

A Identificação de equivalentes tradutórios em corpora comparáveis

Stella E. O. Tagnin
Universidade de São Paulo

Introdução

Um dos grandes problemas do tradutor é, sem dúvida, encontrar traduções satisfatórias para determinados termos ou expressões que ocorrem em seu trabalho. Nas últimas décadas, a Lingüística de Corpus tem se mostrado de grande utilidade nesse sentido, por oferecer a possibilidade de investigar textos autênticos similares, usando o computador (BOWKER, 2002, LAVIOSA, 2002, OLOHAN, 2004). Esses textos são compilados de forma criteriosa, segundo os objetivos da tradução, formando o que se denomina um **corpus**. No caso da tradução, são dois os tipos de corpora mais úteis: o paralelo, constituído de originais e respectivas traduções, e o comparável, composto de textos similares, originais, nas duas línguas de trabalho do tradutor. Qualquer um deles pode ser explorado com ferramentas eletrônicas, fornecendo um grande número de dados a partir dos quais o tradutor pode selecionar o termo ou expressão mais adequado ao seu contexto.

Quando se fala em encontrar “equivalentes tradutórios” num corpus, pensa-se, em geral, em corpora paralelos (FRANKENBERG-GARCIA, 2002a, 2002b, 2006). Embora isso talvez fosse mais fácil, na maioria das vezes não é um procedimento possível devido à escassez de recursos desse tipo, em especial de corpora paralelos nas áreas técnicas. Por que então não recorrer a corpora comparáveis, ou seja, corpora bilíngües constituídos de textos sobre os mesmos assuntos, de gênero, tipologia, extensão e data de publicação comparáveis? É o que se pretende discutir a partir do material disponível no site do **CorTec** (Corpus Técnico), parte do projeto **COMET – Corpus Multilíngüe para Ensino e Tradução**¹, composto de cinco corpora comparáveis inglês-português, nas seguintes áreas: Informática (FROMM, 2005), Ecoturismo (MARTINS, 2005), Instrumentos Contratuais (FONSECA, 2005), Culinária (TEIXEIRA, 2005) e Hipertensão Arterial (GINEZI e CASTANHO, 2005). Cada corpus contém aproximadamente 200.000 palavras em cada língua.

Com o auxílio dos recursos da Lingüística de Corpus, apresentaremos, neste artigo, algumas formas de identificar possíveis equivalentes tradutórios, fazendo uso dos três últimos corpora acima mencionados. Para comprovar a validade da metodologia, usaremos também um corpus similar sobre a Copa do Mundo de 2006.

A Equivalência

Embora o termo “equivalência” cause muita polêmica entre os estudiosos da tradução (RODRIGUES, 1998), não se pode deixar de reconhecer a necessidade enfrentada pelo tradutor quando tem de encontrar um termo que “funcione” no texto de chegada como “funciona” no texto de partida. É esse conceito pragmático de equivalência que adotaremos neste artigo.

Outro problema com que se defronta o tradutor é a necessidade de, em se tratando de texto técnico, produzir uma tradução fluente a fim de garantir melhor entendimento por parte do leitor. Empregar um termo não usual com certeza causará certo estranhamento nesse leitor.

¹ http://www.fflch.usp.br/dlm/comet/consulta_cortec.html

Seria possível argumentar que existem dicionários e glossários técnicos, repositórios desses termos. Na realidade, no entanto, essas fontes nem sempre oferecem ajuda, quer pela falta de critério na compilação dos termos e, em geral, pela falta de exemplos de uso, quer por não poderem manter-se atualizados devido ao rápido avanço das pesquisas científicas e tecnológicas.

Um corpus – paralelo ou comparável –, em contrapartida, pode ser constantemente atualizado, além de fornecer exemplos autênticos de uso, o que confere segurança ao tradutor na escolha do termo a empregar.

O tradutor que recorre apenas a fontes de referência convencionais pode acabar empregando um termo “possível”, porém não consagrado (BOWKER, 1999; TAGNIN, 2005). Já um corpus comparável, composto de textos originalmente escritos nas duas línguas, inglês e português no caso em questão, fornecerá o termo efetivamente usado por aquela comunidade. Em outras palavras, o corpus fornecerá o termo mais “provável” de ocorrer no contexto que o tradutor estiver pesquisando, ou seja, o termo mais recorrente. Nem sempre, é óbvio, o tradutor deseja usar o termo mais recorrente, mas, mesmo nesse caso, o corpus lhe será de grande valia justamente para identificar o termo que *não* pretende usar.

