

Tony Berber Sardinha
tony4@uol.com.br

Preparação de material didático para Aprendizagem Baseada em Tarefas com WordSmith Tools e corpora¹

Preparing teaching materials for Task-Based Learning with WordSmith Tools and corpora

RESUMO - Neste trabalho, apresento algumas maneiras de usar o programa de análise de corpus WordSmith Tools (Scott, 1997) na preparação de material didático dentro de uma proposta de Aprendizagem Baseada em Tarefas para ensino de inglês como língua estrangeira. Espero com isso contribuir para divulgar o potencial do programa na área de ensino de línguas estrangeiras.

Palavras-chave: lingüística de corpus, Aprendizagem Baseada em Tarefas, WordSmith Tools.

ABSTRACT - In this paper, I present a few ways in which the WordSmith Tools (Scott, 1997) corpus analysis suite can be used to prepare teaching materials within the framework of Task-Based Learning for teaching English as a Foreign Language. As such, I hope to contribute to spreading the potential of the software among practitioners in the field of EFL.

Key words: corpus linguistics, Task-Based Learning, WordSmith Tools.

Introdução

Neste trabalho, apresento algumas maneiras de usar o programa de análise de corpus WordSmith Tools (Scott, 1997) na preparação de material didático dentro de uma proposta de Aprendizagem Baseada em Tarefas (Task-Based Learning) para ensino de inglês como língua estrangeira (Estaire e Zanon, 1994; Nunan, 2004; Prabhu, 1987; Skehan, 1996; Willis, 1996). Embora tratemos da língua inglesa, não queremos com isso dizer que os procedimentos mostrados aqui sejam válidos apenas para o inglês; muito pelo contrário, eles podem ser aplicados a diversas outras línguas, estrangeiras ou não.

O programa WordSmith Tools auxilia o professor ou preparador de material ao permitir a identificação de padrões léxico-gramaticais. Um padrão léxico-gramatical é uma seqüência recorrente de (pelo menos duas) palavras, dentro de um espaço delimitado (normalmente equivalente a até quatro palavras de distância), que possui um sentido específico (Sinclair, 1991). Dentre os vários tipos de padrão existentes, o que é mais saliente são os agrupamentos ('clusters'), tam-

bém conhecidos por 'bundles' (pacotes) (Biber e Conrad, 1999). Agrupamentos são seqüências fixas de palavras, como 'bom dia' ou 'que possui um'. Esses padrões são bastante salientes devido à sua rigidez e assim se tornam bons padrões para ensinar em aula de inglês, pois são facilmente notados. A existência de padrões em abundância faz com que possamos falar em uma natureza associativa e probabilística da linguagem, como princípios teóricos de organização em larga escala. Dizemos natureza associativa porque as palavras tendem a se associar em grupos padronizados, recorrentes; e falamos em caráter probabilístico porque essas associações não são aleatórias: as palavras se combinam umas às outras com graus diferentes de probabilidade.

Do ponto de vista da sala de aula, há várias razões que salientam a importância de padrões no aprendizado de língua estrangeira. Os padrões são um tipo de 'unidade pré-fabricada' da língua, que parece residir na memória em forma 'bruta', isto é, sem estar decomposto em unidades individuais. Assim, um agrupamento como 'que possui um' parece estar disponível por inteiro na memória, para produção e recepção, na mente dos falantes nativos. Quando

¹ Quero agradecer ao CNPq (Brasília, Brasil) o apoio financeiro (processo 350455/2003-1). Também sou grato às seguintes pessoas pela leitura e comentários: Estela de Jesus Martins, Lilian de Mello Martins e Lucimar Canonico di Santi.

falamos, possivelmente acessamos nosso estoque lexical mental e retiramos de lá tal unidade semi-pronta e a utilizamos *in natura* na fala e na escrita. Do mesmo modo, ao processar a linguagem falada ou escrita, é possível que decomponhamos a seqüência de fala ou escrita em unidades compostas, muitas vezes em forma de padrões. Esse acesso a unidades pré-fabricadas permite uma maior rapidez no processamento, o que, por sua vez, sustenta nossa capacidade natural de falar e escrever fluentemente, de ouvir e entender de modo síncrono com o nosso interlocutor e de (muitas vezes) poder ler com rapidez, sem nos deter a cada palavra individual do texto (Sinclair, 1991). Embora essas sejam habilidades naturais que todo falante nativo possui, em maior ou menor grau, elas não são automaticamente transferidas para a língua estrangeira. Normalmente, os aprendizes de língua estrangeira tendem a ver a linguagem como um conjunto de palavras individuais sustentadas por regras gramaticais; desse modo, a compreensão e a produção se dão, em muitos casos, com base na junção de palavras individuais. O resultado disso é que, por exemplo, a fluência, um aspecto natural da fala na língua materna, é conseguida somente a duras penas na língua estrangeira.

O trabalho com corpora em sala de aula pode ajudar a melhorar essa situação, pois ele naturalmente faz saltar aos olhos essas ‘unidades pré-fabricadas’, que são os padrões léxico-gramaticais. Ou seja, começando por um trabalho de conscientização, o aluno passa a ter uma visão diferente do que é o vocabulário de uma língua, deixando de lado a idéia de que se trata de um conjunto de palavras isoladas que se juntam por meio de regras gramaticais. Ele passa a perceber que as palavras se juntam por meio de atração mútua, via de regra explicadas somente pelo uso e não por regras de gramática, e que essa atração cria os agrupamentos, pacotes e *chunks*, que por sua vez se juntam e formam o tecido da linguagem.

