2004). Apesar de ser comum a prática de criar um *corpus* a partir da coleta manual de textos disponíveis na Internet, ou seja, acessando a página em que cada texto é encontrado para obter seu conteúdo, a ferramenta adotada realiza o processo de forma automática. A BootCat funciona a partir de palavras iniciais que atuam como “sementes” (do inglês, *seeds*). A ideia é que tais palavras, selecionadas no início do processo, identifiquem o domínio que se deseja investigar, no nosso caso, Relações Exteriores. Uma vez que temos esse número inicial, inserimos essas sementes na ferramenta para buscar páginas da web que as contenham. Em nosso trabalho, usamos como sementes unidades lexicais como *Brazil*, *USA*, *foreign policy*, *global warming*, *climate change*, *Doha Round*, entre outras (Figura 2).

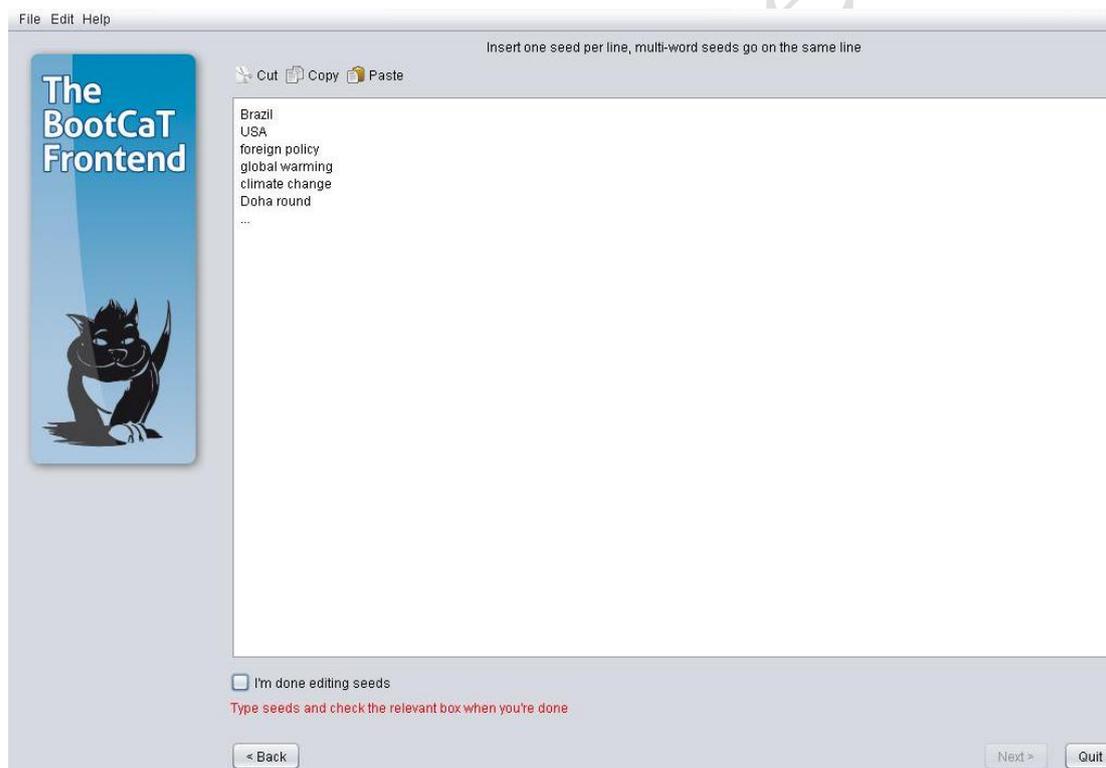


Figura 1: sementes (*seeds*) inseridas na BootCat.

A ferramenta permite que tais palavras sejam organizadas em conjuntos que combinam aleatoriamente uma quantidade específica de “sementes”, conjuntos que o programa chama de *tuples*. Nessa etapa, a BootCat procura páginas que contenham necessariamente todas as palavras que fazem parte de um conjunto. Em outras palavras, na criação de nosso *corpus*, selecionamos dez conjuntos que continham combinações de três

palavras (e.g. *Brazil, USA, foreign policy; global warming, USA, Doha round*; etc). É também possível excluir os domínios de rede que não queremos pesquisar, para evitar que certos conteúdos provenientes de páginas que julgamos inapropriadas apareçam entre os resultados da busca. Uma vez que pretendíamos averiguar com mais cuidado o conteúdo das páginas encontradas, não recorremos ao recurso de limitar domínios. O resultado que a ferramenta nos fornece é uma lista de endereços URL que contém os conjuntos de *tuples* determinados anteriormente. É importante ressaltar que um dos procedimentos que tornam a ferramenta eficiente é a possibilidade de gerar novas sementes a partir das páginas encontradas nos resultados. Consequentemente, pode-se reiniciar o processo a fim de coletar mais dados. No entanto, uma vez que havíamos adotado um número que julgávamos suficiente de sementes iniciais, consideramos que não foi necessário repetir o procedimento.

