

CORPORA ELETRÔNICOS NA PESQUISA EM TRADUÇÃO

Tony Berber Sardinha
LAEL, PUC/SP

1. A importância do uso de corpora na tradução

Há uma unanimidade entre os pesquisadores da tradução e os lingüistas de corpus em torno da questão da utilização de corpora eletrônicos na tradução: o posicionamento corrente é o de que tanto os estudos tradutológicos como área acadêmica de pesquisa, quanto a prática tradutória, têm muito a ganhar com um contato maior com a Lingüística de Corpus.

Entre os lingüistas de corpus, a posição de Hunston (2002) ilustra bem o pensamento desse grupo:

Corpora ... have more to offer translators than might at first sight be apparent. Not only can they provide evidence for how words are used and what translations for a given word or phrase are possible, they also provide an insight into the process and nature of translation itself. (p. 128).

De certo modo, a utilização de corpus na pesquisa lingüística deixou de ser uma opção. A força da evidência obtida na exploração de corpora tem deixado claro que a linguagem é organizada de

um modo muito mais complexo do que se imaginava. Conforme coloca Hunston (2002, p. 216):

new ideas about language emerge and the old ones may need re-evaluation. Our own roles may change. In some ways ... our lives become more complex, simply because it is much harder to ignore the endless intricacy of language itself.

Para Tognini-Bonelli, ignorar a evidência do que um corpus pode oferecer não é apenas inapropriado, mas também perigoso (Tognini-Bonelli, 2002, p.74), porque a Lingüística de Corpus tem mostrado repetidas vezes quão inexata é a intuição humana no entendimento da linguagem (Sampson, 2001). Este ponto, aliás, já havia sido demonstrado de forma notável por Labov (vide Sampson, 2001), mas ficou ainda mais claro recentemente, conforme os resultados de pesquisas baseadas em corpus se tornaram públicos. Vale ressaltar, contudo, o importante papel que a intuição desempenha como ponto de partida na pesquisa tradutológica com corpus (vide abaixo).

Entre os tradutores, o reconhecimento do valor de corpora eletrônicos não é muito diferente. Para Tymoczko (1998), a influência do uso de corpus na tradução tem vários efeitos positivos, entre eles o de permitir um maior intercâmbio de dados entre pesquisadores e praticantes, além de modernizar a área, trazendo-a mais perto do que se espera da pesquisa contemporânea:

Corpus translation studies change in a qualitative as well as a quantitative way both the content and the methods of the discipline of Translation Studies, in a way that fits with the modes of the information age. (p. 652)

Boa parte do interesse em corpora eletrônicos na área de tradução provém de necessidades reais de tradução automática de grande volume de dados. É o caso da União Européia, que, como parte

de sua existência, tem de lidar com documentos que precisam ser traduzidos para as várias línguas da comunidade. Diversos projetos importantes de tradução assistida por corpus têm sido financiados pela União Européia (Hunston, 2002, p. 123).

2. A lenta integração da Lingüística de Corpus com a Tradução

O cenário apresentado na seção anterior levaria qualquer um a supor que há um casamento duradouro entre corpus eletrônico e tradução. Por exemplo, em 1993, portanto há praticamente dez anos, McEnery e Wilson concluíam seu *working paper* sobre Lingüística de Corpus e tradução (McEnery & Wilson, 1993) com a seguinte frase:

the presence of corpora in translation studies, as well as other areas of linguistic study, seems destined to become ever greater. (p.10)

Laviosa (1998b) estima que desde a publicação do trabalho pioneiro de Baker, em 1993, um “número crescente” de estudiosos e praticantes da tradução tenha adotado metodologias baseadas em corpora para estudar aspectos da tradução.

Entretanto, não é isso que se percebe. O “Cumulative Index of Bibliography of Translation Studies” de 1998 a 2001, organizado por Lynn Bowker, da “School of Translation and Interpretation”¹ (Universidade de Ottawa) registra 499 trabalhos, dos quais apenas dez aparecem na rubrica “corpus-based translation studies”, ou seja, somente 2% do total.

Dados semelhantes constam na lista de trabalhos do CETRA (Centre for Translation and Intercultural Studies da University of Manchester Institute of Technology)², que reúne uma seleção de

trabalhos importantes sobre corpora e tradução. Compreendendo um período de nove anos (de 1993 a 2001), há nela 29 trabalhos, o que corresponde a uma média de pouco mais de três trabalhos por ano. Percebe-se, contudo, um aumento da quantidade de publicações ao longo dos anos: de uma em 1993, para três em 1996, para 13 em 1998³, para nove em 2001.

A lista mostra ainda que esses 29 trabalhos são de responsabilidade (autoria ou organização) de apenas 12 pessoas, o que sugere, ainda, que a pesquisa influente em corpora e tradução esteja pouco disseminada.

A tímida acolhida do uso de corpora pelos pesquisadores em tradução é peculiar, segundo Baker (1999):

given that translation is a pervasive linguistic activity that ought to interest corpus linguists and that corpus linguistics offers translation scholars a powerful set of tools that have already revolutionized the study of language in other spheres. (pp. 281-282).

Mesmo em centros onde a Lingüística de Corpus está altamente desenvolvida, como a Grã-Bretanha, a interface com a tradução ainda é restrita. No Brasil, ainda é recente a tentativa de união dessas duas áreas. Nesse sentido, números especiais como o presente volume desempenham um papel especialmente fecundo.

Obviamente, não é o caso de imaginar que a utilização de corpora na tradução tenha de ser a opção *default* do tradutor ou pesquisador, assim como ela não o é na lingüística. Contudo, essa menor expansão da pesquisa baseada em corpus no âmbito dos estudos tradutológicos em relação à lingüística nos faz pensar quais seriam algumas das causas para que a expansão prevista por McEnery e Wilson ainda não se tenha viabilizado.

O pouco relacionamento entre a Lingüística de Corpus e a tradução deve-se a três possíveis razões principais. Duas delas foram apontadas por Baker (1999). A primeira é o preconceito dos lingüistas de corpus em relação ao texto traduzido, pois o vêem como

um tipo de texto desviante, não representativo da linguagem. Por isso, o texto traduzido não é incluído em corpora de uma dada língua ou variedade. Entretanto, não é o fato de a tradução ser preterida, simplesmente, que acarreta o maior problema, já que, de fato, o texto traduzido é fruto de condições de produção e recepção peculiares⁴, diferentes daquelas em que outros textos são produzidos. O maior problema é justamente o preconceito dos analistas de corpus, que fica patente em diversas manifestações. Por exemplo, Baker cita Sinclair, segundo o qual há uma “inevitable distortion” na tradução. Outro importante lingüista de corpus que demonstra o mesmo tipo de visão negativa é Aarts, que, ao tratar da questão da comparação entre línguas via corpus, considera que “an intrusive factor in such corpora is the translation activity itself” (Aarts, 1998, pp. ix-x, apud Baker, 1999, p. 283).

A segunda razão para a lenta integração entre Lingüística de Corpus e tradução é a imagem negativa da lingüística (em geral) perante os tradutores e pesquisadores da área. Segundo Baker (1999), por muito tempo a pesquisa em tradução foi vista como apenas uma aplicação de teorias provenientes da lingüística, deixando de lado questões sociais e ideológicas inerentes à atividade tradutológica. Assim, as possibilidades oferecidas pela Lingüística de Corpus podem ainda estar sendo encaradas como mais uma manifestação daquela visão arcaica de pesquisa. Entretanto, Baker (1999) ressalta que não se trata da mesma situação: a Lingüística de Corpus oferece um olhar novo sobre a linguagem, que tem o poder de mudar os paradigmas da pesquisa lingüística (Tognini-Bonelli, 2001, 2002).

Essa opinião é compartilhada por Tymoczko (1998), segundo a qual, o uso de corpus eletrônico enriqueceria a pesquisa em tradução, com ênfase na descrição e não na prescrição, podendo:

reengage the theoretical and pragmatic branches of Translation Studies, branches which over and over again tend to disassociate, developing slippage and even gulfs. (p. 7)

É necessária uma mudança de atitude de ambos os lados, tanto do pesquisador em tradução quanto do lingüista de corpus. O primeiro deve tentar enxergar as mudanças operadas no pensamento lingüístico pela Lingüística de Corpus e refletir em até que ponto elas engendram um quadro conceitual e metodológico de valia para a tradução. O segundo deve perceber o valor do texto traduzido como um objeto de pesquisa em si, não como algo inferior ou desviante de uma norma.

A terceira hipótese aqui levantada como tendo influência na demora do estreitamento dos laços entre Lingüística de Corpus e tradução diz respeito a um obstáculo muito claro: o acesso à tecnologia. Por tecnologia, entendem-se dois elementos. Em primeiro lugar, a tecnologia enquanto corpora propriamente ditos, especificamente os de maior interesse para a tradução, como os paralelos e/ou comparáveis. Esses são reconhecidamente mais raros e difíceis de coletar do que os corpora monolíngües. Em segundo lugar, entende-se a tecnologia como os programas de computador para exploração desses corpora específicos para a tradução, como os alinhadores e concordanceadores paralelos, que são menos numerosos, poderosos e, indiscutivelmente, de acesso mais restrito.