Com a ajuda de corpus e do programa WordSmith Tools, podemos começar a propor algumas maneiras de lidar com essa questão do uso coerente de corpus e texto autêntico na sala de aula.

Preparação de material com corpora

Há diversas maneiras de produzir atividades de ensino de língua estrangeira com corpus. Lewis (1997) elenca vários tipos de exercícios que podem ser criados para ensinar colocações. Esses exercícios podem ser integrados a diversos tipos de unidades didáticas, sem se filiar a nenhuma linha teórica de aprendizagem específica.

No âmbito do ensino de inglês geral (‘General English’), há várias pesquisas sobre como integrar atividades produzidas com base em corpora na sala de aula. Condi de Souza (2005) desenvolveu atividades online com corpora de complementação do livro didático adotado em escola de inglês. Vicentini (2006) também preparou atividades de complementação ao livro didático com DVDs,

aproveitando um corpus de uma série de televisão americana. Bértoli Dutra (2002) criou atividades de ensino a partir de corpora de letras de música. Por fim, Amarante (2005) preparou material para ensino de pronúncia do inglês, suprindo lacunas do livro didático a partir de pesquisa com corpora.

Do mesmo modo, na esfera de Inglês Instrumental, há pesquisas que visaram à produção de material instrucional com corpora. Ferrari de Oliveira (2004) desenvolveu unidades de ensino para um curso de comércio exterior, com uma base sócio-interacionista. Barbosa (2004) preparou uma unidade de curso de inglês a distância para alunos da área de negócios, fundamentado na abordagem experiencial (Kohonen, 2001).

No caso da língua espanhola, também há pesquisas voltadas à produção de material didático com corpora. Jacobi (2001) preparou material didático para ensinar palavras polissêmicas problemáticas do espanhol. Alonso (2006) desenvolveu atividades de ensino de falsos cognatos, para alunos de segundo grau e universitários.

Uma tendência geral de todos esses trabalhos é a necessidade de enquadrar as atividades de ensino em uma proposta coerente de preparação de material didático, a fim de que as atividades não sejam apenas exercícios soltos de manipulação da língua como um fim em si mesmo.

O material desenvolvido com corpora se encaixa em várias teorias. Porém, a meu ver, esse tipo de material é mais coerente com propostas pedagógicas de cunho construtivista, em que o aluno passa a descobrir os padrões de co-ocorrência de que o corpus dispõe, constrói e testa suas próprias hipóteses, como é o caso do Data Driven Learning (Johns, 1991), em que o aluno passa a desempenhar o papel de pesquisador, analisando concordâncias e tirando suas próprias conclusões sobre o uso e o sentido das palavras estudadas. Do mesmo modo, atividades com corpora me parecem afinadas com propostas que defendam a ligação do trabalho em sala de aula com o mundo fora da escola. Uma dessas propostas é a da Aprendizagem Baseada em Tarefas (Task Based Learning), que em algumas de suas vertentes entende tarefa como uma prática social, uma atividade que desempenhamos genuinamente como parte de nossos afazeres cotidianos, de modo profissional ou não. Tais tarefas se concretizam discursivamente, de modo mais ou menos rotinizado, e recebem nomes socialmente aceitos. Elas são, do ponto de vista lingüístico-discursivo, gêneros (Bhatia, 1993; Eggins *et al.*, 1997; Martin, 1997; Meurer *et al.*, 2005; Swales, 1990): eventos comunicativos institucionalizados em determinados grupos sociais (Motta-Roth, 2006). Alguns exemplos de gêneros são reuniões de trabalho, artigos acadêmicos, mensagens de *e-mail* de reclamação/informação, *blogs*, receitas culinárias, contratos bancários, encontros de serviço (‘service encounters’) variados como a interação para comprar pão e leite na padaria, pagar contas no banco, ter uma consulta médica, pegar um táxi, entre muitos outros.

A seguir, ilustrarei como se poderia empregar os achados da análise de corpus mostrada acima na preparação de uma atividade de ensino dentro de um modelo de Aprendizagem Baseada em Tarefas, ligada a um gênero. Há várias propostas chamadas de Task Based Learning na literatura, como as de Willis (1996), Prabhu (1987), Ellis (2003), Skehan (1996) e Estaire e Zanon (1994), cada um com uma definição diferente de tarefa e com subtipos de tarefa. Esta proposta seguirá a de Estaire e Zanon (1994), pois ela se afina com a visão de tarefa que apresentei acima. Uma tarefa comunicativa é:

a piece of classroom work during which learners' attention is principally focused on meaning rather than form, that is, on what is being expressed rather than on the linguistic forms used for expressing it. (...) [It] is a piece of classroom work which, as far as possible, resembles activities which our students or other people carry out in everyday life, thus reproducing processes of everyday communication (Estaire e Zanon, 1994, p. 14).

Essa definição enfatiza o papel da troca e construção de significado na constituição da tarefa. Ou seja, o foco dela não é a linguagem, mas a consecução de uma atividade que existe além dos limites da sala de aula. Embora a linguagem seja um elemento constitutivo da tarefa, pois sem ela a atividade não aconteceria, não é o objetivo da tarefa. Assim, uma tarefa comunicativa pode ser algo como 'fazer uma matéria de jornal', 'escrever um texto de *blog*', 'realizar uma entrevista de emprego', entre outras, mas não 'aprender o *present perfect*' ou os tipos de futuro em inglês.