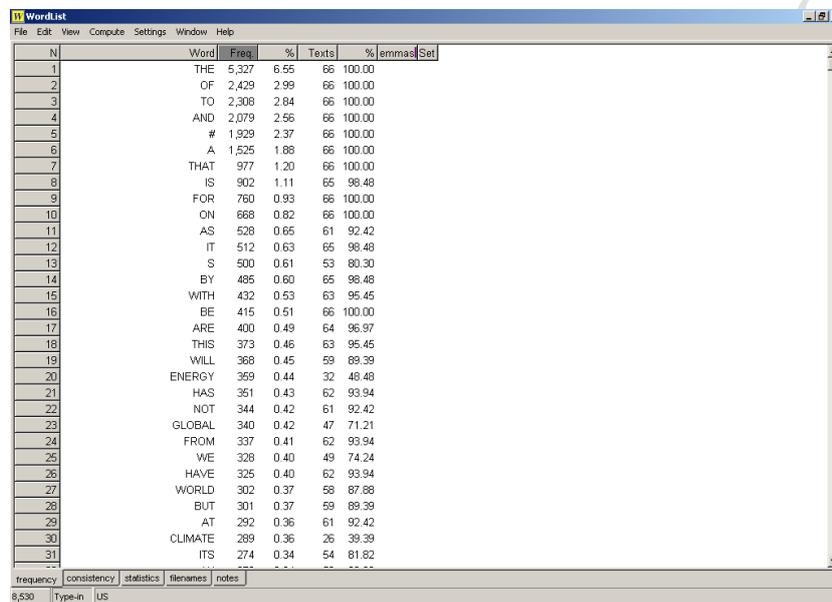
Como mencionamos anteriormente, uma forma de coletar textos que comporiam o *corpus* seria vasculhar páginas da Internet e baixar o seu conteúdo. Optamos por adotar a ferramenta BootCat devido a algumas de suas vantagens. Primeiramente, não conhecíamos em profundidade a área de Relações Exteriores. Não poderíamos, portanto, acessar diretamente páginas que contemplassem tal domínio de conhecimento, uma vez que não estávamos familiarizados com seus textos. Nesse aspecto, valemo-nos de palavras ou termos encontrados nas redações dos alunos para usar como sementes na busca de textos. Outra vantagem está relacionada ao propósito da coleta do *corpus*. Como se tratava de um *corpus* “descartável” (VARANTOLA, 2002), ou seja, construído para uma tarefa específica, não foi necessário seguir os critérios rígidos que se estabelecem ao construir um *corpus* de especialidade para fins lexicográficos ou terminográficos. Em outras palavras, foi uma tarefa bem menos trabalhosa. Por fim, em consequência do limite de tempo, a ferramenta foi de grande valia, pois o prazo para apresentação do material para o curso não permitia uma coleta demorada. Com a automatização que a ferramenta proporciona, obtivemos de maneira rápida e objetiva o *corpus* que serviu como recurso para a etapa seguinte, a criação do material didático.

À época de nosso trabalho utilizamos a interface da BootCat disponível no *site* da ferramenta. Hoje ela está apenas disponível para *download*<sup>1</sup>. Desenvolvimentos mais recentes estão relatados em Bernardini & Ferraresi (2013).

---

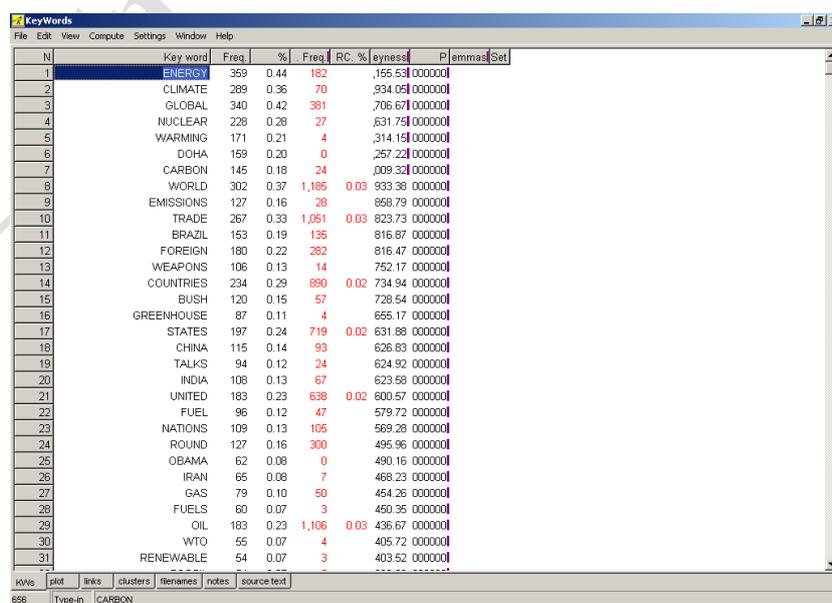
<sup>1</sup> BootCat: <http://bootcat.sslmit.unibo.it>

O produto resultante da utilização da BootCat, o *corpus* de Relações Exteriores é composto por sessenta e nove artigos integrais escritos em inglês, provenientes de jornais, revistas e *sites* especializados, num total de 58.104 palavras. Os textos compreendem o período entre 1999 e 2009. O passo seguinte foi determinar o vocabulário mais recorrente nesse *corpus*. Para isso foi feita uma Lista de Palavras (*WordList*) (Figura 2) com o software WordSmith Tools, versão 5 (SCOTT, 2007) que foi, em seguida, comparada a uma lista de palavras de um *corpus* jornalístico de 2 milhões de palavras a fim de evidenciar as palavras mais típicas (palavras-chave ou *KeyWords*) do *corpus* de Relações Exteriores (Figura 3).