Em relação ao primeiro ponto, a dificuldade de acesso a corpora paralelos ou aos de tradução é notória (Bowker, 1998). Os corpora de tradução são de difícil compilação, já que muitos tipos de texto traduzidos são disponíveis somente em papel, o que torna muito custosa a sua transferência para mídia eletrônicas, além dos problemas costumeiros de liberação de direitos autorais que são comuns à coleta de corpus. Bowker (1998) sugere o emprego de corpora monolíngües como alternativa. Outra saída é a captura em massa de textos em formato eletrônico da Web; Berber Sardinha (2002) descreve meios de automatizar e agilizar essa alternativa. No caso de corpora paralelos, há outro nível de complicação que é o alinhamento, necessário para que programas concordanceadores específicos sejam capazes de capturar as instâncias dos termos de busca nas línguas existentes no corpus. Mihailov e Tommola (2001)

propõem uma série de procedimentos com o software de banco de dados Access para automatizar a coleta e organização dos dados. Berber Sardinha (2001b) apresenta maneiras de alinhar corpora usando recursos da Internet. Corness (2002) ilustra meios de fazer o alinhamento com o software Multiconc. Austermühl (2001, pp.134 ss) ensina como o alinhamento pode ser feito com o Translator's Workbench, da Trados.

Em relação ao segundo aspecto da tecnologia, o problema da disponibilidade e aprendizado de manuseio de software é facilmente percebido em *workshops* de exploração de corpus destinados ao público da tradução (como, por exemplo, o levado a cabo no II Encontro Internacional de Tradutores, realizado em Belo Horizonte, cujo tema foi "Translating the New Millennium: Corpora, Cognition and Culture"). Há, invariavelmente, uma grande procura por esse tipo de evento, e as intervenções do público deixam claro que há uma demanda reprimida por ajuda especializada.

Em suma, o pesquisador ou tradutor que deseje fazer incursões na exploração de corpora para a investigação da tradução enfrentará o problema da maior escassez de recursos para sua área, da necessidade de aprender a utilizar software especializados, além de necessitar executar algumas tarefas comuns da Linguística de Corpus, como a organização, formatação e exploração de corpus.

3. Algumas contribuições dos Estudos da Tradução com corpora

Mesmo dadas as dificuldades apontadas na seção anterior, a área de Estudos da Tradução com corpora (Corpus-Based Translation Studies, ou CTS) tem feito muitas contribuições para o entendimento dos processos envolvidos na tradução. Será feito, nesta seção, um apanhado de alguns estudos importantes, sem a intenção de esgotar o assunto.

A quantidade de trabalhos que se vê na área de Estudos da Tradução com corpora, embora aquém do que se esperava, reflete o potencial da pesquisa com corpora eletrônicos. O recém-lançado volume de Laviosa (2002) é a principal resenha do campo dos estudos da tradução com corpus. O manuscrito apresenta um panorama detalhado da área, inserindo os estudos baseados em corpora dentro de uma perspectiva maior dos estudos da tradução. Recomenda-se que o leitor que deseje ter uma visão global da abrangência e da profundidade dos estudos de tradução com corpora consulte esse volume, além de Laviosa, 1998a, para uma visão mais sucinta. Outros volumes a sair, na mesma linha, são Olohan (no prelo), Bowker e Pearson (2002) e Somers (no prelo). Em termos de coletânea, a mais importante da área até o momento é o volume especial de 1998 da revista *Meta* (Laviosa, 1998c), que contém artigos de vários especialistas, tanto no tocante a aspectos teóricos quanto aplicados.

Afora as publicações, há encontros especializados de tradução onde são apresentados trabalhos sobre tradução e corpora. Os seminários anuais da TELRI (Trans-European Language Resources Infrastructure), organizados desde 1995, acolhem pesquisa da área. A conferência “Research Models In Translation Studies”, realizada em 2000, no UMIST⁵, reuniu grandes nomes da área de estudos da tradução com corpora. No Brasil, o grande evento da área foi o VIII Encontro Nacional de Tradutores, realizado juntamente com o II Encontro Internacional de Tradutores, em Belo Horizonte, em 2001. O tema do congresso foi “Corpora, Cultura, Cognição”. Houve apresentação de vários trabalhos, tanto por pesquisadores brasileiros quanto estrangeiros na sub-área especialmente dedicada aos Estudos de Corpora. Os Seminários de Corpora, realizados na USP em 1999 e 2001. No âmbito do InPLA (Intercâmbio de Pesquisas em Linguística Aplicada)⁶ realizado anualmente na PUC/SP, em três simpósios dedicados à Linguística de Corpus, a pesquisa com tradução teve destaque, inclusive com um simpósio exclusivo para a tradução e corpus em 2000. No encontro do GEL

(Grupo de Estudos Lingüísticos) de 2002, foram apresentadas sessões de comunicação sobre tradução e corpus eletrônico. Um encontro futuro de destaque é o *Corpus-based translation studies – Research and Applications*, que acontecerá em 2003 em Pretória, na África do Sul⁷.

Resumir toda a pesquisa desenvolvida com corpus eletrônico no campo da tradução fugiria em muito ao escopo deste trabalho. Contudo, para fins de uma sistematização mínima, é possível propor algumas áreas de concentração, entendidas como temas que têm recebido atenção especial por parte de pesquisadores. A primeira grande área que fica evidente é aquela voltada à tecnologia. A área de compilação de corpora bi- ou multilíngüe é obviamente uma área forte neste setor (p.ex. Avance de Souza, 2002; Botley, McEnery, & Wilson, 2000; Frankenberg-Garcia, 2002; Johansson & Oksefjell, 1998; Laviosa, 1997; Tagnin, 2002). Os trabalhos nessa área visam muito mais à coleta e à organização dos dados do que propriamente a sua exploração. Ainda na linha da infraestrutura de pesquisa, para chamar assim, estão os trabalhos que tratam de alinhamento automático de corpora paralelos (Hofland & Johansson, 1998; Melamed, 2001; Veronis, 2000). Esta é ainda uma área muito especializada, que exige conhecimentos de informática e estatística muito superiores ao que a maioria dos tradutores e pesquisadores de linguagem possuem. Daí ser um campo restrito, cujos resultados, em termos práticos, ainda são preliminares (não há nenhum software amigável e amplamente disponível, para o ambiente Windows, por exemplo, que faça o alinhamento de corpora com sucesso).

Outro assunto de interesse no campo da informática é a criação e uso de concordanceadores paralelos e outras ferramentas para lidar com corpora paralelos e alinhados (Austermühl, 2001; Barlow, 1995; Bowker, 2002; Ebeling, 1998; Santos & Oksefjell, 2000). Dentro da área tecnológica, há ainda os trabalhos que tratam da questão da tradução automática, que têm usado corpora paralelos para adquirir automaticamente material para informar modelos de “sistemas de tradução com memória” ou “baseados em exemplo”

(“memory-based machine translation” (MBMT) ou “example-based machine translation” (EBMT)”); p.ex. Guidère, 2002). Por fim, os trabalhos que visam ao ensino do uso de programas de computador ou de sistemas *online* (Austermühl, 2001; Berber Sardinha, 2001) é área que tem recebido menor interesse na literatura, o que é lamentável, visto que isso pode ser uma das razões que tem dificultado a ampliação do uso de corpora na tradução.

O outro setor em que se pode agrupar a literatura da área é o da exploração de corpora na pesquisa e formação de pessoal. Os trabalhos que tratam da análise das escolhas lingüísticas são os mais numerosos. A pesquisa aqui encampa escopos variados, tais como o julgamento das escolhas no nível da palavra (Avance de Souza, 2002; Berber Sardinha, 1997; Kenny, 2001; Santos, 1999), no nível do discurso (Ghadessy & Gao, 2001; Hasselgard, 1998; Lopes, 2000; Siqueira, 2000), bem como a reavaliação da questão específica da equivalência (Partington, 1998; Teubert, Tognini-Bonelli, & Volz, 1998). Outra área importante é a do estudo do processo tradutório de tradutores profissionais (Aston, 1999; Baker, 1999), com vistas à melhoria da formação dos aprendizes (Bowker, 2001; Uzar & Walinski, 2001) e treinamento de tradutores (Bowker, 1999; Gavioli & Zanettin, 1997; Tagnin, 2000)⁸.

Os dois setores citados acima não atuam em separado, mas interagem de modo profícuo. Os aportes fornecidos pelos avanços no campo da tecnologia influenciam a disponibilidade de recursos para a pesquisa na linguagem da tradução propriamente dita. De seu lado, os direcionamentos na pesquisa com a linguagem sugerem os caminhos de interesse que a tecnologia deve trilhar, que, por sua vez, ao concretizar os instrumentos, enriquece a pesquisa, num ciclo virtuoso. A ponte entre os dois setores, contudo, ainda é frágil, conforme dito acima, já que os trabalhos escritos ou presenciais que tratam do treinamento no uso de ferramentas computacionais ainda são escassos.