Nas seções seguintes, discutiremos a metodologia empregada para identificar possíveis “candidatos” a equivalentes e, em seguida, aboná-los ou não, sempre recorrendo ao corpus. Vale lembrar que essa metodologia também pode ser empregada pelo terminólogo bilingüe, uma vez que a busca de equivalentes é o ponto em comum entre esses dois profissionais.

O CorTec

Para explorar os corpora que constituem o CorTec, o site conta com três ferramentas (Fig. 1): um Gerador de Lista de Palavras, que produz listas das palavras contidas nos corpora e apresenta-as por ordem de frequência ou alfabética; um Concordanciador, que apresenta o contexto de ocorrência de um item, permitindo buscas por palavras ou expressões (Expressão ou palavra igual a), por prefixos ou início de palavras (Começando com), por sufixos ou terminações (Terminando com) ou ainda partes de palavras (Contendo) – vide Fig. 2 –, e ainda um Gerador de N-gramas, que apresenta combinações com 2, 3 ou 4 palavras.



Figura 1: Cortec: Seleção das Ferramentas



Figura 2: Cortec: Configuração do Concordanciador

Para nossas pesquisas, o primeiro passo foi obter uma lista de frequência do vocabulário de cada corpus estudado, em ambas as línguas. A seguir, foram feitas concordâncias com as palavras que se pretendia estudar para identificar seus contextos de ocorrência. Como cada problema demandava uma estratégia de pesquisa distinta, descrevemos abaixo os procedimentos seguidos em cada um dos casos.

Instrumentos Contratuais

No corpus de Instrumentos Contratuais observamos, a partir da lista de frequência, que o termo *contrato* era a palavra de conteúdo de maior frequência, com 1832 ocorrências – como era de se esperar, já que a palavra de conteúdo (em oposição a palavras gramaticais) mais frequente costuma indicar a área em que se insere o corpus –, enquanto seu cognato em inglês, *contract*, ocorria apenas 186 vezes. Isso levou-nos a supor que nem sempre *contrato* seria traduzido por *contract*. De fato, a hipótese confirmou-se no corpus em inglês: a palavra de conteúdo mais frequente era *agreement*, com 1724 ocorrências. Por meio das concordâncias foi possível confirmar essa equivalência, ou seja, que ambos ocorriam em contextos similares:

1a For the purposes of this **Agreement**, all merchantable Logs 6" in diameter o
1b Constitui objeto do presente **contrato** o intercâmbio eletrônico de documentos

2a 12.2 This **Agreement** may be cancelled by either party, at it
2b 13.2 - A rescisão deste **contrato** implicará retenção de créditos decorren

3a The term of this **Agreement** shall expire June 30, 2001, (the "Term")
3b O presente **contrato** terá prazo de (xxx), iniciando-se no di

Por outro lado, observou-se que a palavra que co-ocorria mais frequentemente com *contract*, isto é, seu colocado, era *adhesion*, o que gerava a colocação *adhesion contract*. Uma rápida busca na internet, usando o Google, revelou 42.000 ocorrências para “adhesion contract”, mas apenas 890 para “adhesion agreement”. É importante atentar, no entanto, que grande parte

dessas últimas não provêm de sites de países cuja língua oficial é o inglês:

4. **EMBRATEL - Residential - Adhesion Agreement**

The present document, hereinafter referred to as **Adhesion Agreement**, defines the conditions and relations that shall prevail between Empresa Brasileira de ...
www.embratel.com.br/Embratel02/cda/portal/0,2997,RE_I_5233,00.html -

5. **ADHESION AGREEMENT**

when necessary, for the payment of conventional penalties under the terms of the **Adhesion Agreement** and of this agreement, as ...
www.scotiacapital.com/Products&Services/CR_Mexico/Clearing_Contract_eng.pdf -

6. **Tractebel Energia (Legal Aspects)**

(47kB) Privacy Policy Part of the **Adhesion Agreement**. (46kB) **Adhesion Agreement** - English version not available Regulates the use of Online Purchase ...
www.tractebelenergia.com.br/.../empty.asp?P=624&VID=default&SID=206736320449512&S=1&C=21600

7. **NATLEX**

10(1-3) and Article III of the first part) entered in force with the entry in force of the **Adhesion agreement** of Slovakia in European Union ...
www.ilo.org/dyn/natlex/natlex_browse.details?p_lang=es&p_country=SVK&p_classification=22&p_or...