N	Word	Freq	%	Texts	%lemmas	Set
1	THE	5,327	6.55	66	100.00	
2	OF	2,429	2.99	66	100.00	
3	TO	2,308	2.84	66	100.00	
4	AND	2,079	2.56	66	100.00	
5	#	1,929	2.37	66	100.00	
6	A	1,525	1.88	66	100.00	
7	THAT	977	1.20	66	100.00	
8	IS	902	1.11	65	98.48	
9	FOR	760	0.93	66	100.00	
10	ON	668	0.82	66	100.00	
11	AS	528	0.65	61	92.42	
12	IT	512	0.63	65	98.48	
13	S	500	0.61	53	80.30	
14	BY	485	0.60	65	98.48	
15	WITH	432	0.53	63	95.45	
16	BE	415	0.51	66	100.00	
17	ARE	400	0.49	64	96.97	
18	THIS	373	0.46	63	95.45	
19	WILL	368	0.45	59	89.39	
20	ENERGY	359	0.44	32	48.48	
21	HAS	351	0.43	62	93.94	
22	NOT	344	0.42	61	92.42	
23	GLOBAL	340	0.42	47	71.21	
24	FROM	337	0.41	62	93.94	
25	WE	328	0.40	49	74.24	
26	HAVE	325	0.40	62	93.94	
27	WORLD	302	0.37	58	87.88	
28	BUT	301	0.37	59	89.39	
29	AT	292	0.36	61	92.42	
30	CLIMATE	289	0.36	26	39.39	
31	ITS	274	0.34	54	81.82	

Figura 2: Lista de Palavras do *corpus* de Relações Exteriores.



N	Key word	Freq	%	Freq	RC	%	eyness	P	lemmas	Set
1	ENERGY	359	0.44	182			.155	531	0.000000	
2	CLIMATE	289	0.36	70			.934	051	0.000000	
3	GLOBAL	340	0.42	381			.706	671	0.000000	
4	NUCLEAR	228	0.28	27			.631	751	0.000000	
5	WARMING	171	0.21	4			.314	151	0.000000	
6	DOHA	159	0.20	0			.257	221	0.000000	
7	CARBON	145	0.18	24			.009	321	0.000000	
8	WORLD	302	0.37	1,185		0.03	.933	38	0.000000	
9	EMISSIONS	127	0.16	28			.868	79	0.000000	
10	TRADE	267	0.33	1,051		0.03	.823	73	0.000000	
11	BRAZIL	153	0.19	135			.816	87	0.000000	
12	FOREIGN	180	0.22	282			.816	47	0.000000	
13	WEAPONS	106	0.13	14			.752	17	0.000000	
14	COUNTRIES	234	0.29	890		0.02	.734	94	0.000000	
15	BUSH	120	0.15	57			.728	54	0.000000	
16	GREENHOUSE	87	0.11	4			.655	17	0.000000	
17	STATES	197	0.24	719		0.02	.631	88	0.000000	
18	CHINA	115	0.14	93			.626	83	0.000000	
19	TALKS	94	0.12	24			.624	92	0.000000	
20	INDIA	108	0.13	67			.623	58	0.000000	
21	UNITED	183	0.23	638		0.02	.600	57	0.000000	
22	FUEL	96	0.12	47			.579	72	0.000000	
23	NATIONS	109	0.13	105			.569	28	0.000000	
24	ROUND	127	0.16	300			.495	96	0.000000	
25	OBAMA	62	0.08	0			.490	16	0.000000	
26	IRAN	65	0.08	7			.488	23	0.000000	
27	GAS	79	0.10	50			.464	26	0.000000	
28	FUELS	60	0.07	3			.460	35	0.000000	
29	OIL	183	0.23	1,106		0.03	.436	67	0.000000	
30	WTO	55	0.07	4			.405	72	0.000000	
31	RENEWABLE	54	0.07	3			.403	52	0.000000	

Figura 3: Palavras-chave do *corpus* de Relações Exteriores.

Em seguida foram geradas linhas de concordância com as palavras ou padrões que se pretendia discutir no curso (quadro 2).

Quadro 2: Linhas de concordância para *global*.

There are many, many other costs, ranging from	<b>global</b>	warming to the toxic pollution of the
human beings as occupying a place within the	<b>global</b>	community which allowed them the
Design Association links designers in a	<b>global</b>	network to provide information,
the farming of sea vegetation as a potential	<b>global</b>	disaster. [p] His triumphant
the rain forests, protect the whales, reduce	<b>global</b>	warming, take a metaphorical axe to
findings. 76 think that people are causing	<b>global</b>	warming, but only 41 think there is
IS DESIGNED TO HELP THE MALDIVES LIVE WITH	<b>GLOBAL</b>	WARMING, NOT TO HELP PREVENT IT. WHEN
GLOBAL WARMING SPREAD TROPICAL DISEASES? If	<b>global</b>	warming does take off and climate
Gulf having concluded that the potential for	<b>global</b>	effects is very small. For global
see Table 4.2). The first version is	<b>global</b>	in scope and historically dilated; the
of Bangladesh now, they will be tenfold under	<b>global</b>	warming. One scenario suggests that 1

No quadro acima já é possível identificar a colocação *global warming*, que se repete seis vezes em 11 linhas.

Os exercícios foram preparados a partir das dificuldades detectadas durante a correção das redações, assim como do vocabulário mais frequente revelado pelas palavras-chave. Preparado o material, ele foi compilado em formato de apostila e distribuído aos candidatos.

## 5. Na sala de aula

O curso foi dividido em quatro sessões de 3 horas cada. As sessões foram realizadas em quatro dias subsequentes.