Por sua vez, as tarefas de apoio ('enabling task'):

act as support for communication tasks. Their purpose is to provide students with the necessary linguistic tools to carry out a communication task. Though they can be as meaningful as possible, their main focus is on linguistic aspects (grammar, vocabulary, pronunciation, functions, discourse) rather than on meaning. They are overt language meaning experiences, whose aim is to enable students to communicate as smoothly and effectively as possible (Estaire e Zanon, 1994, p. 15).

Com base nessas definições, Estaire e Zanon (1994, p. 5) apresentam um modelo para desenvolvimento de uma unidade didática, pautado nesse princípio da centralidade do sentido. O modelo é composto de seis itens: tema, tarefa comunicativa, objetivos, conteúdo, seqüência didática e avaliação. Esses itens são ilustrados na proposta a seguir:

Nível: Intermediário ou avançado

- (1) Tema: Infecções fatais como Aids e gripe aviária no mundo atual.
- (2) Tarefa comunicativa (a ser desenvolvida no final da unidade): Escrever um texto para *blog* a respeito dessas infecções e como evitá-las.
- (3) Objetivos. Durante a tarefa os alunos desenvolverão a habilidade e o conhecimento necessários para:

- Discutir com os colegas a respeito dessas infecções.
- Propor maneiras de evitá-las.
- Identificar as regiões de risco no mundo.
- Escrever um texto para ser publicado em um *blog* alertando os leitores sobre os perigos dessas infecções, os locais de risco e como evitá-las.

(4) Conteúdo

- Temático: Causas, sintomas e meios de transmissão dessas infecções, tratamentos possíveis, epidemias e regiões vulneráveis do mundo.
- Lingüístico:

Padrões de vocabulário relacionado a infecções.

O gênero *blog* e seus sub-gêneros.

(5) Seqüência didática:

- O professor apresenta o tema.
 - *Tarefa de suporte*: Alunos lêem o texto 'Update: Bird flu kills two more in Indonesia' e apresentam os pontos principais. Professor esclarece pontos relevantes e tira dúvidas.
 - *Tarefa de suporte*: Alunos fazem exercícios de aprendizado do vocabulário relacionado ao texto e ao tema em geral, como 'infected', 'bird flu' e 'HIV', por meio de concordâncias. A concordância lacunada de 'infected', mostrada acima, poderia ser empregada neste ponto.
 - Professor apresenta o gênero *blog*.
 - *Tarefa de suporte*: Alunos analisam o gênero *blog* e identificam os tipos de texto (sub-gêneros) comuns nesse meio.
 - Professor aponta algumas características lingüísticas do gênero *blog*.
 - *Tarefa de suporte*: Alunos trabalham com concordâncias em cima desses aspectos lingüísticos.
 - *Tarefa comunicativa*: Alunos escrevem um texto para um *blog* (real ou fictício).
 - Professor recolhe os textos e dá *feedback*.
 - Alunos e professor publicam os textos. Se for um *blog* real, os textos devem ser publicados; se for um *blog* fictício, os textos podem ser colados num mural ou na própria sala de aula.
- (6) Avaliação:
- Qualidade das discussões, das análises, dos textos, etc.

Desenvolvendo o material

Após a escolha do tema, nossa recomendação é que o professor parta para a escolha de um texto que seja relevante para a aula ou material que deseje desenvolver. A seguir, é preciso analisar o texto, começando com uma lista de palavras com 'clusters activated' do texto autêntico em questão. Essa opção faz com que a lista seja montada com 'clusters' em vez de palavras isoladas.

O texto que empregamos nesta análise foi retirado da Internet, do sítio: http://www.foodconsumer.org/777/8/Update_Bird_flu_kills_two_more_in_Indonesia.shtml. O trecho inicial do texto é o seguinte:

Update: Bird flu kills two more in Indonesia

Two more siblings in Indonesia died of bird flu, officials said Friday, citing the rest results were preliminary and needed to be confirmed by a laboratory accredited by the World Health Organization.

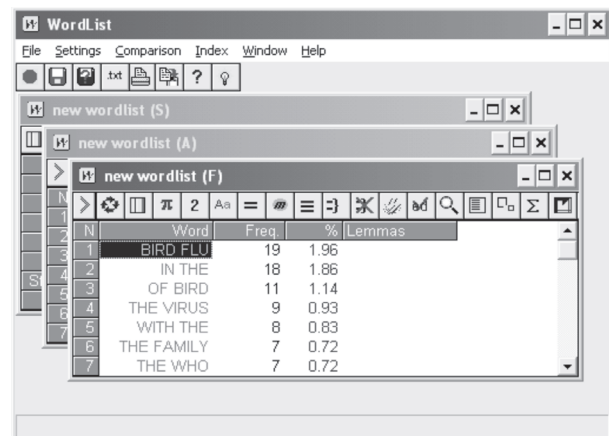
The report came as the WHO and officials from the Ministry of Health in Indonesia have struggled to search for the source of infection for a cluster of eight bird flu cases in a family in the village of Kubu Sembelang [...]

A recomendação que fazemos é de que sejam feitas duas listas, uma com *clusters* de duas palavras e outra com *clusters* de três palavras. As Figuras 1 e 2 mostram a tela do Wordlist com *clusters* de duas e três palavras, respectivamente.

Em seguida, é preciso observar as duas listas e encontrar *clusters* que possam ser trabalhados com os alunos, de acordo com os objetivos da unidade ou da aula, como por exemplo 'bird flu' e 'got infected with'. O primeiro parece interessante pois é um grupo nominal, uma estrutura que algumas vezes é problemática para alunos brasileiros. O segundo *cluster* possui de interessante o fato de ser um exemplo de voz passiva com 'get'. Além disso, ambos são frequentes dentro do texto, o que ajuda para que sejam internalizados.