Figura 3: Parte dos resultados gerados pelo Google para *adhesion agreement*, em que se observa que os textos não provêm de sites de países cuja língua oficial é o inglês.

Observamos, assim, que não basta identificar os termos de maior frequência nas duas línguas para considerá-los “equivalentes”; é preciso analisar o contexto em que ocorrem, assim como seus colocados. Dessa forma, foi possível estabelecer que a colocação consagrada é “adhesion contract” e não “adhesion agreement”, que ocorre com maior frequência em textos escritos por não nativos da língua inglesa, indicando tratar-se, possivelmente, de traduções, um forte indício de que esses “termos” devem ser evitados por não corresponderem à forma consagrada no inglês.

Culinária

Na Culinária optamos por estudar equivalentes adverbiais, uma vez que já se observara a alta frequência dessa categoria em textos da área (Moraes & Teixeira, 2006). Para este estudo selecionamos o advérbio mais recorrente no corpus em inglês, *finely*, e seus cinco colocados mais frequentes: *chopped*, *sliced*, *diced*, *grated* e *shredded*, identificados pelas linhas de concordância ordenadas alfabeticamente pela primeira palavra à direita do advérbio. Embora essa ordenação não seja possível no CorTec, o site oferece a possibilidade de baixar (Fazer o download) os resultados das buscas para a máquina do pesquisador, de modo que esse material possa ser analisado com ferramentas adequadas. No nosso caso, usamos o WordSmith Tools (Scott, 1996), versão 3.0, para fazer essa ordenação. Esse programa conta com um conjunto de ferramentas para análise de corpus, das quais destacamos a WordList, semelhante ao Gerador de Lista de Palavras do Cortec, e o Concord, semelhante ao Concordanciador do Cortec, mas com muito maior flexibilidade na apresentação dos dados. O mesmo procedimento foi adotado para o corpus em português.

86	ets	68 z)	Soy sauce - 1 tbsp	Onion - 1 medium,	finely chopped	Celery - 3 sticks,	finely chopp
87	1/2 o	264	apples - 450g (1 lb),	peeled, cored and	finely chopped	Onions - 225g (8 oz),	finely ch
88	fl oz)	396	Milk - 600 ml (1 pint)	Onion - 2 tbsp,	finely chopped	Celery - 2 tbsp,	finely chopped

Figura 4: Seleção de linhas de concordância geradas pelo WST para *finely*, ordenadas pela 1ª. palavra à direita.

Como um dos possíveis equivalentes para *chop* é “picar”, e para *chopped*, “picado, picada, picados, picadas”, fizemos uma concordância para “picad*”, que apresenta todas as palavras que começam com “picad”. Notamos que há apenas 29 ocorrências em que o colocado é “fino” ou algum derivado seu: “picad* fin*” (18 ocorrências), “finamente picad*” (11 ocorrências). Observem-se, inclusive, as formas incorretas *picadas finas*:

sem sementes e picado 1 cebola pequena, **picada fino** 2 col. (124 sem sementes
 291 dÍvias, espinafre) Molho 3 cebolinhas **picadas finas** 1 col. (chá) de coentro fres
 8 kg de toucinho fresco 2 cebolas grandes **picadas finas** 4 dentes de alho amassad
 6 entes 4 pãezinhos franceses 4 salsichas **picadas fino** 4 colheres (sopa) de creme
 822 e picados fino 1 dente de alho grande, **picado fino** Sal e pimenta-do-reino 1/3 –
 uros, mas firmes, sem pele e sementes e **picados fino** 1 dente de alho grande, pica

Figura 5: Seleção de algumas linhas de concordância geradas pelo WST para *picad**, ordenadas pela 1ª. palavra à direita.