### 5.1 Primeiro dia: noções de Convencionalidade e Linguística de *Corpus*

Como dissemos, os candidatos foram primeiramente apresentados aos princípios da Linguística de *Corpus*, que parte da observação de grandes quantidades de textos para descrever padrões que se repetem. Ao contrário de abordagens anteriores que privilegiavam a intuição do falante ou pesquisador, ou seja, bastava a forma ser gramaticalmente correta para ser considerada aceitável, isto é, possível. Já a Linguística de *Corpus*, observando a linguagem autêntica em uso, pretende descrever aquilo que é mais provável de ocorrer, justamente por se configurar como um padrão recorrente. Esses padrões podem ser classificados em diversas categorias convencionais que variam de colocações como *foreign policy* e *climate changes*, ou binômios como *Republicans and Democrats*, até unidades fraseológicas mais extensas como *on the other hand* (TAGNIN, 2013; BARNBROOK, MASON e KRISHNAMURTHY, 2013).

Em seguida, os candidatos foram apresentados a linhas de concordância e instruídos na forma de como ler esse material a fim de identificar padrões recorrentes e sentidos diversos de uma mesma palavra (TRIBBLE). Num primeiro momento, os alunos estranham esse formato por não apresentar períodos completos. Mas o estranhamento desaparece quando lhes é explicado que uma leitura vertical identifica padrões recorrentes, como vimos no quadro 2, e uma leitura horizontal das linhas salienta sentidos diversos, como se observa no quadro 3, que apresenta três sentidos diferentes de *issue*: exemplar de uma revista (linhas 2, 4 e 5), ‘assunto’, ‘questão’ (linhas 1, 3, 6 e 8) e, finalmente, como o verbo ‘anunciar’ ou ‘emitir’ (linha 7). Uma leitura mais cuidadosa ainda irá revelar, na linha 9, a colocação verbal *take issue with*, que significa ‘discordar’.

Quadro 3: Linhas de concordância para *issue*.

1.	eld convictions and stirs ancient taboos. The <b>issue</b> needs to be tackled with
2.	only commitment is to choose one item from each <b>issue</b> of the magazine minimum
3.	a parting of the ways over a money or moral <b>issue</b> . Let common sense prevail-
4.	Wedgwood Designs For Brides', before the next <b>issue</b> of Brides is published,
5.	fixtures. The first league table will appear in <b>ISSUE</b> 176 - ie: in two weeks'
6.	the disaster aid programme is a humanitarian <b>issue</b> which transcends politics.
7.	be dragged into it. I think if the Americans <b>issue</b> a unilateral ultimatum to
8.	the biggest obstacle to negotiations on that <b>issue</b> . It's a view echoed by Mr.
9.	incompetent. If a hospital or the state took <b>issue</b> with Ken's directives about

Como não se esperava que os candidatos construíssem, de imediato, um *corpus* especializado que atendesse suas necessidades, apresentamos alguns *corpora* que podem ser consultados *on-line*, dentre eles três *corpora* que fazem parte da plataforma de Mark Davies, da Brigham Young University, quais sejam: o *Corpus of Contemporary English*<sup>2</sup> (COCA), o *British National Corpus*<sup>3</sup> (BYU-BNC) e o *Corpus do Português*<sup>4</sup>. Outros *corpora* foram o Lácio-Ref<sup>5</sup>, um *corpus* de português com 9 milhões de palavras, o COMPARA<sup>6</sup>, um *corpus* bilíngue inglês-português com textos literários e o projeto COMET<sup>7</sup>, que é constituído de um *corpus* técnico, o CorTec<sup>8</sup>, com textos em inglês e português em mais de 20 áreas técnicas, e um *corpus* de traduções, o CorTrad<sup>9</sup>, com originais em inglês e português e respectivas traduções. Como tínhamos acesso à internet, demonstramos como fazer buscas em cada um

<sup>2</sup> [www.americancorpus.org](http://www.americancorpus.org)

<sup>3</sup> [corpus.byu.edu/bnc/](http://corpus.byu.edu/bnc/)

<sup>4</sup> [www.corpusdoportugues.org](http://www.corpusdoportugues.org)

<sup>5</sup> [www.nilc.icmc.usp.br/lacioweb](http://www.nilc.icmc.usp.br/lacioweb)

<sup>6</sup> [www.linguateca.pt/COMPARA](http://www.linguateca.pt/COMPARA)

<sup>7</sup> [www.fflch.usp.br/dlm/comet](http://www.fflch.usp.br/dlm/comet)

<sup>8</sup> [www.fflch.usp.br/dlm/comet/consulta\\_cortec.html](http://www.fflch.usp.br/dlm/comet/consulta_cortec.html)

<sup>9</sup> [www.fflch.usp.br/dlm/comet/consulta\\_cortrad.html](http://www.fflch.usp.br/dlm/comet/consulta_cortrad.html)

desses *corpora*. A maioria dos candidatos mostrou-se bastante entusiasmada com esses recursos.

## 5.2 Segundo dia: trabalho com concordâncias

A partir dos problemas identificados durante a correção, foram elaborados os exercícios que compuseram a apostila recebida pelos candidatos. Um dos aspectos mais recorrentes foi o uso indevido de *what* em lugar de *which*, como no trecho abaixo:

Another consequence to Brazilian diplomacy – and to the government as a whole – is the increase in international pressure to enhance its forests protection, what would give rise to the enlargement in Itamaraty’s scope of action.

A fim de conscientizar os candidatos para a diferença entre *what*, *which* e *that* geramos 60 linhas de concordância para cada um desses elementos, das quais transcrevemos dez de cada abaixo (quadro 4). As instruções pedem que os candidatos comparem os três conjuntos para descobrir a forma como cada um dos elementos é usado:

Quadro 4: Linhas de concordância para *what*, *which* e *that*.

I. Compare the three sets of concordance lines below and see if you can tell the different ways in which the search words are used.

