

## USO DE CORPORA NA FORMAÇÃO DE TRADUTORES

(Using Corpora in Translator Training)

Antonio P. BERBER SARDINHA  
(PUC-SP)

**ABSTRACT:** *This paper tackles the issue of using corpora in translator training, focussing more specifically on the question of awareness raising. The paper presents a discussion on the role of corpora in translation, their applicability in professional development, and their importance in leading to a better understanding of how language is constituted. Two example analyses are offered and detailed, so that they are applicable to contexts in which computational resources are scarce. The analyses center around the linguistic choices in a translated newspaper text and in the Brazilian version of a slogan from an American advertising campaign. It is suggested that these activities may be carried out with translation students, in such a way that they enable students, while they explore electronic corpora, to become aware of both the complexity and the specificity of the linguistic choices involved in the process of translation.*

**KEY-WORDS:** *Corpus Linguistics; Awareness raising; Electronic corpora; Translator training.*

**RESUMO:** *O presente trabalho aborda a questão do uso de corpora na formação de tradutores, enfocando mais especificamente a questão da conscientização. O trabalho apresenta uma discussão sobre o papel de corpora na tradução, sua aplicabilidade na formação profissional, e sua importância para o melhor conhecimento da constituição da linguagem. São oferecidos dois exemplos de análises, detalhadas a fim de serem aplicáveis em contextos em que haja poucos recursos de infraestrutura. As análises centram-se em pesquisas sobre as escolhas lingüísticas de um texto jornalístico traduzido para o português, e da versão brasileira de um slogan de uma campanha publicitária americana. É sugerido que essas atividades possam ser desenvolvidas com alunos de tradução, de tal modo que elas forneçam condições para que os alunos, ao explorarem corpora eletrônicos, possam se conscientizar da complexidade e da especificidade das escolhas lingüísticas envolvidas no processo tradutório.*

**PALAVRAS-CHAVE:** *Lingüística de Corpus; Conscientização; Corpora eletrônicos; Formação de tradutores.*

## 1. Introdução

O uso de tecnologia computacional na tradução vem se tornando mais freqüente. Há programas de tradução automática, bancos terminológicos, dicionários eletrônicos, bibliotecas informatizadas, além de diversos outros recursos (Austermuehl, 2001; Bowker, 2002). Nesse cenário, os corpora eletrônicos ganham cada vez mais destaque, como recursos de grande utilidade para o tradutor, seja na sua prática profissional ou como pesquisador. Corpora eletrônicos são coletâneas de textos escritos ou transcrições de fala reunidos em formato de arquivo legível por computador.

O papel dos corpora eletrônicos na pesquisa em tradução tem sido enfatizado por muitos estudiosos, dentro e fora da área dos Estudos Tradutológicos. Entre os pesquisadores da tradução, Mona Baker, como a maior divulgadora do uso de corpora no entendimento do produto e dos processos envolvidos em tradução, vê o corpus eletrônico como um instrumento revolucionário, que permite enxergar aspectos da linguagem do texto traduzido, em particular, de modo muito mais rico e abrangente do que por outros meios (Baker, 1993, 1995, 1996). Ela é vista como 'mother of corpus-based descriptive translation studies' (Laviosa, 2002:18). Seu trabalho teve papel decisivo na implantação de um programa de pesquisa fundado na exploração de corpora que deu vazão a um novo paradigma no âmbito dos estudos da tradução (Laviosa, 2002:21-22).

Vários autores destacam o crescimento da pesquisa em tradução baseada em corpora, em boa parte devido à influência de Mona Baker. Bowker (Bowker, 2000:19), por exemplo, afirma que 'the use of corpora in the discipline of translation is enjoying increased popularity'. No entanto, segundo Berber Sardinha (no prelo-a), quando se olha de fato a produção e publicação de trabalhos nessa área, o que se observa é crescimento aquém do que os comentários dos especialistas nos permitiriam supor. Dentre as possíveis causas desse cenário estão a falta de recursos (corpora e software), a dificuldade de acesso à infraestrutura de tecnologia (computadores e redes) e, talvez principalmente, a capacitação pessoal. O conhecimento de informática e os recursos necessários para a pesquisa com corpora eletrônicos parecem estar distantes da maioria dos tradutores formados e dos aprendizes de tradução.

Uma das maneiras pelas quais talvez seja possível mudar esse cenário é por meio do uso de corpora na formação de novos tradutores e pesquisa-

dores. Com esse trabalho de base, novas gerações desses profissionais e estudiosos estarão familiarizados com a tecnologia e o aparato teórico do uso de corpora na tradução. Para efetivar essa proposta, contudo, é preciso ser realista: a demanda de recursos tecnológicos deve ser compatível com o que está disponível dentro das circunstâncias oferecidas pelas universidades brasileiras. Mesmo levando em conta que a situação que conhecemos possa ser contingencial, uma proposta para ser viável nesse cenário deve, portanto, fazer exigências mínimas de infraestrutura, exigindo apenas corpora e ferramentas básicas. Isso parece razoável até porque se espera que os aprendizes possam acessar esses recursos fora do local de ensino também, particularmente em suas casas. Obviamente, aqueles centros que dispõem de mais recursos não devem se limitar a esse mínimo, podendo (e devendo) oferecer aos seus formandos uma gama mais ampla possível de corpora e programas de computador.

Esse trabalho tenta fazer uma contribuição justamente nesse sentido: sugerir maneiras de como é possível fazer pesquisa, num contexto de formação de novos tradutores, com corpora eletrônicos usando recursos gratuitos já existentes na rede, apenas com uma conexão e navegador de Internet. O ponto defendido aqui, nesse sentido, é o de que o conhecimento gerado e adquirido por meio de pesquisa, tanto do ponto de vista do produto quanto do processo relativo ao manuseio das ferramentas, é enriquecedor e fundamental para a formação do tradutor.

O uso de corpora na formação de profissionais e pesquisadores de tradução vem recebendo cada vez mais atenção. Segundo Laviosa (2002:22):

'More recently the corpus-based approach has been developed and applied in translator training where experimental and empirical research go hand in hand with innovative teaching programmes which directly involve students in designing, creating, exploring and exploiting corpora for improving the quality of their translations.'

Bowker (1999), por exemplo, relata o uso de corpora na conscientização de futuros tradutores no tocante à apreciação das especificidades da linguagem técnica. Segundo ela, os alunos devem ter em mente que textos de Línguas para Fins Específicos impõem vários desafios aos aprendizes de tradução, entre eles o do conhecimento específico de áreas tecnológicas. Comentando o comportamento típico de alunos de tradução relativo a diferenças entre Línguas para Fins Gerais (Languages for General Purposes,

LGP) e Línguas para Fins Específicos (LSP, Languages for Special Purposes), ela afirma:

‘Because initially the students are neither aware of nor interested in the distinction between LGP and LSP, they frequently equate the acceptability of a given translation with its grammatical and/or semantic correctness. In other words, they are often unaware of the importance of factors such as context, text type, register and idiom, and hence they may rely incorrectly on their LGP instincts in LSP situations. This type of behaviour inevitably leads to the production of inappropriate translations, and when the trainer tries to point out such errors, cries of ‘But I found *that* term in the dictionary!’ or ‘But you *can* say that in English!’ can be heard echoing around the translation classroom.’ (Bowker, 1999:161; ênfases no original)

O foco do trabalho de Bowker (1999) é na questão da conscientização, ou seja, em como fornecer condições, por meio da exploração de corpora, para que os alunos tomem consciência da diferença entre seu conhecimento sobre a língua em geral e o conhecimento específico exigido para a tradução de textos especializados. A definição geral de conscientização usada nesse trabalho a vê como ‘a person’s sensitivity to and conscious awareness of the nature of language in human life’ (Donmal, 1985, apud. Thornbury, 1997:x), com ‘human life’ entendido aí como englobando a prática da tradução. Ou, como coloca Thornbury (1997:x, ênfase no original), ‘explicit *knowledge* about language’. Uma maneira de propiciar meios que permitam ao indivíduo construir esse conhecimento sobre a língua, desenvolvendo sensibilidade e consciência acerca da natureza de aspectos da língua envolvidos na tradução, deve ser através do caminho da descoberta:

‘working out something for oneself pays greater dividends in terms of memory and understanding than simply having it explained. In other words, an inductive – rather than a deductive – approach...’ (Thornbury, 1997:xi)

Uma abordagem dedutiva, por outro lado, consistiria no professor de tradução selecionar elementos e explicá-los, geralmente por meio de regras, ou, como coloca Thornbury (1997:xi), ‘get hold of an up-to-date grammar and read it from cover to cover’. Entretanto, tal direcionamento não seria eficaz nem em relação à memória, nem em relação à língua. A pesquisa com corpora eletrônicos deixa claro que muito da informação que as gramáticas dizem a respeito do funcionamento das línguas é distorcido,

incompleto ou mesmo mal-concebido, quando confrontados com a evidência advinda de corpora (Sinclair, 2001). Sem falar que a análise de corpora permite descobrir aspectos novos da linguagem, que nem foram contemplados em gramáticas ou dicionários (Hunston, 2002; Partington, 1998).

Tendo isso em mente, o apelo de Bowker (2000), na verdade, é mais amplo: seu trabalho deixa claro que (1) a tomada de consciência acerca das especificidades da língua é um aspecto importantíssimo da capacitação do indivíduo como tradutor, e (2) esse processo de conscientização pode e deve ser feito por meio da exploração de corpora eletrônicos. Partindo desses pressupostos, o presente trabalho visa a enfocar a questão da conscientização, mostrando como é possível desenvolver atividades para cursos de tradução que contemplem o manuseio de corpora eletrônicos fazendo, porém, poucos exigências relativas a recursos de infraestrutura de pesquisa, conforme já exposto antes. Vale lembrar que os contextos referidos por Laviosa (2002) e Bowker (2000) nas pesquisas citadas acima compreendem acesso a recursos variados, como corpora de textos específicos, corpora paralelos ou até mesmo alinhados e concordanceadores bilíngües. Para um trabalho de conscientização com alunos iniciantes na Lingüística de Corpus e nos estudos tradutológicos com corpora, contudo, tais recursos não são estritamente indispensáveis.

As pesquisas descritas abaixo, portanto, devem ser entendidas como fundamentalmente aplicáveis a alunos iniciantes, com conhecimentos básicos de informática, tais como navegar na Internet, e pouco ou nenhum conhecimento de Lingüística de Corpus. O intuito ao apresentá-las é o de mostrar algumas possibilidades de tópicos de pesquisa que podem ser explorados dentro de um cenário onde há limitação de recursos disponíveis. Embora a meta principal a ser atingida com elas seja a de propiciar uma maior conscientização sobre as especificidades das escolhas lingüísticas no texto traduzido e no texto-fonte, outro aspecto importante que elas suscitam é o de fazer com que os alunos 'ponham a mão na massa'. 'Mexer com os dados' é um elemento-chave da Lingüística de Corpus e, por conseguinte, da pesquisa em tradução com corpora eletrônicos. A discussão das questões teóricas e conceituais envolvidas são importantes, mas é impossível engajar-se nessa abordagem sem de fato manusear os dados, incluindo aí muitas ações que compõe a rotina do analista de corpus, tais como confrontar-se com evidências, descartar ocorrências indesejadas, notar padrões recorrentes, formular hipóteses, sistematizar e avaliar os achados, etc.

A estratégia básica das pesquisas mostradas abaixo está centrada no uso de corpora monolíngües (abundantes na Internet) para avaliar as escolhas feitas em um texto traduzido. O uso desse tipo de corpus, no lugar de corpora paralelos, deve-se ao fato de esses últimos não serem disponíveis com facilidade na Internet. É possível montar um corpus paralelo próprio, com textos da Internet (Berber Sardinha, 2001), mas isso não é necessário para os propósitos de conscientizar os alunos da importância do uso de corpora no entendimento da língua e do processo tradutório.

O que torna possível investigar a tradução com corpora monolíngües é o fato de que as investigações apresentadas aqui centrarem-se na análise de pares de textos compostos por um original (em inglês) e sua tradução (para o português). Assim, embora os corpora sejam monolíngües, o foco continua sendo o texto traduzido, mais especificamente, algumas escolhas feitas pelo tradutor presentes no texto-alvo tendo em vista o texto-fonte. Há portanto, quatro elementos envolvidos: (1) um texto-fonte, (2) sua tradução, (3) um corpus monolíngüe na língua do texto-fonte, e (4) outro corpus monolíngüe na língua para o qual o texto foi traduzido. Os elementos 1 e 2 são os *textos focais* da pesquisa, podendo ser, em essência qualquer par de textos que se julgue de interesse investigar, tendo em vista os propósitos pedagógicos da atividade. A análise é de cunho contrastivo, comparando as escolhas feitas nos textos focais com aquelas encontradas nos corpora monolíngües. Esse trabalho contrastivo possui um lugar nos estudos descritivos da tradução, ao permitir, segundo Hoey e Houghton (1998:49), 'with equal facility explanations of past translating decisions and guidance to prospective ones'.

As análises buscam saber até que ponto as escolhas encontradas no texto traduzido são adequadas, tendo em vista a padronização tanto da língua-fonte quanto da língua-alvo. Padronização é entendida aqui como o conjunto de colocações, coligações, preferências e prosódias semânticas (Hunston, 2002) a partir das quais a léxico-gramática se constitui; em outras palavras, é a materialização do princípio idiomático (Hunston, 2002), visível nas evidências cumulativas e recorrentes fornecidas por corpora eletrônicos.

Os corpora monolíngües servem como fonte indispensável para conhecer a padronização das duas línguas envolvidas na tradução. A

conscientização, espera-se, se dará justamente no confronto entre (1) as escolhas feitas pelo autor do texto-fonte, (2) as escolhas feitas pelo tradutor, (3) como o aluno enxerga essas escolhas do ponto de vista da adequação ao conteúdo, contexto, estilo, registro, etc., (4) o que o aluno intui sobre as línguas A e B e (5) como as línguas A e B se revelam a partir das evidências dos corpora, segundo a tipicidade, probabilidade, frequência e padronização das escolhas lingüísticas. É esperado, no final, que os alunos percebam o grau de especificidade que algumas escolhas implicam, os vários sentidos que existem para uma mesma expressão, a adequação desses sentidos no contexto em que se insere o texto-fonte e o texto-alvo, como a intuição costuma ser falha (insuficiente ou enganosa) e como corpora eletrônicos podem oferecer uma quantidade ampla e rica de informação vital para o tradutor, para melhor ajudá-lo na consecução dos trabalhos. É preciso enfatizar que o propósito não é encontrar erros de tradução para criticar o trabalho deste ou daquele tradutor, ou de questionar a qualidade de editoras ou veículos específicos.

## 2. Corpora usados na pesquisa

Os corpora usados nas análises a seguir estão todos disponíveis na Internet, gratuitamente:

Corpus	Língua	Extensão	Composição
Corpus NILC	Português brasileiro	26,7 milhões de palavras	Linguagem escrita apenas: textos jornalísticos (principalmente), didáticos, cartas
Corpus COBUILD online	Inglês britânico e americano	45 milhões de palavras	Linguagem escrita e falada: textos jornalísticos, revistas, livros, conversação, rádio

O corpus NILC é acessado por meio do portal do projeto Processamento Computacional do Português, em <http://www.portugues.mct.pt>, e o corpus COBUILD está disponível no site <http://titania.cobuild.collins.co.uk/form.html>.