Por outro lado, o advérbio com que “picad*” mais co-ocorre é “bem” (79 ocorrências): “bem picada, bem picado” etc. O mais interessante foi também encontrar “picadinh*” (96 ocorrências, sendo 10 de “bem picadinha”). Dessa forma, o corpus indica que as melhores equivalências para *finely chopped* seriam “bem picad*” ou “picadinh*”, ou seja, o advérbio *finely* transforma-se em “bem” ou na forma diminutiva do adjetivo.

2 cebolas médias **bem picadas**
 ½ dente de alho **bem picado**
 junte os tomates pelados **bem picados**.
 Calabresa **picadinha**
 100 g de bacon **picadinho**
 2 dentes de alho **picadinhos**
 Polvilhar salsa **bem picadinha**
 ½ cebola **bem picadinha**

Figura 6: Seleção de algumas linhas de concordância geradas pelo WST para *picad**, ordenadas pela 1ª. palavra à esquerda

Outra alternativa revelada pelo corpus é a ocorrência do verbo “picar” com a locução adverbial “bem fininh*” (8 ocorrências):

Pique a cebola bem fininho
 Descasque o ovo, **pique-o bem fininho**
 3 dentes de alho descascados e **picados bem fininho**

Figura 7: Seleção de linhas de concordância para “bem fininho”

O próximo colocado é *sliced*. Em português emprega-se, em geral, “cortar em fatias” e não o verbo “fatiar”, de modo que a forma equivalente a *finely sliced* que o corpus revela é, na grande maioria das vezes, “cortado em fatias (ou rodela) finas”. Novamente desaparece o advérbio, substituído pelo adjetivo.

Calda 4 laranjas descascadas **cortadas em fatias finas**
 200 g de cebola **cortada em fatias finas**
 1 pepino sem sementes **cortado em fatias finas**
 6 rabanetes, **cortados em fatias finas**
 Juntar as batatas **cortadas em fatias finas**.
 Decore a quiche com um alho-poró cru **cortado em rodela finas**.
 1 cebola média **cortada em rodela finas**
 400 g de lingüiça portuguesa **cortada em rodela finas**

1 pimentão vermelho médio **cortado em rodelas finas**
Corte em rodelas fininhas

Figura 8: Seleção de linhas de concordância geradas para “cort*” seguida de “fin*” até 3 casas à direita

Fenômeno similar ocorre com *finely diced*, em que o verbo equivalente em português é “cortar em cubos”. O tamanho dos cubos é indicado por um adjetivo, nesse caso “pequeno”: “corte em cubos pequenos”, ou pelo diminutivo do substantivo: “corte em cubinhos”:

1/2 xíc. (chá) de bacon **em cubinhos**
500 g de peito de frango cozido e **cortado em cubinhos**
5 tomates grandes sem pele e sem sementes, **cortados em cubinhos**
1 berinjela **cortada em cubinhos**
100 g de bacon **em cubos pequenos**
1 berinjela média **cortada em cubos pequenos**
2 xícaras de abacaxi **cortado em cubos pequenos**

Figura 9: Seleção de linhas de concordância geradas para “cub*”

Finely grated refere-se, em geral, a queijos². Como, no Brasil, o queijo costuma ser “ralado fino” (apenas 2 ocorrências, sendo uma para queijo parmesão e outra para gruyère), isso nem sempre é especificado, sendo mais comum especificar-se “ralado grosso” (32 ocorrências). Assim, *finely grated parmesan cheese* poderia ser traduzido apenas por “queijo parmesão ralado”. Em outras palavras, por razões culturais, o advérbio desaparece.

2 col. (sopa) de queijo parmesão **ralado fino**
80g de queijo gruyère **ralado fino**
2 colheres (sopa) de parmesão **ralado**
80g de parmesão **ralado**
1 xícara de queijo prato **ralado grosso**
2 xícaras de queijo mussarela **ralado grosso** (200 g)
Para polvilhar 50 g de queijo parmesão **ralado grosso**

Figura 10: Seleção de linhas de concordância geradas para “ralado*”

Já quando se refere a hortaliças ou chocolate, encontramos “ralado/a (bem) fino”.