4. A influência de Mona Baker na pesquisa em tradução com corpora

O grande impulso inicial na pesquisa em tradução com corpora foi dado por Mona Baker, em várias publicações (Baker, 1995, 1996, 1998, 1999). Em seu trabalho em homenagem a John Sinclair (Baker, 1993), ela estabelece os alicerces da exploração de corpora para fins tradutológicos. Num outro trabalho (Baker, 1995), ao eleger quatro categorias (simplificação, explicitação, normalização e estabilização), ou “universais”, como relevantes para a pesquisa em tradução, ela estabelece um programa de pesquisa que seria seguido por outros investigadores. Esses universais são definidos como “características que tipicamente ocorrem em textos traduzidos ... e que não são resultado da interferência de sistemas lingüísticos específicos” (“features that typically occur in translated texts ... and which are not the result of interference from specific linguistic systems”, Baker, 1993, p. 243). Mais especificamente, essas categorias são:

- Simplificação (“simplification”): a linguagem usada nas traduções tende a ser mais simples do que a dos originais, possivelmente como tentativa de facilitar a leitura da tradução. Seria possível investigar esse princípio, por computador, comparando estatísticas advindas de corpora relativas à variação lexical (como a razão forma-ocorrência, ou “type-token ratio”) e ao tamanho da frase.
- Explicitação (“explicitation”): a linguagem usada nas traduções tende a explicitar a informação, até mesmo onde o original deixa aspectos implícitos. Uma possível medida disso, em corpora, seria a estatística de tamanho do texto. Os textos traduzidos seriam mais longos, por conta da maior quantidade de palavras possivelmente necessárias para tornar uma informação mais explícita.

- Normalização (“normalization”): a linguagem das traduções tende a utilizar em excesso algumas das características mais comuns da língua-alvo, minimizando os aspectos criativos ou menos comuns da língua-fonte. Um exame das escolhas lexicais em textos originais e em suas respectivas traduções pode revelar a normalização se indicar, por exemplo, que as escolhas mais ‘marcadas’ (ou criativas) dos originais tiverem sido traduzidas por outras menos marcadas.
- Estabilização (“levelling out”): Textos de um corpus de traduções tendem a ser mais semelhante entre si, no tocante a diversos aspectos lingüísticos, do que textos de um corpus de originais. Uma evidência disso, num corpus, poderia ser obtida por meio da verificação das médias, por exemplo, da razão forma-ocorrência (“type-token ratio”), para um corpus de tradução e para seu correspondente na língua-fonte. A hipótese prevê que os textos traduzidos teriam médias dessa razão mais parecidas entre si do que os textos originais.

Mona Baker é diretora do Centre for Translation and Intercultural Studies, da UMIST, que tem se destacado como centro de referência para estudos de tradução com corpora. O CTIS tem formado pesquisadores atuantes (como Sara Laviosa, Dorothy Kenny, Maeve Olohan) que têm contribuído para a aplicação da agenda de pesquisa de Baker (como, por exemplo, Kenny, 1997 e Laviosa, 2001). A influência de Mona Baker, contudo, é visível na produção advinda de outros centros (p.ex. Scott, 1998), inclusive do Brasil (Magalhães, 2001; Magalhães e Pagano, neste número).

Ao mesmo tempo em que essas hipóteses são acolhidas, começam a aparecer críticas aos princípios colocados por Baker (1995). Steiner (2001), por exemplo, acredita que haja uma distância muito grande entre o tipo de dado que é exigido para estudar esses princípios e o tipo de dado normalmente fornecido por um corpus eletrônico:

The linguistic phenomena in terms of which the hypotheses have been formulated are very low-level and in some cases questionable. Counting words with or without lemmatization, words per sentence, percentages of form words vs. function words, [...] and a few other phenomena of this order of concreteness is methodologically too far removed from the level at which one would want to formulate hypotheses about properties of texts. No relevant model of textuality is formulated anywhere near this low level of representation. (p. 8)

Em decorrência disso, seria um erro, para o autor, comparar diretamente as estatísticas relativas à quantidade de palavras em textos escritos em inglês e em alemão, visto que no alemão a morfologia (declinação etc.) e os compostos nominais, por exemplo, são muito diferentes, o que afeta diretamente a contagem de ocorrências (*tokens*) e formas (*types*). Por exemplo, no alemão usa-se uma palavra para designar a expressão inglesa “taxi driver” (“Taxifahrer”); além disso, outras palavras (artigos, pronomes) podem ser grafadas de vários modos dependendo do caso que assume na frase. Ele propõe, como alternativa, que as análises de corpora entre originais e traduções sejam feitas usando etiquetagem e lematização, quer de modo automático ou manual.

Hansen e Teich (2001) também criticam os universais propostos por Baker, na mesma linha de Steiner, mas notam, ainda, que é preciso incorporar a eles a possibilidade de interferência da L1 no resultado da tradução (a lei de interferência de Toury, 1995). Isso as leva a propor um outro princípio para os textos traduzidos, mais especificamente, o de “visibilidade” (“shining through”), ou seja, o de que “in translations, the source language tends to shine through” (p. 3). De certo modo, a visibilidade é o oposto da normalização. Enquanto a normalização prevê a influência em excesso da língua-alvo, exacerbando suas características, em detrimento da criatividade do original, a visibilidade acontece quando houver influência marcante da língua-fonte, ou seja, quando características da língua do texto original ainda permane-

cerem em excesso no texto traduzido, comprometendo a sua “naturalidade”.

Para viabilizar o tipo de pesquisa proposto, Hansen e Teich (2001) propõem a incorporação, ao desenho de pesquisa, de um corpus comparável de textos originais, na mesma língua dos textos traduzidos. Esse terceiro corpus funciona como um termo de comparação, para permitir que se verifique a hipótese de interferência. Para ilustrar, podemos pensar no caso dos gerúndios (p.ex. “opening a new document”, “printing a file”, etc.) no inglês. Em português, o uso desse tipo de gerúndio é condenado. Assim, se em textos traduzidos do inglês para o português houver muitos gerúndios, surge a suspeita de “visibilidade”. Para investigar essa questão, é preciso tabular a quantidade de uso de gerúndios entre os textos originais em inglês. Supondo que os resultados indiquem que eles são abundantes, então repete-se o processo com um corpus de português de textos originais, comparáveis. Se os resultados indicarem uma baixa frequência de gerúndios em português, então a hipótese de visibilidade estaria confirmada, ou seja, a tradução “teria sofrido interferência” do inglês.

5. Revisitando a equivalência

Uma conseqüência da difusão do uso de corpora nas diversas esferas citadas acima é a revisão de conceitos estabelecidos. Isso é típico da exploração de corpora, tendo ocorrido também na lingüística propriamente dita por meio da Lingüística de Corpus. Um dos conceitos mais tradicionais do campo da tradução é o de equivalência. Embora esse conceito já tenha sido questionado (p.ex. Hatim & Mason, 1990, que preferem falar em termos de “adequação” do que de “equivalência”) e revisto (Halliday, 1992 fala em equivalência como sendo “equivalence of function in context”), há um grande interesse nesse assunto do ponto de vista da pesquisa com corpora (p.ex. Teubert, Tognini-Bonelli, & Volz, 1998), a qual, em geral, oferece evidências concretas da sua imprecisão.

Entre os estudos disponíveis centrados em corpora que revisitam o conceito de equivalência está o de Partington (1998). Ele observa que a palavra “approve” é apresentada como o equivalente mais próximo de “sanzionare” em cinco dos principais dicionários bilíngües de italiano-inglês. Ao observar concordâncias, entretanto, ele percebeu que o sentido predominante de “sanzionare”, num corpus de italiano, é o de “penalizar” (“penalize”), ou seja, o contrário do que os dicionários indicavam. Partington (1998) levanta a hipótese de que “sanzionare” deve estar sofrendo uma mudança radical de sentido por conta da influência da mídia, onde se vêem freqüentes referências a “sanções” de todos os tipos (econômicas, políticas, etc.).

Salkie (2002), a partir de seus achados em um corpus paralelo de alemão e inglês, sugere que a equivalência seja repensada segundo alguns tipos. O primeiro tipo é a equivalência que é “tradutoriamente ambígua” (“translationally ambiguous”). Um exemplo seria a palavra alemã “kaum”, cujo equivalente imediato, segundo o dicionário, é “hardly”. Segundo o corpus, embora a tradução por “hardly” seja minoritária (26%, em registros de não-ficção), as demais se agrupam em poucos outros conjuntos, entre os quais os de “negative expression”, “little” e “almost + negative” são os mais freqüentes. O segundo tipo é a equivalência “tradutoriamente vaga” (“translationally vague”), que tem como exemplo o verbo inglês “contain”: ele aparece num corpus paralelo de francês traduzido como “contenir” (58% dos casos). As demais opções existentes no corpus encampam muitas outras palavras, todas de baixa freqüência. A relação entre “contain” e “contenir” é vista como vaga porque há muitos outros equivalentes possíveis.

Diante da variação existente na relação entre itens de duas línguas por meio da tradução, Salkie (2002) sugere que seja empregado o conceito de “modulação” para servir de critério para a determinação de quais correspondências são equivalências e quais não são (p. 60). O conceito de modulação é entendido, segundo o autor, como “maneiras diferentes de ver a mesma situação” (p. 57).