1	sp of procedure, someone who knows	<b>what</b>	an official class is. If I don
2			Do you know
3	thing. We stop to look at	<b>what</b>	anti-lock brakes are for? They
4	e Quality Street But	<b>what</b>	appears to be a roadside memor
5	g dismissed. "What can I say?" he said. "I was dr		are the local schools like?
6	will do to computers	<b>what</b>	can I say?" he said. "I was dr
7	both Max and I by the arm . . . " What? Did the copy ed		computers did to slide rules?
8	to kids like Lionel, I don't know	<b>what</b>	Did the copy ed
9	e way its comedy steadily darkens. What else could one exp		does," Lewis said. "It
10	o libraries and bookstores to ask "what else do you have?"		What else could one exp
1	ary Road'' -- a feature with	<b>which</b>	I myself have no complaint. '
2			VW, which struggled in the early 1990s,
3	Image] The Senate vote,	<b>which</b>	struggled in the early 1990s,
4	words similar to the old version,	<b>which</b>	may come Friday, will
5	planted at the right time,	<b>which</b>	refers to "an
6	lebrity, particularly that	<b>which</b>	is October and November.
7	an Football Champion-ship, specify	<b>which</b>	inhabits the House of Windsor
8	he most valued in society,	<b>which</b>	prize you're af
9	alker--the aptly named Sea Shanty,	<b>which</b>	might be fine if the exercise
10	known as Liverpool rummy,	<b>which</b>	is smaller than the P
1	ependent BBC has accepted,	<b>that</b>	there should be no tactless me
2	or in chief of Cruise Week, agrees	<b>that</b>	cruise ships
3	imity and calmly authoritative air	<b>that</b>	the reader can almost
4	n't built in a day and all	<b>that</b> ,	I still believe Mr I can turn
5	ay presented a Democratic analysis	<b>that</b> ,	she said, showed that the Rep
6	ter the Labor Department announced	<b>that</b>	consumer prices edged up just
7	week, announcing	<b>that</b>	gold-gold collisions had produ
8	ople's Lottery. Camelot can argue	<b>that</b>	it was carrying
9	Gore himself has argued	<b>that</b>	his dual origins gave him a pe
10	rats who have been arguing	<b>that</b>	budget surpluses won't be enou

Após alguns minutos, pedimos que apresentassem suas conclusões, que íamos transcrevendo no quadro. Ao final da discussão, projetamos um resumo dos diferentes usos em *slides* do PowerPoint, que foram disponibilizados ao público ao término do curso. O quadro 5 mostra parte de um dos slides salientando duas das funções de *what*.

Quadro 5: Parte do *slide* apresentando algumas funções de *what*.

- **Pronome**
- 5 g dismissed. "What can I say?" he said. "I was dr
- 23 "What is the most suffering? What is less suffering?" As
- 34 sed, "Scientific Dilemmas: What's Lurking Out There in Space?"
  
- **Introdução de uma oração que complementa verbo, preposição ou conectivo**
- 6 will do to computers what computers did to slide rules?
- 13 The CHC put together a case about what had happened to the family.
- 15 Everybody in this room knows what happened there. That Ja

Outro problema recorrente nas redações foi o uso inadequado de preposições, em especial *on*. Uma concordância com uma seleção de ocorrências com essa preposição foi apresentada aos candidatos solicitando que identificassem os padrões em que ocorre. O quadro 6 traz um recorte dessa linhas:

Quadro 6. Seleção de linhas de concordância para a preposição *on*.

45	dinner invitations. "On the other hand," Bloodworth-Thom
48	ect of spending last night on the streets. 'There were
49	troom. "TV violence is not on trial. Professional wrestling is
50	all nine of their number on the Judiciary Committee vote aga
51	plan to detail their objections on the Senate floor before the fina
52	Hughes, Mr Mehmet had told police on at least ten occasions that he k
53	will invite us to reflect on the beastliness of the Germans o
55	inter. Alexander said on NBC's "Meet the Press" that ther
56	out on the streets,' Ms Casey said on Today . 'I reach a stage
57	g or deliver the Socratic seminars on litter laws and good cit
58	leased from prison and sent on community service programmes ins
59	t... different. Policies are short on ideas and entail more
60	0 witnesses, including specialists on voting rights and voting
61	o monitor and control big spending on television and print ads be used
63	he ANC, with its eccentric stances on Aids and press freedom, proved t
64	gest Bush administration statement on the subject to date, Fle
67	y. "My personal views on abortion are well known," Ashcro
68	d order issues. However, his views on certain issues -

Além de identificarem o conector *On the other hand*, também observaram o uso da preposição com o verbo *spend* (*spending last night on the streets*, *big spending on television and print ads*), com o adjetivo *short* (*short on ideas*), com o substantivo *views* (*views on abortion*, *views on certain issues*), entre outros. No entanto, o que lhes chamou a atenção foi *including specialists on voting rights* (linha 60), pois acreditavam que a preposição correta

seria *in*. De fato, é o que os dicionários apresentam. Por exemplo, o *Macmillan English Dictionary for advanced learners of American English* (2006) define *specialist* como

Someone whose training, education, or experience makes them an expert in a particular subject: *a web design specialist* **a**. a doctor who is an expert in a specific area of medical work: *Both her husband and brother are specialists in hand surgery.* (p. 1350)

Nota-se que, tanto na definição, quando emprega o sinônimo *expert* (*expert in a particular subject, expert in a specific área of medical work*), quanto no exemplo na acepção **a**, aparece a preposição *in*: *specialists in hand surgery*.