1 cebola **ralada fino**
1 1/2 xícara de cenoura **ralada fino**
4 xícaras de repolho **ralado fino**
Cobertura de chocolate **ralado bem fino**

Figura 11: Seleção de linhas de concordância geradas para “ralad*”

O verbo *shred* significa “rasgar, esfrangalhar; cortar em pedaços ou tiras” (Houaiss, 1997). Em inglês, é comumente empregado em relação a repolho, alface³ etc. No corpus em português, a expressão que co-ocorre com essas hortaliças é “cortar em tiras (finas)/tirinhas” ou apenas “cortar bem fininho”. No entanto, encontramos “rasgar” usado com alface: “folhas de alface lavadas e rasgadas”. Mais uma vez, observa-se praticamente a inexistência do advérbio, ao qual passa a corresponder sua forma adjetiva ou a forma diminutiva do substantivo: “tirinha(s)”.

² mas também pode ocorrer, entre outros, com *lemon zest* (“casca de limão”).

³ mas também com carnes: *shredded chicken*, por exemplo.

10 folhas de alface **cortadas em tiras finas**
Corte a alface **em tiras**.
 substitua o agrião por alface **cortada em tirinhas**.
 4 xícaras de repolho **cortado em tirinhas**
Corte o repolho **em tirinhas bem finas**.
 1/2 repolho roxo **cortado bem fininho**
 e couve **cortada bem fininho**.

Figura 12: Seleção de linhas de concordância geradas para “cortad*”

Em resumo, *finely* não costuma ser traduzido por “finamente”, mas, em geral, por “bem” seguido do adjetivo “fin*” ou pelo adjetivo na sua forma diminutiva.

Talvez esse estudo possa ser replicado para outros advérbios terminados em “mente”, assim como para outras áreas técnicas, uma vez que sua tradução representa um problema principalmente estilístico: em inglês é comum uma seqüência de advérbios terminados em “ly”, mas em português uma seqüência similar com advérbios terminados em “mente” nem sempre é desejável.

Hipertensão Arterial

No corpus de Hipertensão Arterial foram identificados os equivalentes de colocações (ou termos) que contivessem os vocábulos *coração/heart* ou o radical *cardi-*.

O primeiro passo foi gerar as linhas de concordância para *heart*, que produziu 768 ocorrências. Como o CorTec não indica os colocados, essas concordâncias foram examinadas com o programa WST, que permitiu identificar *failure*, *rate* e *disease* como os mais freqüentes. Tentamos, então, encontrar um equivalente para *heart failure*, partindo de uma concordância para “coração”, que apresentou apenas 207 ocorrências, em nenhuma das quais foi possível identificar algum equivalente. Partimos, então, para o radical latino “cardi-”, em português, que apareceu em 1873 instâncias, sendo 817 ocorrências para “cardiovascular(es)”. Como não sabíamos se o equivalente de *failure* seria uma palavra feminina ou masculina, observamos mais uma vez a concordância para *heart* em busca de outro colocado que pudesse fornecer uma pista, e notamos que um termo recorrente era *congestive heart failure*.

1	g mo	290	hythmias requiring treatment, worsening heart failure , and myocar
2	t ventr	289	re hospitalized due to new or worsening heart failure and who wi
3	sudde	182	sudden cardiac death) in patients with heart failure . [3-6] Steven
11	inuria	365); acute pulmonary edema, congestive heart failure , left ventricul
12	e was si	312	ocardial infarction, stroke, congestive heart failure , coronary by
13	hows	495	h hypertension, obesity, and congestive heart failure , the increas
14	, cere	515	ion, coronary heart disease, congestive heart failure , cerebrovas
15	wever,	503	re involved. For example, in congestive heart failure and obesity,

Figura 13: Seleção de linhas de concordância geradas para *heart failure*

Uma concordância para “congestiv*” revelou, de imediato, “insuficiência cardíaca congestiva”, de modo que “insuficiência cardíaca” apresentou-se como um forte candidato a equivalente de *heart failure*.