Embora seja um conceito reconhecidamente vago, para o autor ele é útil porque situa a questão da equivalência também no plano textual, e não somente no plano abstrato do sistema lingüístico, já que se admite que as modulações variem de acordo com o texto, o contexto, o ponto de vista do tradutor, etc. Operacionalmente, para o autor são modulações aqueles itens atestados em um corpus que, mesmo apresentando disparidade (não serem “semantically equivalent”, p. 59), possam ser encarados como resultado de o tradutor ter usado uma maneira diferente de entender a situação-fonte. Desse modo, traduções consideradas dentro do que seria admissível diante da modulação são aceitas, enquanto “traduções em que o texto-fonte e o texto-alvo diverjam mais radicalmente do que a modulação” (“translations where the source text and target text diverge more radically than modulation”) seriam consideradas casos de não-equivalência. Como exemplo de modulação, o autor oferece “contain” e “reposer” (“wicker baskets containing the dough” e “panetons d’osier où reposait la patê”).

Correspondência mútua (ou traduzibilidade, “*translatability*”) é um conceito introduzido por Altenberg (1999) para estimar o grau de equivalência entre itens de línguas diferentes. Ele funciona por meio da verificação da quantidade de equivalências observadas. No caso da tradução, são necessários dois corpora paralelos: um formado por uma língua A como partida e uma língua B como chegada, e outro formado pela língua A como chegada e a língua B como partida. Por exemplo, para estimar a correspondência mútua de “give” e “dar”, seria preciso um corpus de textos originais de português com suas respectivas versões para o inglês e outro de textos em inglês original com suas traduções para o português. Seria observado, então, o grau de correspondência entre as traduções atestadas de “give” (tendo o inglês como língua partida) para “dar” e de “dar” (tendo o português como língua de partida) para “give”.

Os estudos com corpora sobre equivalência apontam, em geral, para um entendimento da equivalência como correspondência condicionada ao contexto, o que vai ao encontro da visão de outros teóri-

cos. Para Chesterman (1998, p. 31), equivalência é “equivalence in context”. Halliday (1992), nesse sentido, lembra que o que há são equivalentes em potencial, que por sua vez estão “contextualmente condicionados”, de tal forma que a escolha de um ou outro, na tradução, implicará a ativação de sentidos específicos oriundos da presença costumeira desse item num contexto maior.

6. O papel especial da intuição nos Estudos de Tradução com corpora

Ao mesmo tempo em que a pesquisa em tradução com corpus discute o conceito de equivalência, ela também faz repensar o papel da intuição. Vale lembrar que a intuição é um dos aspectos do conhecimento lingüístico mais criticados pela Lingüística de Corpus.

Quando se trata de corpora de textos traduzidos, a relação entre evidência e intuição assume contornos específicos, que merecem consideração especial. Um corpus de tradução é, antes de mais nada, uma coletânea de textos escritos em larga medida a partir da intuição e da experiência do tradutor, numa situação especial de produção (“a process of text-induced text production”, segundo Hansen & Teich, 2001, p. 2). As escolhas feitas pelo tradutor para passar o texto da língua-fonte para a traduzida foram validadas por ele tendo como critério, via de regra, sua experiência prévia como usuário da língua-alvo. Mesmo nos casos em que ele tenha consultado dicionários e outras traduções (suas ou de outros), ou mesmo corpora, a sua intuição foi um elemento-chave na realização da tradução do texto, pois sem ela ele provavelmente não saberia nem mesmo por onde começar a fazer a consulta ao dicionário ou ao corpus. Desse modo, pode-se dizer que um corpus traduzido é essencialmente um registro das escolhas lingüísticas, muitas delas conscientes, feitas a partir da intuição dos tradutores responsáveis pelas traduções contidas no corpus:

the use of a translation corpus ... will ... provide us with a set of possible translation pairs that have already been identified and used by translators, in other words, verified by actual translation usage. (Tognini-Bonelli 2002, p. 81).

Ao contrário do que possa parecer, na pesquisa com corpus voltada à questão da equivalência, a intuição do pesquisador ou do tradutor possui um peso considerável. É ela que dá o pontapé inicial na pesquisa:

the initial hypothesis positing one or more tentative matches between two or more *prima facie* units of meaning in SL and TL has to rely on the translator's intuition or past experience. (Tognini-Bonelli, 2001, p. 134)

Para Altenberg e Granger (2002), a equivalência, ao envolver a competência do tradutor, passa também a incorporar a intuição :

ultimately, the notion of equivalence is a matter of judgment, reflecting either the researcher's or the translator's bilingual competence. (p. 18).

O papel da análise de corpus nesse processo não é o de confirmar ou desautorizar a intuição do analista, calcada em sua competência, mas sim o de refinar "initial assumptions of similarity" (Altenberg & Granger, 2002, p. 16).

Apesar de a intuição ser um ponto de partida importante, o corpus proporciona os elementos vitais necessários para que a pesquisa desvende aspectos não contemplados nos pressupostos iniciais. Ao observar a evidência exibida pelo corpus, o analista pode evitar o risco da circularidade na pesquisa (Altenberg & Granger, 2002, p. 17), que acontece quando os resultados da análise não vão muito além da intuição inicial (Krzyszowski, 1990, p. 20). O corpus pode

enriquecer a pesquisa, fornecendo os vários elementos que Chesterman (1998) julga essenciais na pesquisa contrastiva:

the analysis has added explicitness, precision, perhaps formalization, (...) added information, added insights, added perception. (p.58)

7. Metodologias para análise de corpora na área da tradução

Ao tomar a posição de não discriminar a intuição como componente da pesquisa com corpus, a tradução busca desenvolver metodologias de análise de corpus próprias, que tendem a distanciá-la da *Linguística de Corpus*. Além disso, à medida que o uso de corpora se torna mais comum na pesquisa em tradução, começam a surgir trabalhos que propõem metodologias específicas de exploração de corpora para fins tradutológicos (p.ex. Bowker, 2001; Kenny, 1998; Scott & Scott, 2000; Tognini-Bonelli, 2001).

Uma dessas metodologias é proposta por Tognini-Bonelli (2001, 2002), que define uma seqüência de procedimentos e uma unidade de análise específicas para a exploração de corpora bi- ou multilíngües. Vista de modo amplo, a sua proposta se aplica tanto aos estudos tradutológicos quanto à *Linguística Contrastiva*, entretanto ela destaca a sua pertinência para a tradução.

Antes de apresentar a metodologia, é importante ressaltar que ela se insere num contexto maior em que a autora faz uma série de propostas que visam ao reconhecimento da importância da pesquisa com corpora em geral. Em primeiro lugar, a autora distingue dois tipos de abordagem na pesquisa linguística com corpora eletrônicos. A primeira, e a mais comum, segundo ela, pode ser chamada de “baseada em corpus” (“corpus-based”). Nessa abordagem, o papel do corpus é o de ser um depósito de exemplos para ilustrar uma teoria ou conceitos previamente estabelecidos. Por exemplo, a distinção clássica entre léxico, de um lado, e sintaxe,

de outro, que é amplamente questionada pela Lingüística de Corpus, não o é normalmente em estudos que seguem a corrente “baseada em corpus”. Mesmo quando os exemplos chegam a incluir elementos que possam questionar os preceitos seguidos na pesquisa, esse questionamento normalmente não acontece. A autora completa alertando:

This approach does not allow for the fact that the enormous amount of evidence now available is bound to challenge language description and offer fascinating new insights into language (...) To start, therefore, with units derived from traditional descriptions, often based on very little evidence, is not only not sufficient anymore, it is dangerous. (Tognini-Bonelli, 2002, p.74)

O outro tipo de abordagem é o “movido a corpus” (“corpus-driven”). A metodologia envolvida nesse tipo de pesquisa visa à descrição abrangente dos dados, sem a intenção de selecionar exemplos para ilustrar elementos oriundos de uma teoria específica. Em primeiro lugar vem a evidência fornecida pelo corpus:

The theoretical statements, as well as the comments or recommendations made, arise directly from, and reflect, the evidence provided by the corpus. (...) Linguistic description is arrived at, step by step, from the observation of language usage; recurrent language events and frequency distributions are expected to form the basis of linguistic categories. (Tognini-Bonelli, 2002, p.75).

Há ainda o que se costuma chamar de abordagem “presa ao corpus” (“corpus-bound”), em que a pesquisa fica limitada ao que o corpus informa, sem levar em conta o conhecimento prévio do pesquisador, seu direcionamento na análise dos dados, sua interpretação, etc (cf. Bowker, 2001).

A segunda colocação importante de caráter mais amplo feita por Tognini-Bonelli (2002) diz respeito à definição de uma unidade de análise (“currency”) para a análise de corpora: “a unidade funcionalmente completa” (“functionally complete unit”), que é uma unidade de sentido, de caráter fraseológico, a respeito da qual há evidências no corpus. Essa unidade não possui delimitação prévia, podendo ser composta de duas ou mais palavras, com posições fixas ou não. Pode até mesmo acontecer de esta unidade não ter realização (“translation to zero”), ao se considerar a correspondência entre duas línguas.