Além desses, outro tipo de corpus usado aqui é o conjunto de textos que compõe a World Wide Web. A exploração do conteúdo da WWW como um corpus imenso, multilíngüe, extremamente valioso e barato está começando a ser reconhecida e valorizada (Kilgarriff e Grefenstette, in

press). Um problema em considerar a Web como corpus é que seu conteúdo e extensão são desconhecidos. O buscador Google, por exemplo, reporta 2,5 bilhões de páginas indexadas, em 2002. Outras estimativas falam já em cerca de 5 bilhões de páginas acessáveis publicamente. Em termos de palavras, contudo, as previsões são mais incertas, mas os seguintes números já foram projetados:

Língua	Palavras estimadas*	% do total de páginas da rede**
Inglês	47,3 bilhões	43%
Português	925 milhões	2,5%

\*Grefenstette e Nioche (2000)

\*\* Global Reach, <http://www.netsizer.com/daily/TopCountry.html>.

Como se percebe, os valores em quantidade de palavras são muito maiores do que qualquer outro corpus em existência. O Bank of English, da língua inglesa, possui cerca de 400 milhões de palavras, enquanto o Banco de Português registra aproximadamente 220 milhões. Além disso, a tabela indica que os recursos disponíveis para o inglês são muito maiores do que para o português, o que significa que a quantidade e a variedade da evidência que as buscas em inglês retornam são em geral maiores. De qualquer modo, essas são apenas estimativas, e volume da Web está em constante expansão.

A bem da verdade, contudo, a WWW não é um corpus. Isso se dá por várias razões: ela não foi planejada para ser objeto de estudos lingüísticos, sua composição exata não é conhecida, não é possível selecionar parte dos arquivos apenas, etc. Além disso, a Web é limitada no que se refere a dados de fala. Mesmo assim, para os propósitos de encontrar evidências de uso de certos padrões, a Web constitui-se num ótimo recurso para análise de corpus, principalmente quando outros corpora propriamente ditos não estão disponíveis. Sabendo disso, vários autores e programadores já oferecem recursos para a exploração da Web como se fosse um corpus (p.ex. Fletcher, 2001). O serviço WebCorp ([www.webcorp.org.uk](http://www.webcorp.org.uk)), por exemplo, é um concordanceador para a Web que utiliza sites de busca (como o Google), visita as páginas indicadas pelo buscador, retira as citações segundo o termo de busca digitado pelo usuário e retorna os resultados em formato de concordância.



A Internet é ideal, porém, para obtenção dos textos focais envolvidos na primeira das pesquisas relatadas abaixo. Há uma abundância de textos com suas respectivas traduções na web. No Brasil, há portais que disponibilizam textos de jornais, por exemplo, publicados em inglês e suas versões para o português.

A seguir, são detalhadas os dois exemplos de pesquisas, sugeridas por esse trabalho como possíveis dentro dos propósitos e do contexto detalhados na discussão acima. A primeira enfoca um par de textos jornalísticos; a segunda, a versão brasileira de um slogan de campanha publicitária de uma multinacional de entregas (*courier*).

### 3. Análise 1: Texto jornalístico

A primeira pesquisa segundo as linhas descritas acima refere-se à análise das escolhas léxico-gramaticais de um texto de jornal, de um gênero que poderia ser descrito como ‘reportagem sobre lançamento de produto’, publicado no jornal The New York Times. O texto foi escolhido por se tratar de um tema curioso: banheiros japoneses. A versão brasileira desse texto foi disponibilizada no portal UOL, na seção ‘Jornais’.

Depois de baixados os dois textos, o primeiro passo consiste num alinhamento manual de ambos, com base nas divisões entre parágrafos. Isso pode ser feito num editor de texto, como o Word, usando formato de tabelas. Como no quadro a seguir.

O formato básico do alinhamento consiste em deixar paralelos os parágrafos dos dois textos, de tal modo que o início de cada parágrafo do texto-fonte coincida com o início do parágrafo correspondente do texto-alvo. A própria disposição em paralelo dos dois textos já fornece muita informação interessante para o aprendiz, como por exemplo de que os textos possuem extensões diferentes, que os parágrafos podem ou não ser correspondentes, etc. As diferenças e similaridades das divisões de parágrafo e da extensão deles podem ser discutidas com os alunos, trazendo em si só informações relevantes que podem iniciar um processo de conscientização acerca do papel *cumulativo* das escolhas lingüísticas (Berber Sardinha, 1997) no texto traduzido.

<p>Original em inglês, NYT Japanese Masters Get Closer to the Toilet Nirvana By JAMES BROOKE</p>	<p>Versão brasileira, NYT UOL Japão cria vasos sanitários com ar-condicionado James Brooke</p>
<p>NARA, Japan - Japan's toilet wars started in February, when Matsushita engineers here unveiled a toilet seat equipped with electrodes that send a mild electric charge through the user's buttocks, yielding a digital measurement of body-fat ratio.</p>	<p>Em Nara (Japão) A guerra dos banheiros no Japão começou em fevereiro, quando os engenheiros Matsushita introduziram assentos de privada equipados com eletrodos. Os assentos enviam leve descarga elétrica às nádegas do usuário, tirando uma medida digital da quantidade de gordura do corpo.</p>
<p>Unimpressed, engineers from a rival company, Inax, counterattacked in April with a toilet that glows in the dark and whirs up its lid after an infrared sensor detects a human being. When in use, the toilet plays any of six soundtracks, including chirping birds, rushing water, tinkling wind chimes, or the strumming of a traditional Japanese harp.</p>	<p>Nada impressionados, engenheiros de uma empresa rival, Inax, contra-atacaram em abril. Lançaram um assento que brilha no escuro e levanta a tampa quando detecta a presença de um ser humano, com sensor de luz infravermelha. Quando em uso, o vaso toca seis diferentes trilhas, incluindo piar de pássaros, água correndo, sinos de vento ou harpa tradicional japonesa.</p>
<p>Another factor stimulating toilet research is the fact that Japan's population is peaking and the number of households is expected to start declining by the end of the decade. Some money can be made by exporting toilets to countries with comparatively primitive toilet cultures, like China and Vietnam.</p>	<p>Outro fator são as pesquisas demográficas. Elas indicam que população do Japão começará a decair, junto com o número de casas, até o final da década. Pode-se fazer dinheiro exportando privadas para países com culturas sanitárias comparativamente primitivas, como China e Vietnã.</p>

Depois de alinhado, os alunos devem sublinhar, nos dois textos, aquelas escolhas lingüísticas que lhe chamaram a atenção, destacando escolhas que parecem discrepantes, ou seja, aquelas escolhas do tradutor que não parecem condizentes, apropriadas ou consistentes com o significado, estilo, registro, gênero, etc. do texto original.

A primeira questão que pode ser trabalhada está logo no início do texto: 'guerra dos banheiros' é uma tradução adequada para 'toilet wars',

naquele contexto? A idéia é pegar ‘toilet wars’ porque (1) é um grupo nominal, e grupos nominais são notoriamente problemáticos para traduzir, (2) seu equivalente citado no texto, ‘guerra dos banheiros’, parece incontroverso, pois é talvez a única tradução possível, e (3) é usado figurativamente, já que banheiros não lutam. Outras razões podem ainda ser arroladas para justificar esse item, mas a idéia principal é a de que a escolha dessa ou daquela palavra não importa muito, já que o foco não é no item lingüístico em si, mas na conscientização sobre a problemática envolvida nas escolhas lingüísticas.