Já o *Oxford Collocations dictionary for students of English* (2002), na entrada para *specialist* em que elenca as preposições com que o substantivo co-ocorre, dá exemplos com ambas as preposições:

~**in** She is a specialist in eighteenth-century English painting. ~**on** a specialist on the history of this city. (p. 733)

No entanto, não oferece qualquer explicação para diferenciar o uso de uma ou outra. Numa tentativa de identificar alguma diferença, buscamos outros exemplos, que apresentamos em *slides* (Quadros 7 e 8):

Quadro 7: *slide* com exemplos para *specialist in*.

➤ Specialist <b>in</b>
<ul style="list-style-type: none"> <li>• Charles K Hole, specialist <b>in restoration of antique furniture</b></li> <li>• Dental Association, Specialist <b>in Community Medicine</b></li> <li>• another specialist <b>in cancer medicine</b> is</li> <li>• Garratt &amp; Co. is a specialist <b>in assisting firms</b> to make</li> <li>• Maria Balinska, a specialist <b>in North American affairs</b>,</li> <li>• David Willis, a specialist <b>in US affairs</b>, reports.</li> <li>• officer, a specialist <b>in Oriental languages</b></li> <li>• becoming a specialist <b>in his chosen area</b>.</li> </ul>

Quadro 8: *slide* com exemplos para *specialist on*.

➤ Specialist <b>on</b>
<ul style="list-style-type: none"> <li>• Jonathan Fryer, a specialist <b>on southern African affairs</b>, explains.</li> <li>• Jacques Bekaert, a specialist <b>on Indochina</b>,</li> <li>• the well-known Daily Express specialist <b>on defence and intelligence</b>,</li> <li>• the cla's top specialist <b>on the Soviet Union</b>.</li> <li>• specialist <b>on animal behaviour</b>.</li> <li>• A specialist <b>on defense issues</b>, he covered US</li> <li>• Gary Milhollin, a specialist <b>on nuclear proliferation</b></li> <li>• Dr. Klaus Peter Treiter, a specialist <b>on European issues</b>,</li> </ul>

Embora não se possa estabelecer uma diferença categórica, pois, por exemplo, as duas preposições ocorrem com *affairs*, é possível postular que há uma tendência para o uso de *in* quando se refere a uma área (*câncer medicine, Community medicine, oriental languages*) ou a serviços a serem prestados (*restoration of antique furniture, assisting firms*), enquanto *on* ocorre quando se trata de assuntos (*issues*) específicos (*Indochina, Soviet Union, animal behaviour, defense issues, nuclear proliferation, European issues*).

Com essa questão, os candidatos se familiarizaram com uma forma de esclarecer dúvidas, ou seja, recorrer a um *corpus* para buscar exemplos e compará-los.

Uma palavra que apresentou vários erros colocacionais foi *opportunity*. Um dos problemas foi a preposição que a segue: *to* ou *for*. Para essa questão, o exercício foi comparar concordâncias com uma e outra preposição. Mas também ocorreram erros com os verbos que co-ocorrem com *opportunity*. Além disso, notamos que os candidatos pouco variavam os adjetivos com os quais a palavra co-ocorria, privilegiando, na maioria dos casos, *good* ou *great*. Dessa forma, foi-lhes apresentado um exercício em que deveriam identificar os adjetivos que qualificavam *opportunity* e os verbos que co-ocorriam com esse substantivo. O quadro 9 apresenta uma seleção das linhas de concordância do exercício:

Quadro 9: Seleção de linhas de concordância para *opportunity/opportunities*.

20	Patrick Byrne, however, the biggest	<b>opportunity</b> is in picking through
21	decades. "The NHS now has the best	<b>opportunity</b> it has ever had to bri
22	aid. "We think he gives us the best	<b>opportunity</b> right now to win a bas
23	eir slowness in grasping commercial	<b>opportunities</b> offered by the inter
24	d was later revived, and gave early	<b>opportunities</b> to young directors i
25	s neighbourhood, makes people enjoy	<b>opportunities</b> and ideas which were
26	the time and they will seize every	<b>opportunity</b> to get into the game."
27	vels. Amidst the risks are exciting	<b>opportunities</b> , too. A clear signal
28	ng year, and a range of fascinating	<b>opportunities</b> have recently arisen
29	e chance, they will seize the first	<b>opportunity</b> to leap out of the tra
30	No Beat of Drum (1966), which gave	<b>opportunity</b> for her to have her pr
31	y scoring chances. "We were getting	<b>opportunities</b> and were not putting
32	ement years could provide a golden	<b>opportunity</b> for Chile. It could be
33	d to take advantage of every golden	<b>opportunity</b> . The Boilermakers did
34	ng, they simply spurned more golden	<b>opportunities</b> than any Premiership
35	ove to manage," he said. "If a good	<b>opportunity</b> comes up, I would jump
36	p, I would jump at it. I had a good	<b>opportunity</b> here, and it didn't wo
37	place where you can find some good	<b>opportunities</b> ." The tight labor ma
38	igned this offseason. "It's a great	<b>opportunity</b> for me to come in ther
39	attitude. "The states have a great	<b>opportunity</b> to change public healt
40	se free hours are no longer a great	<b>opportunity</b> but become filled with
41	teachers, and in some cases greater	<b>opportunities</b> to get involved in y
42	small school, you will have greater	<b>opportunities</b> for knowing your tea
43	training courses, may offer greater	<b>opportunities</b> for customizable pub
44	s him the chance, he will seize his	<b>opportunity</b> with both hands. "Thin
45	ut taking advantage of the historic	<b>opportunity</b> that will present itse
46	can take, but also what interesting	<b>opportunities</b> you might have to se
47	cow market for promising investment	<b>opportunities</b> . Logoshin ascribes C
48	e says, is about family and "missed	<b>opportunities</b> ." So often, she says
49	But it is also a tale of missed	<b>opportunities</b> - how an impressive t
50	try's recent expansion present more	<b>opportunities</b> for career advanceme

As colocações identificadas foram resumidas no seguinte *slide*:

Quadro 10: *Slide* com o resumo das colocações para *opportunity/opportunities*.