A metodologia sugerida por Tognini-Bonelli (2001) é composta de três etapas:

- (1) Identificação da padronização formal de L1 e das funções correspondentes. Por função entende-se o resultado da descrição dos padrões de um item de interesse na L1 segundo seu ambiente colocacional, coligacional, de preferência e de prosódia semânticas (vide abaixo).
- (2) Identificação de equivalentes tradutórios *prima facie* para cada função. Nessa etapa, o analista propõe um ou mais possíveis equivalentes para cada padrão (função) encontrado na etapa anterior. Os candidatos podem ser propostos a partir da intuição, fazendo uso da competência tradutória do analista, ou mesmo da consulta a dicionários, mas a melhor opção, segundo a autora, é buscar os candidatos em um corpus paralelo. Ela mesma reconhece que a disponibilidade de corpora paralelos é restrita e que, portanto, essa etapa pode ser executada sem recurso a um corpus.
- (3) Identificação da padronização formal de L2 e das funções correspondentes. Essa etapa é semelhante à primeira, com exceção do fato de aqui a descrição ser feita em um corpus da segunda língua presente na pesquisa.

A autora exemplifica a sua metodologia com o estudo de *in the case of* e seu correspondente no italiano. A primeira etapa consistiu na análise das ocorrências dessa locução prepositiva num corpus de inglês nativo. A análise mostrou que a locução parece desempenhar uma função de especificidade, já que ela é seguida costumeiramente de um artigo definido. Além disso, a análise indicou a presença de prosódia semântica neutra (vide abaixo). Nesse ponto, a autora já tem definida uma “unidade de sentido funcionalmente completa” (Tognini-Bonelli, 2002, p.85), que comporta tanto a padronização típica (*in the case of + the*), quanto à preferência semântico-pragmática da expressão (neutralidade). Na segunda etapa, a autora propôs como equivalente *prima facie* no italiano a expressão *nel caso di*. A autora se baseou, para chegar a esse equivalente inicial, na sua intuição e experiência de falante bilíngüe. Na terceira etapa, foi feita a análise de *nel caso di*, a partir da exploração de um corpus de italiano. Os resultados indicaram, assim como no inglês, a presença do artigo definido em posição próxima (... *degli, dei, del, etc*), o que significa o desempenho de uma função de especificação. A análise no nível da preferência semântica mostrou que a locução apresenta seleção pela tecnicidade e por vocabulário acadêmico-literário, e no tocante à prosódia semântica a expressão parece ser neutra. A análise concluiu, então, que o ponto em comum mais forte entre as duas locuções, nas duas línguas, é a expressão de especificidade e a neutralidade prosódica; além disso, não há discrepância no nível da preferência semântica, já que a tecnicidade observada no italiano pode ser fruto da composição diferente entre os dois corpora utilizados.

Outra metodologia específica para os estudos contrastivos de tradução com corpora foi proposta por Kenny (1998). A sua metodologia consiste de três corpora: um paralelo, com textos numa língua de origem e suas respectivas traduções numa língua-alvo, um corpus de referência com textos originais na língua de origem, e outro corpus de referência com textos originais na língua de chegada. Ela sugere que seja feito um mapeamento dos padrões rela-

tivos aos itens de interesse nos dois componentes do corpus paralelo. Entretanto, a autora insiste que seja feita, em seguida, a extração dos padrões dos mesmos itens em cada um dos corpora de referência, para “ascertain their (un)conventionality” (Kenny, 1998). Isso é necessário para que seja possível perceber se os usos observados no corpus paralelo eram típicos ou não. Por exemplo, no caso da palavra “giro”, estudada por ela, primeiramente foi feita a busca das suas ocorrências no componente inglês do corpus paralelo, depois do seu equivalente no componente alemão, seguida da busca de cada item em cada um dos corpora de referência. Os resultados permitiram saber até que ponto os padrões nos textos originais e traduzidos refletiam o uso típico da língua, o que a autorizou a julgar a tradução com mais autoridade, baseando-se na evidência exibida pelo corpus.

8. Prosódia Semântica na tradução

Há um tipo de unidade funcionalmente completa que desafia a questão da equivalência, por refletir um tipo de sentido sutil, mas importante, cuja realização fica particularmente aparente quando é investigada por meio de corpora eletrônicos.

Trata-se da prosódia semântica que tem merecido a atenção de pesquisadores na Linguística de Corpus, com corpora monolíngües (Hoey, 2000; Hunston, 2000; Louw, 1993; Stubbs, 2001). A prosódia semântica é um tipo de padrão que indica um sentido avaliativo ou pragmático, que um trecho do texto assume por meio da presença e co-ocorrência de certos itens. Um exemplo clássico é “set in”, que no inglês possui uma prosódia desfavorável, já que nos trechos onde aparece, vem tipicamente acompanhado de itens negativos, como “decay”, “rot”, “infection”, etc (Sinclair, 1987, p.155-6). Na tradução e em estudos contrastivos, ela tem sido enfocada em alguns trabalhos (Berber Sardinha, 2000a, b; Kenny, 1998).

O conceito de prosódia semântica deriva do entendimento de que o sentido de um item lexical abrange uma extensão de texto mais ampla, “tingindo” o contexto ao seu redor. O estudo da prosódia semântica, assim, não se restringe somente à descrição da palavra em questão, mas também ao contexto maior no qual a palavra em questão está inserida. Em primeiro lugar, esse contexto compreende seu padrão de co-ocorrência; e, em segundo lugar, o tipo de sentido que emana da co-ocorrência e atinge também um trecho maior do seu contexto local, podendo compreender a oração, o período, o parágrafo e até mesmo o texto como um todo. Isso se deve ao fato de que, como já mostraram Hunston e Francis (2000), os padrões não ocorrem sozinhos, mas interligados em seqüências que exibem uma harmonia de sentido, propósito ou posicionamento avaliativo. Além disso, segundo Partington (1998, p. 51), os sentidos das escolhas refletem-se nos diversos níveis da língua: as escolhas no nível da palavra dependem das escolhas no nível do grupo, e vice-versa, que por sua vez dependem das escolhas no nível do período, e vice-versa, num movimento contínuo que chega até os níveis mais altos de organização, como o registro, o gênero e o contexto cultural. Assim, os estudos de prosódia semântica podem ser vistos não somente como um estudo de equivalências no nível da palavra, mas também dos sentidos que emanam dessas escolhas, potencialmente, nos vários níveis da língua.

Um estudo que enfocou a prosódia semântica é Berber Sardinha (2002), que comparou as opções oferecidas por um dicionário bilíngüe para a tradução de “set in” para o português. As buscas feitas num corpus geral e bastante extenso de português indicaram que nenhum desses itens captura a prosódia semântica negativa inerente a “set in”. A primeira opção, “manifestar-se”, possui uma prosódia semântica dupla, sendo negativa somente quando se coloca com itens como “doença”; contudo, o equivalente de “doença”, “sickness” ou “illness”, não são colocados típicos de “set in”. Outra opção presente no dicionário, “estabelecer-se”, demonstrou possuir prosódia neutra. Uma terceira opção, “cair”, somente se aplicaria quando se

tratar da ocorrência de “night” com “set in”. Contudo, a colocação resultante no português, que seria “ao cair da noite” ou “quando a noite cai”, possui, na verdade, uma prosódia semântica predominantemente positiva. Finalmente, o verbo “entrar”, que também é citado no dicionário, não formou colocações com nenhum colocado equivalente de “set in”. Por tudo isso, o autor concluiu que as opções indicadas pelo dicionário são enganosas, por se referirem a colocações que (a) não retêm a prosódia semântica desfavorável do inglês, ou que (b) o dicionário oferece opções de verbos que não formam colocações em português com as palavras em questão.

Kenny (1998), em seu estudo sobre a palavra “giro”, também investigou sua prosódia num corpus paralelo formado por textos originais em inglês e suas traduções para o alemão. Em inglês, “giro” assume um sentido específico de pagamento recebido por desempregados ou aposentados. Ao examinar uma concordância de “giro” num corpus monolíngüe de referência do inglês, a autora notou que esta palavra possui uma prosódia semântica desfavorável, com conotação de pobreza e de injustiça social. Porém, ao inspecionar a tradução desse item num corpus paralelo com as traduções para o alemão, percebeu que “giro” foi usado com outro sentido. O tradutor (de um texto literário) havia interpretado “giro” como talão de cheque. Dessa forma, o trecho original em inglês, que se referia ao fato de um aposentado ter tido o seu “giro” roubado, foi vertido para o alemão com a informação de que o que havia sido roubado fora um talão de cheques. Não bastasse esse erro de tradução no nível denotativo, ainda acarretou mudança na prosódia semântica entre as duas línguas. Ao verificar as ocorrências de “Scheck” (cheque) e “Scheckheft” (talão de cheque) num corpus de referência com textos originais escritos em alemão, a autora observou que as duas palavras possuem uma prosódia semântica neutra. Dessa forma, o “tom” do original em inglês mudou radicalmente quando da passagem para o alemão. A indignação que o leitor poderia sentir ao ler sobre o roubo do dinheiro da pensão de um velho desempregado ficou atenuado no alemão. Esse fenôme-

no, que Kenny chama de “toning down”, é indicativo, segundo ela, de um processo que pode ser chamado de “higienização” (“sanitisation”), e que ocorreria quando a prosódia semântica no texto-fonte é perdida ou significativamente modificada como resultado da tradução.