O próximo passo consiste em verificar, num corpus de português, se ‘guerra dos banheiros’ já foi usado e, caso tenha sido, como foi. No corpus NILC, disponível no site do Processamento Computacional da Língua Portuguesa, uma busca por essa expressão (“*guerra*” “*dos*” “*banheiros*”) não retornou nenhuma ocorrência no corpus. Isso já indica que o termo é incomum no português. Contudo, esse achado não responde a pergunta de se ‘guerra dos banheiros’ é uma alternativa plausível; é preciso continuar a pesquisa, agora alterando o termo de busca, por exemplo, para ‘guerra dos’. Voltando ao mesmo site, o aluno deveria formular a sua busca como “*guerra*” “*dos*”. Dessa vez, aparecem 72 ocorrências. Como nesse site não é possível obter informação sobre os colocados, o usuário precisa observar os padrões diretamente na concordância. Para cursos introdutórios de Lingüística de Corpus e tradução, isso já é suficiente para a maioria das situações. Então, observando as ocorrências do termo de busca na concordância, nota-se que ele ocorre em padrões do tipo:

- (1) guerra dos pilotos da Penske
- (2) guerra dos Windsor
- (3) guerra dos Boers
- (4) Guerra dos Emboabas
- (5) Guerra dos EUA com a Espanha
- (6) guerra dos pneus
- (7) guerra dos telões
- (8) guerra dos superfrangos

Os alunos deveriam agora perceber que:

- ‘Guerra dos’ ocorre em padrões que indicam vários sentidos. Um deles, figurativo, é o de disputa entre pessoas de um certo grupo. Aqui, temos os exemplos 1 e 2 acima.
- A expressão também ocorre como nome de uma guerra propriamente dita: este é o sentido dos exemplos 3 e 4.
- Também ocorre para indicar a batalha em que se envolveram duas (ou mais) facções ou exércitos / marinhas, etc. Este é o exemplo 5.
- Mais importante, a expressão ocorre com nomes de produtos, como nos exemplos 6 a 8. Nesse caso, é usada figurativamente, significando, literalmente, algo como ‘disputa entre os produtores de pneus’. Aí, temos uma metonímia (‘pneus’ por ‘produtores de pneus’, ‘telões’ por ‘produtores de telões’, etc.) e uma metáfora (‘disputa comercial’ por ‘guerra’).

‘Guerra dos banheiros’ encaixa-se nesse último padrão. Portanto, é uma escolha feliz do tradutor, pois é uma expressão aceitável dentro do português, já que reflete uma preferência semântica importante de ‘guerra’, qual seja, a de ‘produtos’. Assim, ‘guerra dos banheiros’ é entendida como ‘disputa comercial entre os produtores de banheiros’<sup>1</sup>, o que é perfeitamente aceitável dentro da padronização de guerra já existente no português.

Uma questão que pode surgir nesse momento é a de se ‘toilet wars’ é uma expressão igualmente idiomática do inglês, ou se o autor a cunhou. O aluno pode efetuar uma busca por ‘toilet war’ em sites como Google ou WebCorp. No Google, o usuário encontra 79 ocorrências desse termo<sup>2</sup>. A conclusão geral, para os propósitos da atividade sugerida aqui, é que a expressão existe em inglês e que portanto o autor não a criou. O sentido imputado a ela pelo autor pode ser novo, mas a verificação disso já levaria o aluno para muito além do objetivo central da atividade, que é o de examinar as escolhas feitas pelo tradutor.

De qualquer modo, aqueles alunos que desejarem matar sua curiosidade não devem ser desencorajados. Entretanto, é importante lembrá-los

<sup>1</sup> Fica claro a partir dessa representação congruente que há uma outra metonímia envolvida aí, qual seja, a de ‘banheiros’ por ‘produtores de peças para banheiros’.

<sup>2</sup> Em setembro de 2002. Como o conteúdo da Web muda, as quantidades também se alteram.

de que o foco da atividade não é esse. Aqueles que desejarem pesquisar o uso de ‘toilet war’ no Google perceberão que não há exemplos claros do tipo ‘produto + war’, que seria similar ao padrão ‘guerra de + produto’ encontrado para a expressão em português. Os casos de ‘toilet war’ parecem referir-se a eventos pessoais, de difícil entendimento. Muitas ocorrências aparecem com pré-modificador, como por exemplo ‘Trans-Siberian Toilet War’ e ‘San Francisco Toilet War’. Para entender a referência dessas expressões, os alunos precisarão entrar nos textos respectivos e os lerem. A primeira expressão parece referir-se a um incidente ocorrido com turistas no banheiro de um trem russo, e a segundo à política de segurança levada a cabo pelas autoridades da cidade norte-americana de San Francisco, que permite que banheiros públicos sejam inspecionados pela polícia na tentativa de encontrar entorpecentes.

A segunda questão que chama atenção no texto é ‘unveil’, em ‘Matsushita engineers here *unveiled* a toilet seat’. Uma escolha *prima-facie* (Tognini-Bonelli, 2001) aqui seria ‘revelaram’.

‘Unveil’ não é um verbo muito freqüente no inglês (1,26 ocorrências por milhão de palavras, segundo o BNC), e portanto os alunos podem não conhecê-lo. Assim, se esse for o caso, é recomendável fazer com eles uma concordância dessa palavra e pedir que induzam o seu significado. Nesses casos, é melhor optar por um site que ofereça concordâncias mais legíveis do que o resultado de busca do Google, tais como ‘COBUILD Concordance Demo’ ou o ‘WebCorp’. Usando o COBUILD, são obtidas 40 linhas de concordância, o que já é suficiente para perceber algumas tendências. Algumas das linhas aparecem abaixo.

1. budget strikes. The government has unveiled a new land policy; for the moment
2. 18p at 632p. Great Universal also unveiled a 5 per cent rise in half-year
3. contagion spread to scientists, who unveiled a plan straight from the pages of
4. Korean-owned company Halla which unveiled a £ 17 million investment
5. flush with cash. [p] Yesterday it unveiled halfyear profits ahead by 4 to &
6. Although New York controversially unveiled its first black woman Mother
7. PLAQUE UNVEILED [/h] A plaque was unveiled in Budapest today in memory of
8. [p] The Rover kv6 engine was unveiled at the Brussels motor show which

Os alunos poderão perceber que:

- ‘unveil’ tem o sentido geral de ‘revelar’. Literalmente, significa ‘retirar o véu’. Esse sentido literal está presente, entre os casos destacados acima, somente na linha 7 (‘a plaque was unveiled’).

- Coisas que são ‘unveiled’ em geral são abstratas, como políticas, aumentos, planos, investimentos e indicações de pessoas para cargos (linhas 1 a 6).

Espera-se, nesse ponto, que os alunos percebam o sentido de ‘unveil’ como o de revelar. O tradutor não usou essa opção, contudo; ele preferiu ‘introduziram’.

O próximo passo para os alunos seria contrastar as seguintes opções: ‘revelar’ e ‘introduzir’, cada um seguido de nome de um produto. Usando corpus NILC no site do Processamento Computacional do Português, uma busca por ‘introduzir’ retorna 266 ocorrências. É possível diversificar a pesquisa buscando outras formas do verbo, mas para os propósitos pedagógicos aqui seguidos, limitar-se apenas ao infinitivo já é suficiente. O aluno pode ir lendo e anotando algumas ocorrências dos colocados de ‘revelar’ na concordância, tomando o cuidado de ignorar as ocorrências de ‘se revelar’. Alguns colocados são os seguintes:

<aspectos, detalhes, dinâmicas, estratégias, formas, informação, intimidades, ironia, jogadores, limites, material, presença, problemas, processos, quantia, segredos>

Percebe-se que o conjunto lexical indica uma preferência por elementos abstratos, o que lança dúvidas sobre o candidato ‘revelar’.

Por sua vez, ‘introduzir’ traz consigo os seguintes colocados (também retirados a partir de uma leitura de parte da concordância da forma ‘introduzir’, que agrega 717 ocorrências):

<cláusula, competitividade, conceitos, concepções, condutas, construtos, elemento, filho, gene, item, leis, melhorias, mudanças, noção, oxigênio, ressalvas, rigor, tema>

Há apenas uma ocorrência de ‘introduzir’ seguido de produto:

Para aplicar é necessário tirar a válvula do pneu e introduzir o produto.

Entretanto, o sentido aqui é o literal de ‘inserir’, que não se aplica ao caso do texto do jornal.