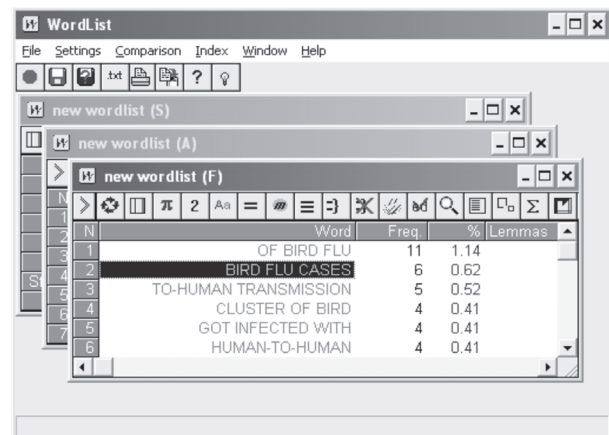
Para trabalhar esses itens com os alunos, o professor pode preparar concordâncias, por meio do Concord. Para tanto, basta iniciar o Concord e digitar em 'Search

Word' o *cluster* exatamente como aparece na lista do Wordlist, como mostra a Figura 3, com 'bird flu'.



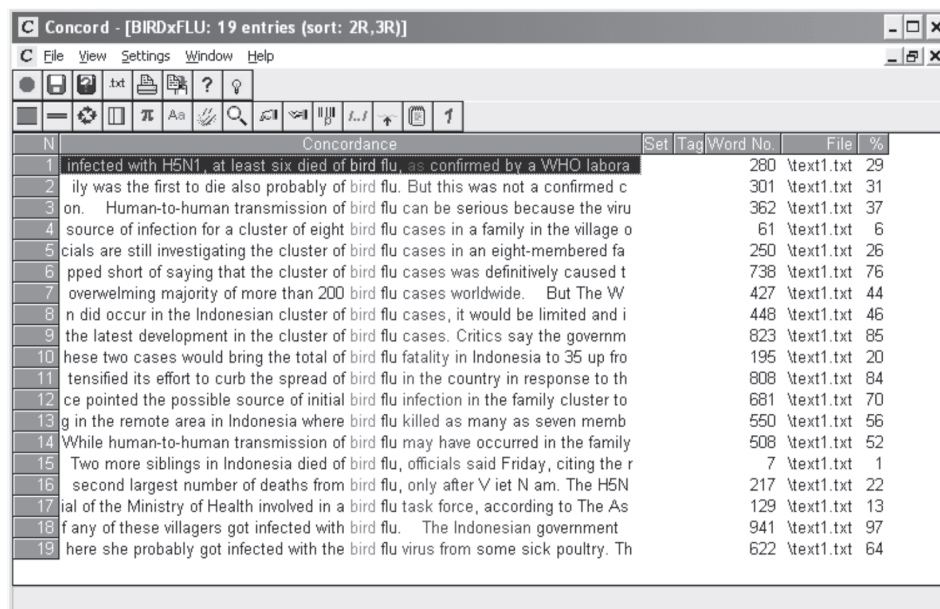
N	Word	Freq.	%	Lemmas
1	BIRD FLU	19	1.96	
2	IN THE	18	1.86	
3	OF BIRD	11	1.14	
4	THE VIRUS	9	0.93	
5	WITH THE	8	0.83	
6	THE FAMILY	7	0.72	
7	THE WHO	7	0.72	

Figura 1. Tela do Wordlist com *clusters* de duas palavras.



N	Word	Freq.	%	Lemmas
1	OF BIRD FLU	11	1.14	
2	BIRD FLU CASES	6	0.62	
3	TO-HUMAN TRANSMISSION	5	0.52	
4	CLUSTER OF BIRD	4	0.41	
5	GOT INFECTED WITH	4	0.41	
6	HUMAN-TO-HUMAN	4	0.41	

Figura 2. Tela do Wordlist com *clusters* de três palavras.



N	Concordance	Set	Tag	Word No.	File	%
1	infected with H5N1, at least six died of bird flu, as confirmed by a WHO labora...	280	text1.txt	29		
2	ily was the first to die also probably of bird flu. But this was not a confirmed c...	301	text1.txt	31		
3	on. Human-to-human transmission of bird flu can be serious because the viru...	362	text1.txt	37		
4	source of infection for a cluster of eight bird flu cases in a family in the village o...	61	text1.txt	6		
5	cials are still investigating the cluster of bird flu cases in an eight-membered fa...	250	text1.txt	26		
6	pped short of saying that the cluster of bird flu cases was definitively caused t...	738	text1.txt	76		
7	overwhelming majority of more than 200 bird flu cases worldwide. But The W...	427	text1.txt	44		
8	n did occur in the Indonesian cluster of bird flu cases, it would be limited and i...	448	text1.txt	46		
9	the latest development in the cluster of bird flu cases. Critics say the governm...	823	text1.txt	85		
10	hese two cases would bring the total of bird flu fatality in Indonesia to 35 up fro...	195	text1.txt	20		
11	tensified its effort to curb the spread of bird flu in the country in response to th...	808	text1.txt	84		
12	ce pointed the possible source of initial bird flu infection in the family cluster to...	681	text1.txt	70		
13	g in the remote area in Indonesia where bird flu killed as many as seven memb...	550	text1.txt	56		
14	While human-to-human transmission of bird flu may have occurred in the family...	508	text1.txt	52		
15	Two more siblings in Indonesia died of bird flu, officials said Friday, citing the...	7	text1.txt	1		
16	second largest number of deaths from bird flu, only after Viet N am. The H5N...	217	text1.txt	22		
17	ial of the Ministry of Health involved in a bird flu task force, according to The As...	129	text1.txt	13		
18	f any of these villagers got infected with bird flu. The Indonesian government...	941	text1.txt	97		
19	here she probably got infected with the bird flu virus from some sick poultry. Th...	622	text1.txt	64		

Figura 3. Exemplo de concordâncias para 'bird flu' preparadas por meio do Concord.