- **Adjectival collocations:** ample, biggest, best, early, exciting, fascinating, golden good, great, greater, historic, interesting, new, rare, perfect
- **Verbal collocations:** offer, give, have, miss, provide, pursue, grasp, seize, take advantage of, spurn, open up
- **Intensifiers:** more, a lot of, plenty of, a wealth of

Alguns dos outros tópicos abordados foram as diferenças entre

- the /a(n) / zero article
- concerning / regarding / regard / regards
- despite / in spite of
- economy / economic / economical
- such / like
- big / large /great
- reduce / decrease
- increase / enlarge / enhance

### 5.3 Terceiro dia: avaliação em sala de aula

O terceiro dia foi dedicado a exercícios de averiguação do aprendizado, sem qualquer tipo de avaliação formal. Um dos exercícios requeria o preenchimento de lacunas em trechos de textos jornalísticos relacionados a assuntos que foram identificados a partir das palavras-chave do *corpus*. Dessa forma o candidato estaria trabalhando com textos pertinentes ao universo de sua atuação futura (Figura 12).

Quadro 11: Exercício de preenchimento de lacuna.

\_1\_ United States and \_2\_ allies have long recognized \_3\_ power of \_4\_ jihadi narrative, but \_5\_ attempts to overcome it have been ill-conceived. \_6\_ Al Qaeda's story justifies \_7\_ violence by claiming \_8\_ mantle of \_9\_ victimhood, but \_10\_ Bush administration responded like \_11\_ boxer, hitting back with \_12\_ "war on terror" and \_13\_ "battle of ideas." \_15\_ war metaphor only played into \_16\_ terrorists' hands, and \_17\_ Obama administration is rightly moving away from it.

No exercício acima o candidato deveria preencher as lacunas com o artigo definido *the* ou indefinido *a(n)*, quando necessário. O texto corretamente preenchido foi apresentado em *slide* (quadro 12):

## Quadro 12: Exercício de preenchimento de lacuna completado.

**The United States and its allies have long recognized the power of the jihadi narrative, but Ø attempts to overcome it have been ill-conceived. Ø Al Qaeda's story justifies Ø violence by claiming the mantle of Ø victimhood, but the Bush administration responded like a boxer, hitting back with a "war on terror" and a "battle of ideas." The war metaphor only played into the terrorists' hands, and the Obama administration is rightly moving away from it.**

Outro tipo de exercício consistia de concordâncias com apagamento da palavra de busca. No exercício abaixo o candidato deveria descobrir qual forma de *concern\** ou *regard\** seria correta.

Quadro 13 : Exercício de preenchimento da palavra de busca<sup>10</sup>.

In the following exercise, try to figure out what the missing word is.

a. Concern\*/regard\*

1	cans support the key recommendations of the Iraq Study Group	the withdrawal of U.S. troops and talking with I
2	elming majority of both U.S. and international media reports	"global warming" as a foregone conclusion is tha
3	irely upon climate models. We all know the frailty of models	the air-surface system." – Atmospheric scientist
4	SCO in place, the future for export may require concessions	the nuclear issue with Iran. • The U.S. is also
5	e South. It would allow developing economies to grow without	to atmospheric limits-and without the budgetary
6	China, Japan, Russia, and India, seek to emulate it in this	, the already voracious military component of glo
7	d countries with diminishing reserves, and security concerns	Iranian profits from inflated oil prices. The an

Como, em alguns casos, *concerning* e *regarding* podem ser usados como sinônimos, foram dadas as duas opções quando pertinente (quadro 14):

## Quadro 14: Exercício de preenchimento da palavra de busca completado.

1	cans support the key recommendations of the Iraq Study Group	<b>concerning / regarding</b>	the withdrawal of U.S. troops and talking with I
2	elming majority of both U.S. and international media reports	<b>concerning / regarding</b>	"global warming" as a foregone conclusion is tha
3	irely upon climate models. We all know the frailty of models	<b>concerning / regarding</b>	the air-surface system." – Atmospheric scientist
4	SCO in place, the future for export may require concessions	<b>concerning / regarding</b>	the nuclear issue with Iran. • The U.S. is also
5	e South. It would allow developing economies to grow	<b>without regard to</b>	atmospheric limits-and without the budgetary
6	China, Japan, Russia, and India, seek to emulate it	<b>in this regard</b>	, the already voracious military component of glo
7	d countries with diminishing reserves, and security concerns	<b>regarding</b>	Iranian profits from inflated oil prices. The an

Por fim, os candidatos receberam nove textos para serem traduzidos para o português, uma vez que, em seu futuro cargo, deveriam estar aptos a compreender e eventualmente traduzir textos em inglês. A correção, a partir de uma tradução proposta, suscitou várias sugestões por parte dos candidatos, em especial quando se tratava de terminologia específica da área. Assim, enquanto a tradução sugerida – e apresentada em slide – apresentava as opções:

“integrantes/membros do Escritório do Representante de Comércio/Comercial dos Estados Unidos / os representantes do comércio dos Estados Unidos”

<sup>10</sup> A fonte dessas concordâncias foi reduzida para se adequar ao espaço da revista. Na apostila, elas aparecem em formato paisagem.

os candidatos sugeriram “integrantes/membros do Escritório de Representação Comercial dos Estados Unidos”, opção não apresentada na tradução proposta.