Outra opção nesse momento é fazer uma busca em outro corpus. Como não há outros corpora de português acessáveis online, é preciso recorrer

aos buscadores de WWW ou ao WebCorp. Google retorna zero ocorrências de *revelar um novo produto* e 37 ocorrências de *introduzir um novo produto*. Uma ocorrência típica desse último caso é ‘introduzir um novo produto no mercado’. Os alunos podem efetuar mais buscas, substituindo ‘produto’ por nomes de produtos verdadeiros ou por palavras relacionadas a eles, como carro, eletrodoméstico, modelo, etc.

Esse resultado indica que embora ‘introduzir seguido de produto’ não seja o padrão mais comum de ‘introduzir’, ele já faz parte da língua portuguesa, referindo-se ao ato de lançar um novo produto no mercado, colocando-o à venda. A propósito, *lançar um novo produto* possui 224 ocorrências no Google, o que pode indicar que talvez fosse uma escolha mais comum.

Tendo agora estabelecido que quando seguidos de ‘produto’ ou ‘nome de produto’, ‘revelar’ é incomum, ‘introduzir’ é relativamente comum e ‘lançar’ é o mais freqüente dos três, os alunos podem apreciar as possíveis diferenças de sentido entre os três verbos e compará-las com o sentido de ‘unveil’. Podem, ainda, contrastar ‘unveil’ com outros verbos com o sentido parecido, como ‘introduce’ e ‘release’, gerando mais concordâncias no COBUILD ou fazendo buscas no Google ou WebCorp. Essas buscas podem ser mais ou menos complexas, com graus diferentes de aprofundamento, dependendo da familiaridade da turma com os instrumentos e mesmo com a língua.

Por exemplo, os alunos podem se perguntar se ‘unveil’ seguido de um produto é comum no inglês ou não. Há várias opções para fazer uma busca nesse sentido. Uma delas é com a lista de colocados de ‘unveil’. Caso o aluno tenha à sua disposição um corpus gravado em seu computador (ou na rede local), poderá usar um programa como WordSmith Tools para fazer uma concordância do item e depois obter a lista de colocados. Outra alternativa é o próprio CD de colocações do COBUILD, que fornece lista de colocados de milhares de palavras do inglês. Contudo, quando esses recursos não existem, como é o caso da situação-alvo descrita aqui, isso não é possível e o aluno terá de recorrer à Internet. Na WWW, o projeto COBUILD fornece listas de colocados em seu site (o mesmo citado acima).

Antes de tentarem responder essa pergunta, é aconselhável fazer um trabalho preparatório com eles sobre modos de aferir o grau de associação entre palavras. No contexto de uso em sala de aula, fica a critério do professor se vale a pena entrar, quando entrar e como abordar o assunto sobre

estatísticas de associação. Abaixo segue uma pequena demonstração do tipo de direcionamento que isso pode tomar.

Na página do site referente ao Collocation Sampler há uma janela para digitação do item a ser pesquisado e duas opções de estatística de associação: Mutual Information (MI) e T-Score. A utilização de estatísticas é inteiramente dispensável para identificação de colocados para os propósitos de observação das escolhas de tradutores em contextos pedagógicos. Mas como no site é obrigatório o usuário escolher uma estatística, é preciso expor em algum detalhe o que isso representa.

Cada estatística tende a trazer resultados diferentes, mesmo que seja apenas no grau de associação reportado por cada uma e não na composição dos colocados. De modo sucinto, a diferença entre as duas é que o MI captura associações em que os itens aparecem juntos *frequentemente em termos relativos* (força de colocação, segundo Hunston 2002), enquanto o T-Score aponta associações em que os itens co-ocorrem *frequentemente em termos absolutos* (certeza de colocação, também segundo Hunston 2002). Isso significa dizer que o MI localiza pares de palavras que aparecem juntos muitas vezes em relação à frequência de um e de outro, mesmo que cada um ocorra poucas vezes. Já o T-Score dá preferência aos colocados que são, por si só, muito frequentes. A razão dessa diferença está na fórmula de cálculo das duas estatísticas – enquanto MI é apenas uma razão entre o valor observado e o esperado, T-Score incorpora ainda o desvio padrão, que é medida de dispersão em relação à média; assim, palavras com alta frequência possuem médias mais altas e menor dispersão.

Por exemplo, os colocados de ‘coffee’ com maior MI incluem:

Colocado	Frequência do colocado sozinho	Frequência do colocado com a palavra de busca	Valor da estatística
percolator	5	3	10.430806
decaffeinated	36	17	10.085275
nescafe	23	7	9.451467
starbucks	113	31	9.301676
Bodum	11	3	9.293189
arabica	12	3	9.167646



Já entre os com maior T-Score estão:

Colocado	Frequência do colocado sozinho	Frequência do colocado com a palavra de busca	Valor da estatística
Cup	20198	309	17.078583
And	1369241	1163	16.637780
Tea	3320	276	16.526319
A	1228514	953	13.560089
Table	8066	150	11.960971
Or	175734	250	10.976745
Shop	6127	100	9.733482

É importante ressaltar primeiramente que a lista de colocados é fornecida com palavras retiradas de um horizonte de até quatro palavras de distância do nóculo (palavra de busca); sendo assim, não é possível ter certeza sobre a exata composição das expressões, nem das suas frequências. O valor da terceira coluna das tabelas acima refere-se à frequência mútua dentro do horizonte, por isso os colocados podem ter ocorrido em qualquer uma das posições disponíveis.

Tendo feito essas ressalvas, como se nota, os colocados com MI maior indicam colocações mais específicas, como ‘coffee percolator’, ‘decaffeinated coffee’, ‘nescafe coffee’, etc. Já T-Score sugere colocações mais frequentes, como ‘cup of coffee’, ‘a coffee’, ‘coffee table’, ‘tea or coffee / coffee or tea’, etc. Para ter certeza da ocorrência dessas colocações, é preciso rodar concordâncias delas no corpus.

Voltando à questão de se ‘unveil’ seguido de um produto é uma colocação comum no inglês, os resultados das estatísticas de associação fornecidas pelo site do COBUILD Collocation Sampler são:

- MI** = plaque, blueprint, memorial, plans, proposals, officially, package, plan, profits, gorbachev, due, expected, figures, details, tonight, latest, results, today, tomorrow, set
- T-Score** = to, will, new, plans, a, today, plan, its, expected, set, is, due, their, plaque, package, proposals, details, next, government, for

Segundo os colocados, a resposta é afirmativa: ‘unveil um produto’ não é muito comum, já que os colocados reiteram o que já havia sido descoberto antes pela interpretação da concordância: uma predominância de palavras abstratas. Contudo, há duas ressalvas importantes nesse momento. A primeira é que isso não significa necessariamente que ‘unveil um produto’ seja uma criação do autor (mesmo que ‘unveil a toilet seat’ o seja). A segunda é que a versão do corpus COBUILD de onde vêm as listas de colocados não é muito extenso (45 milhões) e pode estar desatualizado (note por exemplo o colocado ‘Gorbachev’).

Assim, outra opção é pesquisar na própria Web, com Google ou WebCorp. Como nenhum deles fornece lista de colocados, o usuário deve entrar a palavra ‘unveil’ sozinha ou seguida de um possível colocado. A busca por *unveil* no Google retorna mais de 288 mil ocorrências, e *unveiled*, 744 mil; entre as primeiras, já aparecem ‘unveil plastic plants’ e ‘unveil nanotechnology’, que são exemplos de produtos. Confirmada essa possibilidade, é possível partir diretamente para uma busca, por exemplo, por *unveil new product*, que possui 111 ocorrências, segundo Google. O usuário deve tentar outras alternativas, como por exemplo *unveiled a new product*, que ocorre 461 vezes. Outra possibilidade é *unveiled a new car*, com 39 ocorrências. *Unveiled a toilet* retorna 34 ocorrências, mas são todas iguais, retiradas do texto do New York Times. *Unveiled a new toilet* possui duas ocorrências, e *unveil new toilet* e *unveiled new toilet* não retornam nenhuma citação. As ocorrências dessas expressões incluem:

McDonald's Top Fuel team unveiled a new car in Denver at the Mopar Parts Mile-High Nationals...