8.1 Prosódia semântica de “borderline” e “limiar”

Para ilustrar a análise de prosódia semântica com corpora, no âmbito da tradução, será considerado aqui o item “borderline” e seu equivalente “limiar”, conforme ocorreram num texto em inglês e na sua tradução para o português.

A metodologia empregada aqui é inspirada na proposta de Tognini-Bonelli (2002) discutida acima. São usados corpora monolíngües com a função de corpora de referência para cada uma das línguas. Entretanto, ao contrário de Tognini-Bonelli (2002), não são propostos equivalentes *prima facie* a partir da intuição. Ao contrário, os equivalentes são retirados de um par de textos paralelos (original e versão). A análise indica as unidades de sentido funcionalmente completas, segundo Tognini-Bonelli.

Numa reportagem publicada em 2002 na versão brasileira, em português, do jornal USA Today intitulada “More men just say ‘no’ to working” , aparece o seguinte trecho:

Assim como os Leiser, Bonnell diz que a decisão foi tomada devido a fatores meramente financeiros. A sua mulher ganha mais dinheiro que ele e conta com mais oportunidades para progredir na carreira. Além disso, “ela está no limiar para se tornar uma trabalhadora compulsiva, enquanto eu estou no limiar do vício em ociosidade”, diz Bonnell.

A leitura desse trecho causou “estranhamento”⁹, especialmente no tocante ao item “limiar”. O tradutor usara “limiar” como equivalente de “borderline”, segundo mostra, abaixo, o texto do original:

Like the Leisers, Bonnell says the decision was purely financial. His wife makes more money than he did and has more opportunities for advancement. Besides, 'She's a borderline workaholic, and I was borderline not-workaholic,' he says.

No corpus Banco de Português, que possui 220 milhões de palavras da escrita e fala, há 742 ocorrências de “limiar”. Dessas, nenhuma possui a padronização exibida no trecho em questão, isto é, a coligação “limiar + para < verbo> “ e a colocação “limiar do vício”.

Dessa forma, as escolhas do tradutor foram marcadas, o que deve ser um fator causador da estranheza sentida pelo leitor da versão da reportagem. “Limiar” é normalmente usado em português nos seguintes padrões:

limiar do/da/de	440
limiar entre	26
limiar aeróbico / anaeróbico	22

Os colocados à direita mais frequentes de “limiar” indicam o sentido temporal como predominante:

< século (86), milênio (52), anos/ano (26), era (7), tempo (5), história (4)>

Quando não é usado para se referir a um período de tempo, “limiar” ocorre com substantivos indicativos de situações negativas:

< loucura (7), violência (6), morte (5), dor (4), calúnia (3), rebaixamento (3)>

A opção do tradutor, “limiar do vício”, encaixa-se neste último grupo, embora não seja uma escolha típica.

-
- 4 I need these limits to
keep from working too much.
If you're a > > borderline workaholic, an at-home business might not be the best option for you.
-
- 5 but with her long list of
accomplishments and > > borderline workaholic drive she has managed to raise several businesses from the ashes.
-
- 6 Helton has achieved his
position through hard work and
determination — not to
mention being a > > borderline workaholic who eats, sleeps and lives NASCAR.
-
- 7 I am also a type A, > > borderline workaholic, but I'm TRYING to change this behavior.
-
- 8 But work sets you free,
and for now, Caan, a self-
professed" > > borderline workaholic.
-
- 9 My father was always
busy — it would be easy to
apply cheap pop psychology
and say that he's a > > borderline workaholic.
-
- 10 I am a > > borderline workaholic and I hate being away for a day. I hate being sick and I hate taking time off.
-
- 11 To say that every
member of this side is a > > borderline workaholic would be oversimplifying.
-
- 12 This is a mighty fun
hobby for the > > borderline workaholic.

-
- 13 Under "married?",
Dennis writes, not yet! He is
living in Fair Oaks and
describes himself as a > > borderline workaholic.
-
- 14 He's a > > borderline workaholic, who
hides it by doing much of his work
quietly and in private.
-
- 15 if only briefly, i forget
what it is to be a > > borderline workaholic.
-
- 16 "Not that I know of,
anyway. Just wanted to get an
early start on some things." > > borderline workaholic, he.
-
- 17 College, Tom Ventress
About: > > borderline workaholic, (strike
borderline) with a penchant for
puns would pretty much sum me
up. I work
-
- 18 not really sure I can do
that since I'm a > > borderline workaholic. I have
a wonderful husband whom I
enjoy spending time with.
-
- 19 I'm also a > > borderline workaholic. I still
take tons of supplements and am
on Armour (thyroid) and Cortef
(adrenal).
-
- 20 A > > borderline workaholic, Lay
was on the road working the
nightclub circuit 300 days a year.

pictures,” he moans.² Though now a > workaholic confined within the
 manager Melinda French. [p] > workaholic Gates, 38, is America’s
 briefing the press himself [p] A > workaholic, Gergen rarely spent
 mirror Do you really want to be a > workaholic guilt-plagued mother
 his crew like the conscientious > workaholic he is; gape at Champion
 [p] Passionate, beautiful and > workaholic, her idea of relaxation
 in River Wild (CIC [p] But her > workaholic husband Tom (David
 yes to more than 5, you may be a > workaholic. If so, take some time
 and shot himself through the head. > workaholic Mr Biggs had
 be solved. So you turn into a > workaholic”, see less of your
 musicals with aggressive American > workaholic showbiz monomaniacs
 wonderful places [p] Having been a > workaholic since the age of 15,
 Diane McLellan, is ‘the earnest, Mason-
 After he quit drinking he became a > workaholic son of a Greek priest”.
 Eisen knew for a fact to be a real He is,
 referred to sometimes as a modest > workaholic, spending as little time
 to win,” says Pipe. [p] This > workaholic), sun-browned and
 who discovered that the eminent obviously
 I’m single, but they know I’m a > workaholic, the Chancellor
 > workaholic trainer, who is a bundle
 > workaholic was also secretly a
 > workaholic who gets results.” [p]
 Another

favorite dessert.” If you’re a > workaholic, you may be asked to work through

Em resumo, as escolhas constantes na versão em português parecem acentuar não apenas a prosódia semântica negativa, mas também a impessoalidade. No inglês, a prosódia semântica que tende a emanar do texto é a de interpessoalidade, com menor carga de negatividade. As evidências dos corpora em inglês indicam que as escolhas feitas pelo autor da reportagem em inglês estão compatíveis com a padronização típica de “borderline workaholic”, já que, no texto, a expressão ocorre numa fala em primeira pessoa que se refere a uma situação familiar, inclusive com menção a si próprio (“I was borderline not-workaholic”). Entretanto, na versão em português, o uso de “limiar” mostrou-se fraseologicamente pouco convencional.

Este tipo de análise ilustra o potencial do uso de corpora na avaliação da qualidade da tradução, na linha do que Bowker (2001) propõe. O corpus foi aqui usado como base para o julgamento, sem substituir a participação do avaliador, que, no caso, é o pesquisador:

A corpus should not be seen as a replacement for competence and critical judgement on the part of evaluators, but rather as an aid to help them make sound and objective judgements. (Bowker, 2001, p.361)

A proposta de Bowker inclui a criação de um complexo “corpus de avaliação” (“evaluation corpus”). Entretanto, o procedimento aqui discutido permite a avaliação das escolhas fraseológicas de um texto traduzido, sem necessidade da criação de corpora específicos para esse fim.

9. Comentários finais

Neste trabalho foi feita uma reflexão acerca do papel do uso de corpora eletrônicos nos estudos tradutológicos. Foi observado que, embora haja um grande potencial para a exploração de corpora no campo da tradução, a pesquisa realizada ainda é relativamente pouca. Enquanto a presença de corpora nas “outras” áreas da pesquisa e estudo em linguagem cresceu a passos largos, o mesmo não pode ser dito em relação à tradução: embora tenha havido um aumento da pesquisa e aplicação de corpora na tradução, a maioria dos trabalhos disponíveis na literatura e em encontros especializados da área de tradução não utiliza corpora. Há, contudo, um desenvolvimento gradual, constatado numa crescente produção e na ampliação do grupo de pesquisadores atuantes na área. Em níveis mundiais, a pesquisa em tradução com corpora ainda está aquém do que poderia ser. No Brasil, o mesmo acontece, embora haja sinais de crescimento visíveis na oferta de eventos especializados.

A possibilidade de utilização de corpora na pesquisa em tradução já causou a preocupação de estudiosos da área. Segundo Tymoczko (1998), a adoção de corpora eletrônicos pode trazer consigo uma crença no objetivismo da descrição lingüística, no cientificismo, numa condenação do elemento subjetivo. Ela alerta ainda que a quantificação sem propósito, aquela que visa a “provar o óbvio” e que leva a descrições estereis, deve ser evitada. Basil Hatim (apud Laviosa, 2000) também faz um comentário semelhante, mesmo reconhecendo o valor da pesquisa com corpus. Ele faz uma distinção entre o que está *no* texto (“in the text”) e o que é *do* texto (“of the text”), enfatizando que a pesquisa com corpus não deve se restringir aos aspectos de superfície, à manifestação lingüística, mas deve englobar também o nível textual, que inclui vários aspectos, entre eles a constituição discursiva, questões culturais, a ideologia, etc.