As linhas da concordância estão ordenadas pela primeira à direita. Os padrões que emergem são:

- Colocação: Bird flu cases
- Coligação: Bird flu + Substantivo (fatality, infection, task force).

Fazendo *resort* pela primeira palavra à esquerda, obtemos a tela ilustrada na Figura 4.

O padrão mais saliente é ‘of bird flu’, que em si não é interessante. Mas a partir dele pode-se expandir o alcance dos padrões mais para a esquerda, conseguindo os seguintes padrões:

- Cluster of bird flu cases
- Spread of bird flu
- Died of bird flu
- Got infected with bird flu / the bird flu virus.

Esses padrões são de central importância para entender o texto em questão; além disso, ajudam o aluno a entender muitos outros textos sobre o assunto.

Depois de retirar algumas fraseologias importantes do texto, é preciso passar para a análise de um corpus. A idéia é encontrar no corpus mais instâncias de padrões do texto (1) para reforçar a exposição dos alunos a esses padrões ou (2) para expandir o leque de usos dos padrões do texto.

Em ambos os casos, há duas opções, dependendo do corpus de referência:

Se o professor tiver acesso a um corpus de características (gênero e assunto) semelhantes ao texto da aula, pode fazer buscas de ‘clusters’ do texto tais como eles aparecem. Por exemplo, se o professor tiver um corpus de textos sobre gripe aviária, pode fazer outras buscas por ‘bird flu’ no corpus, para levantar outros padrões.

Se o professor não possuir um corpus específico, mas sim um corpus geral de inglês, provavelmente não

encontrará muitos outros exemplos de padrões do texto em questão, caso se trate de um assunto contemporâneo ou altamente especializado. Por exemplo, no BNC há apenas uma ocorrência de ‘bird flu’ (‘National Institutes of Health have overcome this dilemma by growing a *bird flu* virus which does not cause human disease’). Nesse caso, o professor deve buscar outras formas mais gerais referentes ao vocabulário dos padrões, como ‘flu’ apenas ou ‘infected’.

Para ilustrar, vejamos a palavra ‘infected’, que faz parte de um padrão do texto (‘got infected with’) no British National Corpus. Há 1156 ocorrências dessa palavra no corpus. Ao fazer a concordância, é importante selecionar ‘Activated’ em ‘Ignore Tags’, na aba ‘Tags’ de ‘Settings’. Isso precisa ser feito pois o BNC é etiquetado morfossintaticamente e, para o uso que vislumbramos em sala de aula, as etiquetas são dispensáveis; caso apareçam na tela, tornarão a visualização dos padrões mais difícil. A concordância resultante aparece na Figura 5.

Como se percebe, a concordância está com muitos espaços entre as palavras; estes são os espaços onde estão as etiquetas (‘tags’) que foram ignoradas. Para melhorar

N	Concordance	Set	Tag	Word No.	File
1	touch with around actually infected individuals of with			2,679	\ala0\al0c
2	000 people actually infected)			529	\ala0\al0c
3	s . Then all infected timber is cut			1,632	\ala7\ala7c
4	to their allegedly infected chicken coop			2,608	\ala5\ala5t
5	which is allegedly infected with Salmonella			1,651	\ala5\ala5c
6	cities are already infected . The ef			8,503	\ala0\ala0t
7	Adults are also infected			3,694	\ala0\ala0n

Figura 5. Concordância resultante em pesquisa do vocabulário *infected*.

N	Concordance	Set	Tag	Word No.	File	%
1	overwhelming majority of more than 200 bird flu cases worldwide. But The W			427	\text1.txt	44
2	ial of the Ministry of Health involved in a bird flu task force, according to The As			129	\text1.txt	13
3	source of infection for a cluster of eight bird flu cases in a family in the village o			61	\text1.txt	6
4	second largest number of deaths from bird flu, only after Viet Nam. The H5N			217	\text1.txt	22
5	ce pointed the possible source of initial bird flu infection in the family cluster to			681	\text1.txt	70
6	on. Human-to-human transmission of bird flu can be serious because the viru			362	\text1.txt	37
7	ciala are still investigating the cluster of bird flu cases in an eight-membered fa			250	\text1.txt	26
8	pped short of saying that the cluster of bird flu cases was definitively caused t			738	\text1.txt	76
9	n did occur in the Indonesian cluster of bird flu cases, it would be limited and i			448	\text1.txt	46
10	the latest development in the cluster of bird flu cases. Critics say the governm			823	\text1.txt	85
11	hese two cases would bring the total of bird flu fatality in Indonesia to 35 up fro			195	\text1.txt	20
12	tensified its effort to curb the spread of bird flu in the country in response to th			808	\text1.txt	84
13	While human-to-human transmission of bird flu may have occurred in the family			508	\text1.txt	52
14	infected with H5N1, at least six died of bird flu, as confirmed by a WHO labora			280	\text1.txt	29
15	Two more siblings in Indonesia died of bird flu, officials said Friday, citing the r			7	\text1.txt	1
16	ily was the first to die also probably of bird flu. But this was not a confirmed c			301	\text1.txt	31
17	here she probably got infected with the bird flu virus from some sick poultry. Th			622	\text1.txt	64
18	g in the remote area in Indonesia where bird flu killed as many as seven memb			550	\text1.txt	56
19	f any of these villagers got infected with bird flu. The Indonesian government			941	\text1.txt	97

Figura 4. Tela do Concord após o uso do *resort* pela primeira palavra à esquerda (bird).

o 'layout' da tela, é preciso escolher 'View', 'Spaces Cut'. Isso uniformiza os espaços, conforme mostra a Figura 6.