Ferrari now look to be more of a force in the constructors championship with two cars, and they have also unveiled a new car...

Daewoo unveiled a new car — the Kalos ...

Nokia has today unveiled a new product ...

Network Flight Recorder has unveiled a new product designed to keep hackers from hiding their tracks...

Esses resultados demonstram que ‘unveil’ seguido de nome de um produto (ou da própria palavra ‘product’) faz parte da língua inglesa, notadamente de textos que se referem ao mundo dos negócios, já que grande parte segue o padrão ‘Empresa + unveil + produto’. Este é justamente o padrão que foi usado no texto-fonte, corroborando a escolha ‘introduzir + produto’, que também possui características parecidas de seleção de registro. Além disso, esses resultados complementam a informação

obtida no corpus do COBUILD, adicionando que embora ‘unveil’, na língua geral, tenha como complemento, predominantemente, nomes abstratos, há, na linguagem de ou sobre negócios, a colocação ‘unveil’ seguido de produto.

A conclusão geral a que se chega é que a tradução de ‘unveiled’ como ‘introduziram’ é legítima e perfeitamente adequada.

Nos textos, há várias outras ocorrências que podem ser investigadas usando os procedimentos discutidos acima, tais como:

- ‘Body-fat ratio’ como ‘quantidade de gordura do corpo’, no primeiro parágrafo (vide trecho do texto alinhado, acima).
- ‘whirs up its lid’ como ‘levanta a tampa’; ‘with a toilet that glows in the dark’ como ‘lançaram um assento que brilha no escuro’; e ‘the strumming of a traditional Japanese harp’ como ‘harpa tradicional japonesa’, no segundo parágrafo.
- Apagamento de ‘stimulating toilet research’; ‘number of households is expected to start declining’ como ‘junto com o número de casas’; e ‘Japan’s population is peaking’ como ‘população do Japão começará a decair’, no terceiro parágrafo.

Algumas dessas questões são mais difíceis de pesquisar, como a referente à tradução ‘por zero’ de ‘stimulating toilet research’, pois não há equivalente concreto na língua de chegada para servir de termo de busca.

Além de verificar as escolhas uma a uma, como é sugerido aqui, os alunos podem avaliar o conjunto das escolhas. Por exemplo, um dos casos destacados, o de traduzir o adjunto ‘with a toilet that glows in the dark’ por uma oração independente (‘lançaram um assento que brilha no escuro’), pode ser uma indicação de explicitação (Baker, 1995). Outras escolhas, como ‘whirs up its lid’ e ‘flipping its lid’ (não incluído nos trechos acima) ambas traduzidas por ‘levanta / levantando a tampa’ indicam possível normalização.

#### 4. Análise 2: Campanha publicitária

Um outro tipo de atividade que tende a ser motivante é a pesquisa com traduções de campanhas publicitárias. Os alunos podem fazer análise

de peças inteiras ou dos slogans apenas. Devido ao seu alcance social, mercadológico e aos contornos ideológicos que trazem no seu bojo, slogans são um meio muito eficaz de sensibilizar os alunos acerca das importância que as escolhas lingüísticas possuem na sociedade. As tendências ou problemas descobertos pelos alunos são percebidos como tendo reais repercussões na maneira como a mensagem publicitária pode ter sido apreendida pelo público, sem falar na própria sobrevivência de uma marca, da venda do produto anunciado e da imagem da empresa. Pelo fato de não ser, necessariamente, feita por tradutores propriamente ditos, a versão de slogans não se configura em si numa habilidade a ser adquirida pelos alunos. Por isso, o tipo de atividade descrito abaixo tem um papel fundamentalmente de conscientização acerca do papel das escolhas lingüísticas no impacto da mensagem publicitária, muito mais do que prática com um tipo de texto com que eles provavelmente se envolverão profissionalmente.

O trabalho com peças publicitárias, e slogans em particular, suscita questões importantes sobre a cultura do país que importa a campanha e, também, sobre a cultura do país que a produziu. Fox (1998) menciona uma campanha de chocolates ingleses que usava o slogan 'another load of [marca]' pintado em letras garrafais nos caminhões de entrega do produto. Ela suspeitou dessa construção e pesquisou num corpus grande, descobrindo que 'load of' era tipicamente usado em inglês com complementação desagradável (excrescências!). Daí percebe-se, claramente, que a escolha do slogan, na língua original mesmo, era problemática, pois com certeza os publicitários pensaram em associar a entrega dos chocolates a algo agradável, gostoso, e não a algo repugnante! Berber Sardinha (no prelo-b) notou, num estudo sobre um slogan político da prefeitura de uma grande cidade brasileira, que havia também um conflito entre o sentido desejado pela campanha e o sentido já adquirido pela expressão em outros contextos. A campanha centrava-se no uso da expressão 'tocando para frente'. O autor descobriu, num corpus de grandes proporções de português do Brasil, que essa expressão estava associada, em boa medida, a uma ação motivada por condições adversas, geralmente executada a contra-gosto. Novamente, há uma contradição séria entre o sentido desejado pela campanha, de que a prefeitura estava trabalhando pela cidade, e o que a expressão já possuía na língua, de algo com que alguém se envolve involuntariamente por infortúnio.

O slogan escolhido para a atividade-exemplo mostrada aqui é ‘Consider it Done’, de uma empresa de entregas, cuja versão brasileira é ‘Dito e Feito’. A seguir é apresentado um relato de análise desse slogan, que pode ser refeita ou aproveitada em cursos de formação de tradutores.

Primeiramente, é importante chegar a algum consenso acerca da mensagem almejada pelo slogan. Nesse caso, a mensagem almejada parece clara: a empresa se compromete a entregar o pacote; o cliente, por sua vez, pode tranquilizar-se ao contratar os serviços da empresa – uma entrega contratada é sinônimo de entrega feita. Em inglês, portanto, a mensagem parece perfeitamente coerente com a imagem que se espera de uma empresa desse ramo.

No COBUILD online, a busca por ‘consider it done’ retornou apenas duas ocorrências:

Dayton as this week's clones. Consider it done, Jonathan, and keep your it? Add 15 per cent to the cost and consider it done. A dinner service to match

As duas ocorrências parecem trazer os dois sentidos principais da expressão: o primeiro, figurativo, como promessa a alguém de que algo será executado; e o segundo (de modo menos claro, por falta de co-texto mais amplo) como algo mais concreto, que pode ser tido como concluído, terminado. Esses dois sentidos de ‘done’ são pertinentes à mensagem publicitária, pois se completam: de um lado, expressão de promessa de entrega, e de outro, expressão de algo concluído; aqui reside um grande trunfo dessa expressão em inglês para esse fim: a combinação de duas leituras, a de que a contratação do serviço de entrega é igual à entrega já feita, com a de que a responsabilidade do cliente termina com a contratação, pois a empresa tomará conta do resto.

Alguns alunos podem não ter conhecimento do uso de ‘done’ como ‘terminado’, ‘concluído’, apesar da proximidade com o português. Se esse for o caso, é aconselhável fazer uma concordância de ‘done’. O problema aqui é que ‘done’ apenas como termo de busca retornará muitos casos indesejáveis, de ‘done’ em grupos verbais precedido de verbo auxiliar (has/had/is done, etc.) Uma alternativa é buscar colocações como ‘I’m done’. No COBUILD online, essa busca pode ser formulada como *m + 1done*, que significa ‘m seguido de done como primeira palavra à direita’. Isso retorna

citações como “‘Are you done?’ asked Karen. ‘I’m done. Thanks, Chet.’”, que exemplifica esse sentido de ‘acabado’ (ie, ‘você acabou?’).