Aston (1999), ao comentar o uso de corpora para o treinamento de tradutores, adverte que:

There is as yet little hard empirical evidence to demonstrate the effectiveness of corpora as translation and as learning tools. While learners seem to a large extent enthusiastic about using corpora, it remains to be shown just in what respects, and under what conditions, their performance as translators may improve as a consequence: we cannot for instance exclude the idea that training with corpora may also improve dictionary usage, by instilling greater attention to collocation and register. No research that I am aware of has yet attempted to compare the effectiveness of different types of corpora, or of different learner approaches to them. (p. 313)

Embora sempre haja o risco de que um recurso novo, tecnológico, como é o corpus eletrônico, venha a ser mal empregado ou pouco compreendido, o que se observa, conforme foi relatado neste artigo, não justifica as preocupações de Tymoczko (1998). A pesquisa com corpora tem revelado aspectos muito relevantes da constituição textual do texto traduzido, bem como de sua relação com o texto fonte, além de auxiliar na compreensão mais direta de vários aspectos do processo tradutório, inclusive os culturais. Conforme enfatiza Kenny (1998):

Whereas in the past so-called “linguistic” approaches to translation have been criticised for their inability to say anything about the wider cultural context in which translation occurs, I would like to suggest here that a careful study of collocational patterns in translated text can shed light on the cultural forces at play in the literary marketplace, and vice versa. Culture and language are inextricably bound up in one another, and it makes no sense to suggest that cultural and linguistic approaches to the study of translation can be mutually exclusive. (p. 519)

Até mesmo a questão da subjetividade, outra preocupação de Tymoczko (1998), tem sido trabalhada sem antagonismo, longe da

falsa dualidade “objetivo versus subjetivo”. Conforme exposto neste trabalho, a intuição do pesquisador e do tradutor assumem um estatuto diferenciado no campo dos estudos tradutológicos, em contraste ao que assumem na Lingüística de Corpus, sua área parente.

Por isso tudo espera-se que logo a pesquisa em tradução venha a se valer de forma mais ampla/intensa da exploração de corpora eletrônicos para que possamos testemunhar a concretização das previsões de dez anos atrás.

Notas

1. <http://aix1.uottawa.ca/~lbowker/bibtsweb/01cuinca.html>
2. <http://www.umist.ac.uk/ctis/staff/mona.htm#cetra>
3. Na verdade, os números desse ano incluem artigos do volume especial da Revista Meta, além de uma entrada separada para a própria revista (Laviosa, 1998c), o que causa duplicidade.
4. Segundo Baker (1999, p. 282), essas condições incluem “the fact that it [the translated text] is constrained by a fully articulated text in another language and that translations have a different social and textual status (people approach a translated text with a specific set of expectations)”.
5. http://www.umist.ac.uk/ctis/events/conference_2000.htm.
6. <http://lael.pucsp.br/inpla>.
7. <http://www.umist.ac.uk/ctis/events/unisa.htm>.

8. Há ainda a área da terminologia, em que estão alguns trabalhos constantes neste volume, tais como Varantola, Maia e Tagnin.
9. Hunston (2002) e Fox (1998) ilustram como o “soar estranho” (“sound odd”) é um gatilho para o pesquisador, e como as investigações que resultam dessa impressão inicial podem ser frutíferas.
10. Concordância retirada do COBUILD Direct, em [http:// titania.cobuild.collins.co.uk/ form.html](http://titania.cobuild.collins.co.uk/form.html). Este corpus possui 56 milhões de palavras, de fala e escrita.

Referências bibliográficas

- Aarts, J. (1998) “Introduction”, in S. Johansson & S. Oksefjell (orgs.) *Corpora and Cross-Linguistic Research - Theory, Method, and Case Studies*. Rodopi: Amsterdam, pp ix-xiv.
- Altenberg, B. (1999) “Adverbial connectors in English and Swedish: Semantic and lexical correspondences”, in H. Hasselgard & S. Oksefjell (orgs.) *Out of Corpora - Studies in Honour of Stig Johansson*. Amsterdam/Atlanta, GA: Rodopi, pp 249-268.
- Altenberg, B., & Granger, S. (2002) “Recent trends in cross-linguistic lexical studies”, in B. Altenberg & S. Granger (orgs.) *Lexis in Contrast: Corpus-Based Approaches*. Amsterdam: John Benjamins, pp 3-49.
- Aston, G. (1999) “Corpus use and learning to translate”, *Textus*, 12, 289-314.
- Austermühl, F. (2001) *Electronic Tools for Translators*. Manchester: St. Jerome.

Avance de Souza, E. (2002) "A Lingüística do corpus e sua aplicação na atividade tradutória". Trabalho apresentado no Simpósio 'Lingüística de Corpus', no XII InPLA, PUC/SP, 25 a 26 de abril de 2002.

Baker, M. (1993) "Corpus Linguistics and translation studies: Implications and applications", in M. Baker, G. Francis, & E. Tognini-Bonelli (orgs.) *Text and technology: In honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, pp 233-250.

Baker, M. (1995) "Corpora in Translation Studies: An Overview and some Suggestions for Future Research", *Target*, 7, 223-243.

Baker, M. (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, pp 175-186.

Baker, M. (1998) "Réexplorer la langue de la traduction: Une approche sur corpus", *Meta*, 43, 480-485.

Baker, M. (1999) «The role of corpora in investigating the linguistic behaviour of professional translators», *International Journal of Corpus Linguistics*, 4:2, 281-298.

Barlow, M. (1995) "ParaConc: A concordancer for parallel texts", *Computers & Texts*, 10 .

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, pp. 147-154.

Berber Sardinha, A. P. (2000a) "Prosódia semântica na tradução do Português e Inglês: Um estudo baseado em corpus", in M. das G. Volpe Nunes (org.) *PROPOR '2000 - V Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada. 19 a 22 de novembro de 2000*. São Carlos: NILC / ICMC / USP, pp. 93-104.

Berber Sardinha, A. P. (2000b) "Semantic Prosodies in English and Portuguese: a Contrastive Study", in P. Cantos Gómez & A. Sánchez Pérez (orgs.) *Cuadernos*

de Filología Inglesa: 9, 1. Corpus-based Research in English Language and Linguistics. Murcia: Universidad de Murcia, pp. 93-109.

Berber Sardinha, A. P. (2001) «Linguística de Corpus e tradução: Algumas ferramentas computacionais» Trabalho apresentado no VIII Encontro Nacional de Tradutores e II Encontro Internacional de Tradutores, Belo Horizonte, 23-27 de julho de 2001.

Berber Sardinha, A. P. (2002) *Linguística de Corpus*. São Paulo: Editora Manole.

Botley, S., McEnery, T., & Wilson, A. (orgs.) (2000) *Multilingual Corpora in Teaching and Research*. Amsterdam/Atlanta, GA: Rodopi.

Bowker, L. (1998) "Using specialized monolingual native-language corpora as a translation resource: A pilot study", *Meta*, 43, 631-651.

Bowker, L. (1999) "Exploring the potential of corpora for raising language awareness in student translators", *Language Awareness*, 8:3/4, 160-172.

Bowker, L. (2001) "Towards a methodology for a corpus-based approach to translation evaluation", *Meta*, 46, 345-364.

Bowker, L. (2002) *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.

Bowker, L., & Pearson, J. (2002) *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Chesterman, A. (1998) *Contrastive Functional Analysis*. Amsterdam: John Benjamins.

Corness, P. (2002) "Multiconcord: A computer tool for cross-linguistic research", in B. Altenberg & S. Granger (orgs.) *Lexis in Contrast: Corpus-Based Approaches*. Amsterdam: John Benjamins, pp 307-326.

Ebeling, J. (1998) "The Translation Corpus Explorer - A browser for parallel texts", in S. Johansson & S. Oksefjell (orgs.) *Corpora and Cross-linguistic Research - Theory, Method, and Case Studies*. Amsterdam/Atlanta, GA: Rodopi, pp 101-113.

Fox, G. (1998) "Hocus pocus and graven images: collocation '98" Palestra apresentada no IATEFL 98, UMIST, Manchester, UK.

Frankenberg-Garcia, A. (2002) "COMPARA, a Web-searchable corpus of Portuguese and English parallel texts" Trabalho apresentado no Simpósio "Linguística de Corpus", no XII InPLA, PUC/SP, 25 a 26 de abril de 2002.

Gavioli, L., & Zanettin, F. (1997) "Comparable corpora and translation: a pedagogic perspective", trabalho apresentado em Corpus Use and Learning to Translate Conference, Advanced School of Modern Languages for Translators and Interpreters in Forlì, University of Bologna, Bertinoro, Italy, 14-15 November 1997. (Versão html consultada, disponível em <http://www.sslmit.unibo.it/cultpaps/laura-fede.htm>).

Ghadessy, M., & Gao, Y. (2001) "Small corpora and translation: Comparing thematic organization in two languages", in M. Ghadessy, A. Henry, & R. L. Roseberry (orgs.) *Small Corpus Studies and EFL - Theory and Practice*. Amsterdam/Philadelphia: John Benjamins, pp. 335-361.