1 mHg, presença de **insuficiência cardíaca congestiva**, hemorragia cere- bral, insuficiência
 2 is quando existe **insuficiência cardíaca congestiva** associada. Podem causar insuficiência
 3 tensão severa ou **insuficiência cardíaca congestiva** associada. Limitações do seu emprego
 4 e Cardiologia. A **insuficiência cardíaca congestiva** (ICC) é síndrome que basicamente apre

Figura 14: Seleção de linhas de concordância geradas para “congestiv*”

Nota-se, assim, que a busca pelo colocado pode ser uma estratégia eficaz para identificar termos equivalentes.

Copa do Mundo

Esta seção tem por objetivo demonstrar que o mesmo tipo de pesquisa pode ser feita com corpora do tipo “faça-você-mesmo” (ZANETTIN, 2002), ou seja, compilados pelo próprio pesquisador, conforme seus interesses e necessidades. Nosso corpus foi compilado durante a Copa de 2006, com textos publicados na internet por periódicos brasileiros (*Folha de São Paulo, O Estado de São Paulo, Gazeta Esportiva e Lance*) e americanos (*New York Times, Washington Post, Newsweek International*). O corpus em inglês contém 145 textos totalizando 133.703 palavras, enquanto o corpus em português contém 144 textos, porém apenas 69.418 palavras, o que indica que os textos coletados em português são de menor extensão. Para o estudo desse corpus recorreremos ao WordSmith Tools (Scott 1996), cuja ferramenta WordList permitiu-nos identificar os verbos mais recorrentes nas duas línguas. A partir daí, buscamos seus possíveis equivalentes.

O primeiro termo para o qual procuramos um equivalente em inglês foi “marcar um gol”. Para isso, fizemos uma concordância com *goal* e pedimos que as linhas fossem ordenadas alfabeticamente pelas três primeiras palavras à esquerda, o que nos levou a identificar *score a goal* como possível equivalente:

40	fending champions, they didn't score a goal and went home after the first round.
41	hether the Americans can finally score a goal or two on their own and make thei
42	rip was in 1974, when it failed to score a goal in losing to host West Germany a
49	eam that lost the penalty kicks scores a goal in the overtime period, it wins. Th
50	nt out in the first round without scoring a goal . The tension in the French camp

b

Figura 15: Seleção de linhas de concordância geradas para *scor* a goal*

Durante a pesquisa com “goal”, uma colocação que nos chamou a atenção foi “gol contra”. Fizemos então uma busca inicial com *goal* seguido ou precedido de *against* e obtivemos apenas nove ocorrências (uma delas repetida, o que indica que deve haver um texto repetido no corpus, que deve ser eliminado numa futura revisão):

1	efender Fabio Grosso, who scored the go-ahead goal against Germany in the 119th minute,
2	le in this World Cup for Trinidad. England's first goal against T&T came when Peter Crouch
3	tional speed. But with the exception of the lovely goal against Ghana that he created with a
4	as yielded just one goal in six games -- an own goal against the United States in a first-
5	eded only two goals in the tournament: an own- goal against the United States and Zidane's
6	eded only two goals in the tournament: an own- goal against the United States and Zidane's
7	rship of the World Cup without scoring a solitary goal against moderate opposition. There
8	naldinho is still beneath him, but — " But that goal against Chelsea? That triumph of
9	oals and he won the penalty that produced the goal against Portugal. In that game, he

Figura 16: Linhas de concordância geradas para *goal against*

Observamos, no entanto, que em todas as ocorrências tratava-se de gol marcado contra a outra equipe, e não um gol contra o próprio time. Assim, fez-se necessária outra estratégia de busca. Voltamos ao corpus em português e procuramos as ocorrências de *gol contra*:

1	Ribery, o argentino Maxi Rodríguez - cujo gol contra o México foi o mais bonito da C
2	gola. É muito perigoso e criou a jogada do gol contra Angola jogando atrás”, lembrou.