A questão que fica é se ‘dito e feito’ captura as mesmas mensagens presentes na campanha em inglês. Curiosamente, ‘dito e feito’ também foi usado em recente campanha eleitoral, como ‘slogan’ de um candidato. Isso reforça o apelo do tema.

Pode-se começar pelo site do Processamento Computacional do Português, tendo o corpus NILC como base. O termo de busca, por ser composto, deve ser formulado com cada palavra entre aspas – no nosso caso, foi “*dito*” “*e*” “*feito*”. Foram retornadas apenas duas ocorrências. Buscando nas demais opções de corpora disponíveis no site, os resultados também foram poucos. Além disso, a quantidade de co-texto oferecida era pequena, tornando difícil a interpretação do sentido da expressão.

Por isso, nesse caso uma opção melhor, por ser mais flexível, foi o WebCorp. O usuário pode escolher a quantidade de co-texto que deseja. Embora o motor de busca do WebCorp utilize o Google<sup>3</sup>, ele de fato entra nos textos encontrados e retira o contexto desejado pelo usuário. Para ‘dito e feito’ pareceu aconselhável trabalhar com co-texto amplo, de por exemplo, 25 palavras para cada lado do nóculo, para permitir um entendimento melhor sobre como o termo foi usado.

Utilizando, então, o WebCorp com 25 palavras de contexto, foram registradas 185 ocorrências na Web, retiradas de 120 páginas da rede (14 das quais apresentaram erro). Dessas 185, apenas 27 foram consideradas – as demais caíram em um ou mais destes casos: repetições, trechos incompreensíveis, títulos ou subtítulos, nomes de serviços, produtos, programas de televisão, sites, chats, etc. Esse passo referente à ‘limpeza’ dos dados é muito importante e deve ser enfatizado aos alunos; também é recomendável que o professor auxilie os alunos no processo, já que demanda uma certa experiência no manuseio e leitura de concordâncias.

O próximo passo foi a interpretação das ocorrências, tendo em mente que o objetivo da busca foi o de descobrir como ‘dito e feito’ é usado comparativamente a ‘consider it done’.

---

<sup>3</sup> Como default, mas pode ser mudado.

Cinco casos tinham o sentido literal de ‘algo que foi dito e foi feito’, formados pelo padrão ‘[verbo ser] dito e feito’, como por exemplo:

- Depois que tudo foi dito e feito...
- Quando tudo isso é dito e feito, ...

Em outros oito casos, o sentido predominante era o de cumprimento de algo anunciado ou prometido por si mesmo – alguns desses exemplos seguem abaixo:

- assim que soube que os resultados do Enem poderiam ajudá-la a ter uma nota maior, não teve dúvidas: resolveu estudar para a prova. **Dito e feito.** Graças aos resultados do exame, hoje ela é uma aluna da São Francisco.
- Então porque não saímos às ruas perguntando o que eles achariam importante noticiar? E mais: quais os principais problemas de Londrina? **Dito e feito.** No dia 27 de março, foram ouvidas 39 pessoas no Calçadão e na Biblioteca Pública Municipapl [sic].
- Como ficamos sem isca, ocorreu-me uma brilhante idéia. “Vamos usar a goma de mascar, (o popular chiclete), já que tenho alguns comigo”, lembrei. **Dito e feito.** Ficamos mastigando sem parar aquela borrachinha perfumada e doce. Varas na água novamente e, grudadinho no anzol, aquele chicletinho sem-vergonha!

Os demais quatorze casos expressavam o sentido de acontecimento anunciado ou profetizado, como os exemplos abaixo ilustram:

- Rogério disse que em BH também estava chovendo bastante, mas que eles iriam varrer a chuva de Porto. Foi **dito e feito.** Até o final do show, que foi perfeito, não caiu sequer uma gota d’água.
- O diretor então afirmou: “Essa é a evolução da espécie e a evolução da Avon. Em pouco tempo, as mulheres mandarão na companhia”. **Dito e feito.** Sete anos depois, Andrea tornava-se a primeira mulher a presidir as operações mundiais da Avon, a maior companhia de vendas diretas do mundo.
- Partimos na sexta à noite, para levantar sábado bem cedo para aproveitar o dia, que prometia fazer um belo sol. **Dito e feito.** Sábado amanheceu com um “solão”, um dia muito quente.

Esse último grupo de sentido pode também incluir prenúncio de algo negativo:

- Na hora falei para o Cláudio que esse cara iria atrapalhar um de nós. **Dito e feito**, na largada o cara se enroscou no meu amigo e eu fui embora. Depois da muvuca logo primeira subida formou-se o primeiro pelotão.
- Montamos nas motos e agora o Destroyer foi na frente. Num mato, havia um belo buraco escondido, apenas esperando um prego passar. **Dito e feito**: Destroyer dentro do buraco e novo capote...
- só um dos gêmeos tinha o problema. Era provável que uma mutação tivesse ocorrido em um dos meninos, gerando a doença. **Dito e feito**: a análise do genoma dos dois revelou uma diferença das mais suspeitas, a deleção (apagamento) de 18 pares de bases do gene IRF6.

A análise revelou que há um sentido de ‘dito e feito’ que é comum (o mais freqüente na concordância enfocada) que não é compatível com o sentido de ‘consider it done’. Compreensivelmente, pode-se argumentar que os dois outros sentidos (‘dizer e fazer literais’ e ‘cumprimento de promessa pessoal’) são compatíveis com a mensagem desejada na campanha publicitária, já que tanto ‘consider it done’ quanto ‘dito e feito’ significam que algo que foi dito (ou prometido) foi (ou será) realizado, ou cumprido. Entretanto, o problema é que há também esse outro sentido comum, em português, de expressar o cumprimento de um mal presságio, ou de prenúncio de desastre, doença e outros eventos negativos, com ‘dito e feito’. Em inglês, ‘consider it done’ não atua com essa função. Portanto, há um problema claro na tradução do slogan, pois embora a análise não revele qual sentido predomina, ela indica que um sentido indesejado para a mensagem publicitária está presente. Interpretado assim, o slogan em português poderia significar, de modo absurdo, que uma entrega contratada junto à empresa traz consigo um problema, talvez mesmo um atraso ou acidente!

## 5. Comentários finais

Mais do que analisar escolhas dos tradutores, o trabalho sugerido aqui visa a sensibilizar alunos para a importância do uso de corpora no processo



tradutório. Foi sugerido que as atividades possam ser desenvolvidas com alunos de tradução, de tal modo que elas forneçam condições para que esses alunos, ao explorarem corpora eletrônicos, possam se conscientizar da complexidade e da especificidade das escolhas lingüísticas envolvidas no processo tradutório. Mais do que o conhecimento adquirido sobre o uso ou equivalência desse ou daquele item, pedagogicamente o aspecto mais relevante delas é a possibilidade de conscientização acerca das regularidades existentes na língua.

A familiarização com os procedimentos envolvidos na pesquisa com corpora é outro benefício do tipo de atividade proposto aqui. Esses procedimentos podem ser transferidos depois a outras tarefas mais complexas, com alunos mais experientes, formando, por exemplo, a base do tipo de pesquisa avaliativa de tradução que Bowker (2001), Kenny (2001) e Berber Sardinha (no prelo-a) propõem. Essas habilidades, uma vez adquiridas, também podem ser transferidas para o contexto profissional, a fim de auxiliar o profissional a produzir traduções de maior qualidade (Bowker, 2000).

Conforme dito na introdução, um dos desafios para a popularização da pesquisa em tradução baseada em corpora é a formação de pessoal novo, o que se defronta com problemas dos mais variados referentes à falta de infraestrutura. Problemas com uso de ferramentas e corpora online são conhecidos: dificuldade de conexão, lentidão de acesso, queda de serviço ou servidor, etc. Alguns desses problemas são referentes à infraestrutura do local onde os alunos trabalham, outros fogem ao controle dos usuários, pois são ocorrem no lado de quem está fornecendo o serviço, isto é, os próprios sites que os alunos acessarão. O que se mostrou aqui foi uma maneira de propiciar aos alunos contato com o manuseio de corpora para enfocar questões da tradução com exigências mínimas de recursos. A orientação do trabalho partiu da idéia de que uma proposta que exige recursos mínimos não é sinônimo de uma proposta pobre; pelo contrário, quando se usa os recursos já existentes de forma racional e coerente com os pré-requisitos teóricos e aplicados da área, o resultado tende a compor uma experiência de aprendizado enriquecedora.