Guidère, M. (2002) "Toward corpus-based Machine Translation for standard Arabic", *Translation and Computers*, 19 . (Trabalho disponível online em <http://accurapid.com/journal/19mt.htm>)

Halliday, M. A. K. (1992) "Language theory and translation practice", *Rivista Internazionale di Tecnica della Traduzione*, 0 , 15-25.

Hansen, S., & Teich, E. (2001) "Multi-layer analysis of translation corpora: methodological issues and practical implications". Trabalho apresentado no EUROLAN 2001, the Summer Institute on "Creation and Exploitation of Annotated Language Resources", 30 de julho a 11 de agosto de 2001, Romênia.

Hasselgard, H. (1998) "Thematic structure in translation between English and Norwegian", in S. Johansson & S. Oksefjell (orgs.) *Corpora and Cross-linguistic*

Research - Theory, Method, and Case Studies. Amsterdam/Atlanta,GA: Rodopi, pp. 145-168.

Hatim, B., & Mason, I. (1990) *Discourse and the translator*. London: Longman.

Hoey, M. (2000) "A world beyond collocation: New perspectives on vocabulary teaching", in M. Lewis (org.) *Teaching Collocation - Further Developments in the Lexical Approach*. Hove: LTP, pp. 224-243.

Hofland, K., & Johansson, S. (1998) "The Translation Corpus Aligner - A program for automatic alignment of parallel texts", in S. Johansson & S. Oksefjell (orgs.) *Corpora and Cross-linguistic Research - Theory, Method, and Case Studies*. Amsterdam/Atlanta,GA: Rodopi, pp. 87-100.

Hunston, S. (2000) "Colligation, lexis, pattern, and text", in M. Scott & G. Thompson (orgs.) *Patterns of Text - In honour of Michael Hoey*. Amsterdam/Philadelphia: John Benjamins, pp. 13-34.

Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S., & Francis, G. (2000) *Pattern Grammar - A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam/ Philadelphia: John Benjamins.

Johansson, S., & Oksefjell, S. (orgs.) (1998) *Corpora and Cross-linguistic Research - Theory, Method, and Case Studies*. Amsterdam/Atlanta,GA: Rodipi.

Kenny, D. (1997) "(Ab)normal translations: a German-English parallel corpus for investigating normalization in translation", in B. Lewandowska-Tomaszczyk & P. Melia (orgs.) *Practical Applications in Language Corpora '97 - Proceedings*. Lodz: Lodz University Press, pp. 387-392.

Kenny, D. (1998) "Creatures of habit? What translators usually do with words", *Meta*, 43, 515-523.

Kenny, D. (2001) *Lexis and Creativity in Translation - A Corpus-Based Study*. Manchester: St. Jerome.

Krzyszowski, T. (1990) *Contrasting languages. The scope of contrastive linguistics*. Berlin: De Gruyter.

Laviosa, S. (1997) "How Comparable Can Comparable Corpora Be?", *Target*, 9, 289-319.

Laviosa, S. (1998a) "Corpora in Translation Studies", in M. Baker (org.) *Routledge Encyclopedia of Translation Studies*. London: Routledge, pp. 50-53.

Laviosa, S. (1998b) "The corpus-based approach: A new paradigm in translation studies", *Meta*, 43, 631-651.

Laviosa, S. (org.) (1998c) "L'Approche basée sur le corpus/The Corpus-based approach" Special issue of *Meta* 43(4).

Laviosa, S. (2000) "TEC: A Resource for Studying What Is 'in' and 'of' Translational English". Trabalho apresentado na conferência 'Research Models in Translation Studies', UMIST, Manchester, 28 abril de 2000.

Laviosa, S. (2001) "Simplification: Before and after the advent of corpora". Trabalho apresentado no the European Society for Translation Conference 2001, 30 de agosto a 1 de setembro de 2001.

Laviosa, S. (2002) *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.

Lopes, M. C. (2000) "Homepages institucionais em português e suas versões para o inglês: Uma análise baseada em corpus de aspectos lexicais e discursivos" Dissertação de Mestrado Inédita, São Paulo, LAEL, PUC/SP.

Louw, B. (1993) "Irony in the text or insincerity in the writer: the diagnostic potential of semantic prosodies", in M. Baker, G. Francis, & E. Tognini-Bonelli (orgs.) *Text and technology - Essays in honor of John McH Sinclair*. Amsterdam/

Atlanta, GA: John Benjamins, pp. 157-176.

Magalhães, C. M. (2001) "Corpora-based Translation Studies in Brasil: towards universals of translation?" Trabalho apresentado no the European Society for Translation Conference 2001, 30 de agosto a 1 de setembro de 2001.

McEnery, T., & Wilson, A. (1993) "Corpora and Translation: Uses and Future Prospects", UCREL Technical Papers, University of Lancaster.

Melamed, I. D. (2001) *Empirical Methods for Exploiting Parallel Texts*. Cambridge, Mass: MIT Press.

Mihailov, M., & Tommola, H. (2001) "Compiling parallel and text corpora: Towards automation of routine procedures", *International Journal of Corpus Linguistics*, 6: Special Issue, 67-78.

Olohan, M. (no prelo) *Corpora in Translation Studies*. London: Routledge.

Partington, A. (1998) *Patterns and Meanings - Using Corpora for English Language Research and Teaching*. Amsterdam/Philadelphia: John Benjamins.

Salkie, R. (2002) "Two types of translation equivalence", in B. Altenberg & S. Granger (orgs.) *Lexis in Contrast: Corpus-Based Approaches*. Amsterdam: John Benjamins, pp. 51-72.

Sampson, G. (2001) *Empirical Linguistics*. London: Continuum.

Santos, D. (1999) "The pluperfect in English and Portuguese: What translation patterns show", in H. Hasselgard & S. Oksefjell (orgs.) *Out of Corpora - Studies in Honour of Stig Johansson*. Amsterdam / Atlanta, GA: Rodopi, pp. 283-300.

Santos, D., & Oksefjell, S. (2000) "An evaluation of the Translation Corpus Aligner, with special reference to the language pair English-Portuguese", in T. Nordgård (org.) *Proceedings from the 12th "Nordisk datalingvistikkdager" (Trondheim, 9-10 December 1999)* Trondheim: Department of Linguistics, NTNU, pp. 191-205.

Scott, N. (1998) "Normalisation and readers' expectations: a study of literary translation with reference to Lispector's *Hora da Estrela*". Tese de Doutorado Inédita, University of Liverpool, Liverpool.

Scott, N., & Scott, M. (2000) "Empirical methodology in translation using Corpus Linguistics techniques". Trabalho apresentado na conferência 'Research Models in Translation Studies', UMIST, Manchester, 28 abril de 2000.

Sinclair, J. McH. (org.) (1987) *Looking up: An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary*. London: Collins.

Siqueira, C. P. (2000) "Análise temática em estudos de tradução: O caso dos relatórios anuais de empresas brasileiras". Dissertação de Mestrado Inédita, São Paulo, LAEL, PUC/SP.

Somers, H. L. (org.) (no prelo) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins.

Steiner, E. H. (2001) "Translations English - German: investigating the relative importance of systemic contrasts and of the text-type "translation", in SPIK reports (org.) *Proceedings of the Symposium "Information structure in a cross-linguistic perspective", held at the Norwegian Academy of Science and Letters, November 30 - December 2, 2000* Oslo: Norwegian Academy.

Stubbs, M. (2001) *Words and Phrases - Corpus-based studies of lexical semantics*. Oxford: Routledge.

Tagnin, S. E. O. (2000). "Corpora and the Innocent Translator: How can they help him." Trabalho apresentado no The Lodz Session of the 3rd International Maastricht-Lodz Duo Colloquium on 'Translation and Meaning', Lodz (Polônia), 22 a 24 de setembro de 2000.

Tagnin, S. E. O. (2002) "Um corpus de integração: o projeto COMET" Trabalho apresentado no Simpósio 'Linguística de Corpus', no XII InPLA, PUC/SP, 25 a 26 de abril de 2002.

Teubert, W., Tognini-Bonelli, E., & Volz, N. (orgs.) (1998) *Translation Equivalence - Proceedings of the Third TELRI European Seminar*. Birmingham: TELRI / Institut für Deutsche Sprache / Tuscan Word Centre.

Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam/Atlanta, GA: John Benjamins.

Tognini-Bonelli, E. (2002) "Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach", in B. Altenberg & S. Granger (orgs.) *Lexis in Contrast: Corpus-Based Approaches* Amsterdam: John Benjamins, pp.. 73-96.

Toury, G. (1995) *Descriptive translation studies and beyond*. Amsterdam/Atlanta, GA: John Benjamins.

Tymoczko, M. (1998) "Computerized corpora and the future of translation studies", *Meta*, 43, 652-660.

Uzar, R., & Walinski, J. (2001) "Analysing the fluency of translators", *International Journal of Corpus Linguistics*, 6: Special Issue, 155-166.

Veronis, J. (2000) *Parallel Text Processing : Alignment and Use of Translation Corpora*. Dordrecht: Kluwer.