Com base nessa tela, o professor pode buscar mais alguns padrões que possa ensinar. Um que desponta na tela é 'HIV infected', cujos padrões incluem:

- Are already / to be HIV-infected
- Becoming HIV infected
- HIV-infected staff
- People estimated to be HIV infected²

Conjugando a análise do texto com a do corpus, temos uma visão mais aprofundada da padronização de 'infected':

- Get infected with
- Be HIV-infected
- Become HIV-infected
- HIV-infected [substantivo]

Essa padronização nos mostra que 'infected':

- pode funcionar como verbo (get infected) ou adjetivo (HIV-infected staff);
- pode formar a passiva com 'get', reforçando sua prosódia semântica, pois 'get' tende a ser usado na passiva para indicar sentidos negativos. Além disso, 'get passives' podem indicar uma parcela de culpa do indivíduo na ação ('get drunk' pressupõe a opção da pessoa em se embriagar);
- pode associar-se a substantivos para especificar o tipo de infecção (HIV-infected).

Resumindo nosso percurso exploratório:

- descobrimos que 'bird flu' era um dos principais 'clusters' do texto;
- ainda no texto, a partir da padronização de 'bird flu', chegamos a 'got infected';

- já no corpus, partindo de 'got infected', chegamos a 'HIV-infected' e sua padronização.

Esse tipo de análise exploratória é uma das grandes vantagens de trabalhar com corpus na sala de aula, pois revela aspectos da língua que os dicionários e 'usage guides' não deixam tão claro e com os quais as gramáticas sequer lidam. Além disso, é um percurso de descoberta 'movido a dados', pois foram os dados da primeira descoberta (bird flu) que levaram às descobertas seguintes. Nesse sentido, é um processo pessoal. Cada aluno e cada professor pode obter 'mapas' dessa jornada exploratória diferentes, pois não há necessariamente a necessidade de chegar a 'got infected' a partir de 'bird flu'. Uma outra pessoa poderia ter visto outro padrão junto a 'bird flu' que lhe chamasse a atenção. O importante é que os alunos e o professor descubram padrões atestados nos dados que sejam relevantes para a sua aula, para o nível e para os objetivos do curso.

Para que isso aconteça, é fundamental ganhar experiência na análise de concordâncias, 'treinando os olhos' para perceber padrões que sejam 'interessantes'. Muitas vezes, conforme mostra minha prática pessoal de sala de aula, os alunos se detêm em padrões pouco relevantes. Por exemplo, alunos de nível avançado que estivessem analisando as concordâncias acima poderiam ter notado a presença do padrão 'had/has/have infected'. Esse padrão, embora existente nos dados, nada mais é do que o 'present perfect', que tais alunos supostamente já conheciam. Ou seja, é preciso deixar claro para o aluno que 'repetição não é padrão', para que eles passem a procurar padrões novos por si mesmos.

Outra tendência que professores e alunos 'novatos' em análise de corpus exibem é se deter a padrões

The screenshot shows a window titled 'Concord - [INFECTED: 123 entries (sort: 1L)]'. The window contains a table with columns: N, Concordance, Set, Tag, Word No., File, and %. The table lists various instances of the word 'infected' in different contexts, such as 'and you may die. Your gut can get infected, so you can't absorb the good transfusions...' and 'The chance of getting infected from a pint of blood is less than...'. The table also shows the frequency of each instance across different files and percentages.

N	Concordance	Set	Tag	Word No.	File	%
61	and you may die. Your gut can get infected, so you can't absorb the good transfusions. The chance of getting infected from a pint of blood is less than...			1,972	\a\0\01	22
62	d transfusions. The chance of getting infected from a pint of blood is less than...			577	\a\0\01	7
63	entre &equo. &equo Trade &equo had infected even gentry pursuits, though...			19,316	\a\6\6y	57
64	r tell him he has AIDS. Tom has infected at least 6 people — the...			905	\a\0\01	11
65	typhimurium, which is alleged to have infected the nun's flock. Professor Ja...			2,023	\a\5\50	10
66	kyns communicated enthusiasm. He infected his hearers with the sense th...			14,183	\a\6\68	26
67	in the home not only benefits the HIV infected individual but the whole family...			5,778	\a\0\00	74
68	number of people estimated to be HIV infected is 1.3 million, representing ov...			3,500	\a\0\02	86
69	s. Tragically, more are becoming HIV infected, and many of those we care f...			2,922	\a\0\02	73
70	ources. Seven million are already HIV infected in sub-Saharan Africa, incre...			3,445	\a\0\02	85
71	y with hundreds of young children HIV infected through hazardous medical or...			1,387	\a\0\02	36
72	e African towns and cities already HIV infected, it is hard to know how to be...			1,252	\a\0\02	32
73	a right-wing pressure group to ban HIV-infected staff from working with food h...			1,331	\a\0\0c	4
74	gal and moral duty in dealing with HIV-infected staff, the topic has receded fr...			1,474	\a\0\0c	5
75	catch it by using floppy discs holding infected programmes, or from progra...			10,760	\a\3\30	60
76	tion should be made on floppy discs. Infected programmes activate the virus...			10,715	\a\3\30	60
77	people are being murdered, terrorised, infected with diseases and driven from...			8,927	\a\7\7a	40

Figura 6. Espaço uniformizado após o uso de space cut.