Em outro texto, a lista de sugestões para traduzir o adjetivo *overwhelming* em

The solution to the climate crisis is as simple as it is overwhelming.

foi bem extensa: ‘atordoante’, ‘arrebataadora’, ‘gigantesca’, ‘hercúlea’, ‘homérica’, ‘complexa’, enquanto a tradução proposta foi ‘impressionante’.

Nos exercícios de tradução os candidatos pareciam estar mais à vontade e mais motivados a contribuir com sugestões diretamente relacionadas a sua futura área de atuação.

#### 5.4 Exercícios de versão

O último dia foi dedicado a exercícios de versão para o inglês, que os candidatos deveriam ter feito em casa. Entretanto, poucos o fizeram, certamente por falta de tempo, uma vez que no outro período do dia assistiam a um curso preparatório de português. Porém, como já havíamos preparado uma versão para ser apresentada e, principalmente, para ser discutida com os candidatos, o exercício foi bastante proveitoso, principalmente porque muitas sugestões ou dúvidas puderam ser esclarecidas por meio de consulta ao COCA, uma vez que tínhamos acesso à internet e a tela de busca podia ser projetada para visualização da classe.

#### 6. Resultados

O método de revisão utilizando a Linguística de *Corpus* foi aplicado com resultados satisfatórios, conforme se pode observar no decorrer das atividades. O contato com uma metodologia ainda pouco conhecida no ensino de línguas estrangeiras foi bastante motivante, em especial por se tratar de uma abordagem computacional, universo com que os candidatos estão familiarizados. Em avaliação informal a maioria dos alunos expressou satisfação em relação ao método, que correspondeu às suas expectativas de revisão do conteúdo face ao curto tempo disponível, embora alguns tenham criticado o fato de “sempre se recorrer ao *corpus*” para sanar alguma dúvida. Enquanto para os primeiros a Linguística de *Corpus* representou uma porta para um novo universo, que poderiam explorar por si, para os últimos representava uma falha no conhecimento da ministrante do curso, que, em caso de dúvidas, “tinha de recorrer ao *corpus*”. A esses interessavam “respostas corretas” que deveriam ser

fornecidas pela professora. No geral, no entanto, afirmaram que ficaram estimulados a passar a usar a metodologia em suas vidas profissionais.

## 7. Considerações finais

Como era de se esperar, a Linguística de *Corpus* mostrou-se uma abordagem pedagógica eficaz para o tipo de curso descrito – uma rápida revisão gramatical da língua inglesa, visando também a tradução e a versão de textos da área de Relações Exteriores. O resultado alcançado comprovou que a abordagem pode ser replicada para qualquer área profissional, desde que haja *corpora* disponíveis. Quando não os há, construir um *corpus* para esse fim é tarefa bastante fácil e rápida, com vimos acima. Quanto aos alunos, uma vez familiarizados com a metodologia, esses podem tornar-se pesquisadores capazes de buscar respostas para questões que eventualmente surjam durante seu aprendizado, imergindo-se em um processo autodidata de aprendizagem. Ressalte-se, entretanto, que sempre haverá os que preferem depender do conhecimento do professor a desenvolver sua própria autonomia de aprendizado.

## Referências bibliográficas

- BARNBROOK, G.; MASON, O.; KRISHNAMURTHY, R. **Collocations: Applications and Implications**. Basingstoke: Palgrave Macmillan, 2013.  <http://dx.doi.org/10.1057/9781137297242>
- BARONI, M.; BERNARDINI, S. **BootCat - Bootstrapping corpora and terms from the web**. Proceedings of LREC 2004 Conference. Lisboa: [s.n.]. 2004. p. 1313-1316.
- BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.
- BERNARDINI, S.; FERRARESI, A. Old needs, new solutions - comparable *corpora* for language professionals. In: SHAROFF, S., et al. **Building and Using Comparable Corpora**. Berlin - Heidelberg: Springer Verlag, 2013. p. 303-319.  [http://dx.doi.org/10.1007/978-3-642-20128-8\\_16](http://dx.doi.org/10.1007/978-3-642-20128-8_16)
- HALLIDAY, M. A. K. Categories of the theory of grammar. **Word**, Vol. 17, No. 3 1961. 241-292.
- JOHNS, T. Whence and whither classroom concordancing? In: BONGAERTS, T., et al. **Computer applicatons in language learning**. Dordrecht: Foris Publications, 1988. p. 9-27.
- JOHNS, T. Should you be persuaded: two samples of data driven learning. **ELR**, vo. 4 1991. 27-46.

JOHNS, T. **Data-driven Learning: The Perpetual Challenge**. Proceedings of the Fourth International Conference on Teaching and Language Corpora - Graz, 19-24 July 2000. Amsterdam / New York: Rodopi. 2002. p. 107-117.

MACMILLAN EDUCATION. **Macmillan English Dictionary for advanced learners of American English**. Oxford: Macmillan Publishers Limited, 2006.

OXFORD UNIVERSITY PRESS. **Oxford Collocations dictionary for students of English**. Oxford: Oxford University Press, 2002.

SCOTT, M. **WordSmith Tools**. Oxford: Oxford University Press, 2007.

TAGNIN, S. E. O. **O Jeito que a Gente Diz**. São Paulo: Disal, 2013.

TRIBBLE, C. **Concordances in the classroom**. [S.l.]: [s.n.].

VARANTOLA, K. Disposable corpora as intelligent tools in translation. **Cadernos de Tradução**, 2002. 171-189.

Artigo recebido em: 15.10.2014

Artigo aprovado em: 02.12.2014