Uma crítica que se faz ao tipo de análise com corpora apresentado aqui é que ela vem a complicar a vida do aluno e do profissional. Com ela, eles precisam empenhar-se em análises que são, muitas vezes, laboriosas, demoradas e detalhadas, para ‘descobrirem o óbvio’, ou, como é geralmen-

te posto, 'para descobrir o que o dicionário já dizia'. Em primeiro lugar, não é verdade que as análises confirmam o que os dicionários bilíngües (ou monolíngües) já dizem. Vários estudos já demonstraram que a informação oferecida pelos dicionários é incompleta e falha; o mesmo pode ser dito da intuição. Um estudo que verificou a qualidade da tradução feita com e sem corpora, por alunos, também atestou que os que não tiveram acesso a corpora tiveram mais problemas. De qualquer modo, não se deve encarar o objetivo do tipo de atividade sugerida aqui como o de chegar a um produto ('fazer uma boa tradução'), mas o de melhorar e enriquecer o processo da tradução ('como eu posso aperfeiçoar minha tradução').

Uma das possíveis reações ao trabalho com corpora é o sentimento de o aluno estar muitas vezes 'duvidando de si mesmo'. Parece natural que no início os alunos tenham esse sentimento, ao se confrontarem com a possibilidade de que as escolhas que eles fariam ou que os tradutores já fizeram possam se mostrar inadequadas perante uma análise baseada em corpora. Mas também é de se esperar que esse sentimento dê lugar, com o passar do tempo, a uma segurança maior à medida que o aluno passe a descobrir quantas vezes as escolhas se confirmam. Mais importante, o conhecimento que ele tende a adquirir com o manuseio dos corpora é muito mais variado e rico do que obteria de outro modo.

À medida que o aprendiz (ou mesmo o tradutor) se torna um usuário mais experiente de corpora, percebe que as escolhas lingüísticas não giram em torno de uma questão polarizada, entre certo e errado, de chegar a escolha única, inquestionável. Como mostra Munday (2000), corpora tendem a mostrar várias opções, algumas mais apropriadas ao contexto, outras menos. Não se trata de usar o corpus como uma espécie de tradutor automático, que fornece o equivalente e no qual o usuário deve 'depositar uma fé cega' (Bowker, 1999); quem tem a palavra final é o tradutor, que deve examinar e levar em conta a evidência fornecida pelos corpora. No fundo, é uma questão de perceber e aprender a lidar com probabilidades, que, na verdade (e este é o verdadeiro 'clique' da conscientização!), já estão internalizadas pelos usuários de uma língua; os corpora apenas registram o resultado cumulativo delas. Tomar consciência da existência dessas probabilidades, de como elas se manifestam e como podem ser pesquisadas são práticas profundamente enriquecedoras para o aprendiz de tradutor. Corpora eletrônicos perfazem esse papel como nenhum outro recurso.

E-mail: tony4@uol.com.br  
Recebido em outubro de 2002

**REFERÊNCIAS**

- AUSTERMÜHL, F. 2001. *Electronic Tools for Translators*. Manchester: St. Jerome.
- BAKER, M. 1993. Corpus Linguistics and translation studies: Implications and applications. In: M. BAKER et al (org.). *Text and technology: In honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins.
- \_\_\_\_\_. 1995. Corpora in Translation Studies: An Overview and some Suggestions for Future Research. *Target*, 7: 223-243.
- \_\_\_\_\_. 1996. Corpus-based Translation Studies: the challenges that lie ahead. In: H. SOMERS (org.). *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager*. Amsterdam: John Benjamins.
- BERBER SARDINHA, A. P. 1997. Patterns of lexis in original and translated business reports: Textual differences and similarities. In: K. SIMMS (org.). *Translating Sensitive Texts: Linguistic Aspects*. Amsterdam: Rodopi.
- \_\_\_\_\_. 2001. Lingüística de Corpus e tradução: Algumas ferramentas computacionais. Paper presented at Translating the New Millenium - Corpora, Cognition and Culture, The 8Th National Translation Forum and the 2nd Brazilian International Translation Forum, Belo Horizonte, 23-27 July 2001.
- \_\_\_\_\_. no prelo-a. Corpora eletrônicos na pesquisa em tradução. *Cadernos de Tradução*.
- \_\_\_\_\_. no prelo-b. *Lingüística de Corpus*. São Paulo: Editoria Manole.
- BOWKER, L. 1999. Exploring the potential of corpora for raising language awareness in student translators. *Language Awareness*, 8.3/4: 160-172.
- \_\_\_\_\_. 2000. Towards a methodology for exploiting specialized target language corpora as translation resources. *International Journal of Corpus Linguistics*, 5.1: 17-54.
- \_\_\_\_\_. 2001. Towards a methodology for a corpus-based approach to translation evaluation. *Meta*, 46: 345-364.
- \_\_\_\_\_. 2002. *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- DONMAL, B. G. Org. 1985. *Language Awareness* (NCLE Reports and Papers: 6). London: Centre for Information on Language Teaching and Research.
- FLETCHER, W. 2001. Concordancing the Web with KwiCFinder. Fletcher, W 2001 Concordancing the Web with KwiCFinder,, trabalho apresentado no The Third North American Symposium on Corpus

- Linguistics and Language Teaching, March 23-25, 2001, Boston, MA, disponível em [miniapolis.com/KWiCFinder/FletcherCLLT2001.pdf](http://miniapolis.com/KWiCFinder/FletcherCLLT2001.pdf).
- FOX, G. 1998. Hocus pocus and graven images: collocation '98. Talk given at IATEFL 98, UMIST, Manchester, UK.
- GREFENSTETTE, G. & J. NIOCHE. 2000. Estimation of English and non-English Language Use on the WWW. Proceedings of RIAO'2000, "Content-Based Multimedia Information Access", Paris, April 12-14, 2000, pp. 237-246.
- HOEY, M. & D. HOUGHTON. 1998. Contrastive Rhetoric and Translation. IN: M. BAKER Org. *Routledge Encyclopedia of Translation Studies*. London: Routledge.
- HUNSTON, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- KENNY, D. 2001. *Lexis and Creativity in Translation - A Corpus-Based Study*. Manchester: St. Jerome.
- KILGARRIFF, A. & G. GREFENSTETTE Orgs. (in press). *The Web as Corpus: Special Issue of Computational Linguistics*. Cambridge, Mass: MIT Press.
- LAVIOSA, S. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications* (Approaches to Translation Studies: 17). Amsterdam: Rodopi.
- MUNDAY, J. 2000. Seeking translation equivalents: A corpus-based approach. In: A. CHESTERMAN et al (org.). *Translation in Context: Selected Contributions from the EST Congress, Granada 1998*. Amsterdam / Philadelphia: John Benjamins.
- PARTINGTON, A. 1998. *Patterns and Meanings - Using Corpora for English Language Research and Teaching* (Studies in Corpus Linguistics 2). Amsterdam/Philadelphia: John Benjamins.
- SINCLAIR, J. 2001. Review of The Longman Grammar of Spoken and Written English. *International Journal of Corpus Linguistics*, 6.2: 339-360.
- THORNBURY, S. 1997. *About Language - Tasks for Teachers of English*. Cambridge: Cambridge University Press.
- TOGNINI-BONELLI, E. 2001. *Corpus Linguistics at Work* (Studies in Corpus Linguistics: 6). Amsterdam/Atlanta, GA: John Benjamins.