3 a deu o chute que foi desviado por Petit no **gol contra**. - É ótimo marcar dois gols p
4 Os croatas começaram na frente, com um **gol contra** de Pablo Ibañez, logo
5 zzurra só só foi vazada uma única vez: um **gol contra** do lateral Zaccardo, no empate

Figura 17: Linhas de concordância geradas para “gol contra”

Nas duas primeiras ocorrências temos novamente o gol contra outra equipe, mas nas outras três encontramos exemplos da colocação que buscávamos. Como era mencionado o autor do gol, pedimos uma concordância com um desses nomes, Zaccardo, na qual foi possível identificar *own goal* como possível equivalente para “gol contra” a partir do contexto:

1 oss was sliced into his own net by Cristian Zaccardo. Things got even worse for Italy
2 eable for the Italians, but defender Cristian Zaccardo flailed at it with his left foot and
3 d in the 27th when Italian defender Cristian Zaccardo knocked the ball into his own n
4 ed on," Buffon said. Italy defender Cristian Zaccardo scored an **own-goal** 27 minutes i
5 nt of panic by a reserve defender, Cristian Zaccardo, who put the ball into his own net
6 nced out by a an **own goal** from Cristian Zaccardo, but the real talking point of an e
7 up E clash but an **own goal** from Cristian Zaccardo put the Americans back on level
8 United States three weeks ago on Cristian Zaccardo's **own goal**. The Azzurri had be

Figura 18: Linhas de concordância geradas para *Zaccardo*

Uma nova concordância, dessa vez feita para *own goal*, confirma a equivalência:

1 's opener for Italy was canceled out by a an **own goal** from Cristian Zaccardo, but the r
2 s yielded just one goal in six games -- an **own goal** against the United States in a firs
3 ith a header in their Group E clash but an **own goal** from Cristian Zaccardo put the A
4 imination of his team in the 1998 Cup by an **own goal**, was assassinated when he retur
5 minutes for U.S. soccer there -- we had an **own goal** and a red card,' " said defender
6 nited States tied it five minutes later on an **own goal**. The scoring ended, the drama
7 nder, Andres Escobar, scored an **own goal** to help the United States to a
8 araguay captain Carlos Gamarra scores an **own goal** to give England the lead as the
9 en up one goal in six games, and it was an **own goal**. And the Italians have done it wi
10 e wanted." England had to rely on an early **own goal** by Paraguay captain Carlos Gam

Figura 19: Seleção de linhas de concordância geradas para *own goal*

A pesquisa com corpus não se restringe à confirmação ou não de hipóteses, mas aponta para novos problemas e novos caminhos. Foi o caso também de *overtime*, que apareceu, durante as pesquisas, na seguinte linha de concordância da Figura 15:

49 eam that lost the penalty kicks **scores a goal** in the overtime period, it wins. Th

A dúvida que surgiu foi: *overtime* referia-se ao tempo acrescentado ao final de um jogo em razão de paralisações no seu decorrer ou ao período de 30 minutos, acrescido em virtude de empate no tempo normal da partida?

Uma concordância para *overtime* retornou 37 ocorrências, das quais selecionamos as seguintes, tendo assinalado em negrito as partes que parecem esclarecer a dúvida:

1 as the game extended from regulation to a **pair of 15-minute overtime periods**.
2 Too many games decided on the **free kicks that follow the overtime period**.
3 In the 110th minute, early in **the second overtime session** of a 1-1 tie at a sold-out Olympic Stadium,
4 In overtime, the two became involved again, this time with Zidane
5 One minute into **the injury time** added on to **the 30-minute overtime**,
6 Their patience almost paid off at the start of **the 30-minute overtime** as reserve

Figura 20: Seleção de linhas de concordância geradas para *overtime*

Observa-se, nas linhas 1, 5 e 6, a referência à duração da *overtime*. A linha 2 também fornece uma informação relevante: os lances livres (pênaltis) seguem-se à *overtime*. A linha 3 faz referência à segunda *session* (segundo tempo) da *overtime*.

Assim, o contexto esclarece-nos que *overtime* refere-se ao “período de 30 minutos, acrescido em virtude de empate no tempo normal da partida”.

A outra hipótese, “tempo acrescentado ao final de um jogo em razão de paralisações no seu decorrer”, parece estar contemplada na linha 5, *injury time*, confirmada em 14 linhas de concordância, das quais destacamos:

-
- 1 John Aloisi added one in **injury time**.
 - 2 Zinedine Zidane's substitution in **injury time** could mark his World Cup farewell -- he will m
 - 3 Rahdi Jaidi's header in **injury time** Wednesday gave Tunisia a 2-2 tie with Saudi
 - 4 ended dramatically on reserve Oliver Neuville's goal in **injury time**.
 - 5 Nadj was ejected in first-half **injury time**, and Domoraud got his second yellow card in s
 - 6 and Domoraud got his second yellow card in second-half **injury time**.