² Pode-se notar, de passagem, a falta de uniformização ortográfica em relação ao uso do hífen entre 'HIV' e 'infected'.

gramaticais. Por exemplo, os padrões da concordância de ‘infected’ do corpus poderiam ser expressos por um padrão como ‘verbo + infected’, onde verbo poderia ser substituído por ‘be/get/have’. Esse padrão, novamente, é pouco informativo. Ele simplesmente retrata um aspecto gramatical, qual seja, o grupo verbal formado por auxiliar seguido de verbo principal. O que leva alunos e professores ‘novatos’ a pinçarem esse tipo de padrão é certamente a influência de anos de ensino de gramática, o que por sua vez os levam a ter familiaridade em analisar a língua segundo a gramática estrutural. Embora a análise de corpus possa ser usada para estudo de pontos gramaticais, na verdade ela pode e deve levar o aluno a uma apreciação diferente da língua, baseada na léxico-gramática. Isso, conforme dissemos, leva tempo, pois envolve prática na identificação de padrões em concordâncias. E prática, como se sabe, somente se adquire praticando, por isso corresponde a um longo processo que necessita da intervenção do professor, principalmente no início. Felizmente, o WordSmith Tools é bastante rápido e simples de usar, o que evita gastar tempo demais na mecânica da obtenção das concordâncias, podendo o professor e os alunos direcionarem suas energias para a análise dos dados e para a descoberta dos padrões.

O emprego de concordâncias

As concordâncias têm sido usadas em sala de aula de ensino de línguas há muito tempo. Tribble e Jones (1990) e Johns e King (1991) já mostravam como incorporar atividades com concordâncias ao ensino de inglês há mais de 15 anos. Fox (1998) também ilustrou o uso desse instrumento no ensino de vocabulário. A experiência de empregar concordâncias em sala de aula se mostrou tão frutífera que Willis (1998) até mesmo apresentou uma maneira de usá-las em locais onde não existe computador.

Há em essência dois tipos de concordância que podem ser usadas em sala de aula: completas e lacunadas. As concordâncias completas possuem a palavra de busca visível no centro, ao passo que as lacunadas (‘gapped’) têm um espaço em branco onde originalmente aparecia a palavra de busca. Para criar uma concordância lacunada no WordSmith Tools, basta fazer uma concordância e teclar a barra de espaço (enquanto a concordância estiver sendo exibida na tela). Para retornar a concordância ao formato original, basta teclar a barra de espaço novamente.

Abaixo aparece uma concordância lacunada da palavra ‘infected’ (a largura das linhas foi truncada para caber nesta página; para fins pedagógicos, recomendamos uma largura maior, de no mínimo 100 caracteres):

1 person, or if their _____ blood gets into another
2 when there are a lot of _____ spots on the face and
3 and should a piece of the _____ tissue break off, an
4 anal intercourse, because _____ material from the vagina
5 offence. The _____ man is a 24-year-old

6 smallpox, where a single _____ individual can cause an
7 ILLUSTRATION Bathe the _____ area in a salt solution:
8 seedbed regard that as _____ land, and do not use it
9 changed, they’ve become _____ with this great
10 by one of the women he _____ with the HIV virus.

O tipo de trabalho que cada tipo de concordância proporciona é diferente. Com uma concordância completa, os alunos geralmente trabalham para (a) tentar descobrir os padrões de uso da palavra, ou (b) tentar identificar as diferenças de sentido de uma mesma palavra, baseando-se na existência de padrões diferentes para cada sentido. Numa concordância lacunada, por outro lado, o trabalho essencialmente se resume a identificar a palavra excluída, com base nas palavras próximas à lacuna e nos padrões que circundam a lacuna. Na concordância acima, por exemplo, espera-se que os alunos descubram a palavra ‘infected’ com base no conhecimento temático já ativado na unidade (infecções) e nos padrões:

‘infected [parte do corpo]’: linhas 1 a 4;
‘infected [pessoa]’: linhas 5 e 6;
‘infected [porção de terreno]’: linhas 7 e 8;
‘infected with’ e ‘HIV virus’: linhas 9 e 10.

Em relação à composição das concordâncias, há um debate entre aqueles que defendem seu emprego sem seleção intencional de linhas, podendo no máximo ser aplicado um método de seleção aleatória de linhas (no caso de palavras com muitas ocorrências) e aqueles que reivindicam que sejam apresentadas ao aluno apenas linhas com vocabulário acessível ou com padrões mais visíveis (nesse caso, o preparador de material deve manter na concordância apenas as linhas que julgar mais adequadas). Pessoalmente, acredito que ambas as opções sejam válidas, dependendo da situação de ensino. Para alunos menos avançados e para aqueles com pouca familiaridade com esse instrumento, parece-me mais adequado selecionar as linhas da concordância. Já para alunos mais avançados ou com bastante exposição a concordâncias, a seleção intencional de linhas pode ser desnecessária. De qualquer modo, como regra geral creio ser desejável evitar concordâncias com mais de 30 linhas ou uma página de papel impresso de extensão, devido ao esforço cognitivo que sua análise demanda.

Comentários finais

O programa WordSmith Tools é versátil e poderoso. Neste trabalho, mostramos como ele pode ser empregado na preparação de material didático para ensino de língua estrangeira. Apresentamos uma proposta de material baseado na Aprendizagem baseada em Tarefas (Task-Based Learning) em que figura um exercício com concordância produzido com o WordSmith Tools. Esperamos com isso ter contribuído para divulgar o potencial do programa na área de ensino de línguas estrangeiras.

Referências

- ALONSO, M.C.G. 2006. *Corpus lingüístico e a aquisição de falsos cognatos em espanhol como língua estrangeira*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- AMARANTE, R.P. 2005. *Uma contribuição da lingüística de corpus para a fonologia: Um estudo de colocações e aspectos segmentais das vogais da língua inglesa*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- BARBOSA, M.E.D.C. 2004. *Material didático para o ensino de inglês instrumental online: uma abordagem experiencial baseada em corpus, gênero e tarefa*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- BÉRTOLI DUTRA, P. 2002. *Explorando a lingüística de corpus e letras de música na produção de atividades pedagógicas*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- BHATIA, V.K. 1993. *Analysing Genre: Language Use in Professional Settings*. London, Longman.
- BIBER, D. e CONRAD, S. 1999. Lexical bundles in conversation and academic prose. In: H. HASSELGARD e S. OKSEFJELL (orgs.), *Out of Corpora - Studies in Honour of Stig Johansson*. Amsterdam/Atlanta, GA, p. 181-190.
- CONDI DE SOUZA, R. 2005. *Dois corpora, uma tarefa: O percurso de coleta, análise e utilização de corpora eletrônicos na elaboração de uma tarefa para ensino de inglês como língua estrangeira*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- EGGINS, S.; MARTIN, J.R. e VAN DIJK, T.A. 1997. Genres and registers of discourse. *Discourse as Structure and Process*. London, p. 230-256.
- ELLIS, R. 2003. *Task-based language learning and teaching*. Oxford applied linguistics. Oxford, Oxford University Press, 387 p.
- ESTAIRE, S. e ZANON, J. 1994. *Planning classwork: A task-based approach*. Oxford, Heinemann.
- FERRARI DE OLIVEIRA, J. 2004. *ESP, Lingüística de Corpus e Sócio-Interacionismo na elaboração de uma unidade de um material didático para comércio exterior*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- FOX, G. 1998. Using corpus data in the classroom. In: B. TOMLINSON (org.), *Materials development in language teaching*. Cambridge, p. 25-43.
- JACOBI, C.C.B.D. 2001. *Lingüística de Corpus e ensino de espanhol a brasileiros: Descrição de padrões e preparação de atividades didáticas (decir/hablar ; mismo; mientras /en cuanto/ aunque)*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- JOHNS, T. 1991. From printout to handout: Grammar and vocabulary teaching in the context of Data-Driven Learning. *ELR Journal*, 4:27-46.
- JOHNS, T. e KING, P. 1991. *Classroom Concordancing — ELR Journal 4*. Birmingham, Centre for English Language Studies/University of Birmingham.
- KOHONEN, V. 2001. *Experiential learning in foreign language education*. Applied linguistics and language study. Harlow/New York, Longman, xvii, 182 p.
- LEWIS, M. 1997. *Implementing the Lexical Approach - Putting Theory into Practice*. Hove, LTP.
- MARTIN, J.R. 1997. Analysing genre. Functional parameters. In: F. CHRISTIE e J.R. MARTIN (orgs.), *Genre and Institutions. Social processes in the Workplace and School*. London, p. 3-39.
- MEURER, J.L.; BONINI, A. e MOTTA-ROTH, D. (orgs.). 2005. *Gêneros - Teorias, Métodos, Debates*. São Paulo, Parábola.
- MOTTA-ROTH, D. 2006. O ensino de produção textual com base em atividades sociais e gêneros textuais. *Linguagem em (Dis)curso*, 6(3).
- NUNAN, D. 2004. *Task-based language teaching*. Cambridge language teaching library. Cambridge/New York, Cambridge University Press, xv, 222 p.
- PRABHU, N.S., 1987. *Second language pedagogy*. Oxford English. Oxford/New York, Oxford University Press, 153 p.
- SCOTT, M. 1997. *WordSmith Tools. Version 3*. Oxford, Oxford University Press.
- SINCLAIR, J. 1991. *Corpus, Concordance, Collocation*. Oxford, Oxford University Press.
- SKEHAN, P. 1996. Second language acquisition research and task-based instruction. In: J. WILLIS e D. WILLIS (orgs.), *Challenge and Change in Language Teaching*. Oxford, p. 17-30.
- SWALES, J.M. 1990. *Genre Analysis - English in Academic and Research Settings*. Cambridge, Cambridge University Press.
- TRIBBLE, C. e JONES, G. 1990. *Concordances in the Classroom - A Resource Book for Teachers*. London, Longman.
- VICENTINI, G.P.M. 2006. *A lingüística de corpus e o seriado Friends como base para o ensino de chunks em sala de aula de língua inglesa*. São Paulo, SP. Dissertação de Mestrado. PUCSP.
- WILLIS, J. 1996. *A framework for task-based learning*. Longman handbooks for language teachers. Harlow/Essex, Longman, 183 p.
- WILLIS, J. 1998. Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In: B. TOMLINSON (org.), *Materials development in language teaching*. Cambridge, p. 44-66.

Submetido em: 11/2006

Aceito em: 11/2006

Tony Berber Sardinha
PUCSP