Figura 21: Seleção de linhas de concordância geradas para *goal against*

As linhas 5 e 6 esclarecem, inclusive, que esse período pode ser acrescido tanto ao primeiro tempo quanto ao segundo.

A menção do nome dos jogadores nos exemplos acima pode levar-nos a citações similares no corpus em português.

Buscando por *Aloisi*, encontramos 4 linhas, das quais destacamos

4. **Aloisi** deu números finais ao placar já nos **acréscimos**

Na busca por *Zidane*, encontramos, dentre 54 linhas de concordância, as seguintes:

- 5a. mas acabou levando mais um gol nos **acréscimos**. **Zidane** recebeu livre na esquerda, deu um corte em Puyol,
- 5b. do parecia que mais uma partida seria decidida na **prorrogação**, **Zidane** cobrou uma falta da direita,

em que aparecem os termos “acréscimos” e “prorrogação”. Se ampliarmos o contexto das duas, notaremos que ocorrem no mesmo texto:

E quando parecia que mais uma partida seria decidida na **prorrogação**, **Zidane** cobrou uma falta da direita, a zaga espanhola desviou e Vieira, livre na segunda trave, cabeceou firme, colocando a França na frente. A Espanha partiu para uma pressão nos últimos minutos, mas acabou levando mais um gol nos **acréscimos**. **Zidane** recebeu livre na esquerda, deu um corte em Puyol, e bateu firme, vencendo o goleiro Casillas, seu companheiro de Real Madrid.

podendo-se inferir que “prorrogação” é um possível equivalente tradutório para *overtime*, e “acréscimos”, para *injury time*, conforme se pode confirmar nas linhas de concordância abaixo para esses dois termos.

-
- 1 Nos **acréscimos**, em jogada começada por Lúcio, a bola passou por Kaká e Cafu,
 - 2 a Austrália nas oitavas por 1 a 0 com um gol marcado nos **acréscimos**, fruto de um pênalti duvidoso.

- 3 A Argentina conseguiu marcar o gol nos **acréscimos**, com Messi, mas o árbitro viu
- 4 Aloisi deu números finais ao placar já nos **acréscimos**.
- 5 Quando todos esperavam a prorrogação, os italianos definiram a vitória nos **acréscimos**.
- 6 nos últimos minutos, mas acabou levando mais um gol nos **acréscimos**
- 7 mas conseguiu um gol de pênalti marcado por Totti, nos **acréscimos** da partida.
- 8 Van Bronckhorst acabou expulso nos **acréscimos** por entrada dura em Tiago.

Figura 22: Seleção de linhas de concordância geradas para “acréscimos”

-
- 1 Ahn marcou o gol que eliminou, aos 12min do 2º tempo da **prorrogação**, os italianos nas
 - 2 e sacramentou a vitória argentina sobre o México por 2 a 1, já na **prorrogação**,
 - 3 mas nada disso foi suficiente para evitar que a partida fosse para a **prorrogação**
 - 4 mas viu sua equipe deixar a Argentina virar o jogo na **prorrogação**, com um belo gol de Maxi 5 Maxi Rodriguez, que fez o gol da Argentina aos sete minutos da **prorrogação**.
 - 6 Quando todos esperavam a **prorrogação**, os italianos definiram a vitória nos acréscimos.
 - 7 A etapa final da **prorrogação** foi ainda mais sonolenta do que os primeiros 105 mi
 - 8 sul-americanos definiram o placar com gol de Maxi Rodriguez na **prorrogação**.

Figura 23: Seleção de linhas de concordância geradas para “prorrogação”

A linha 5 da Figura 22, em que ocorrem os dois termos, deixa bem clara a diferença: caso não houvesse vitorioso no tempo normal, inclusive nos “acréscimos”, a partida teria de ser decidida na “prorrogação”.

Considerações finais

O objetivo deste artigo foi demonstrar como a Lingüística de Corpus, por meio de um corpus comparável bilíngüe, pode auxiliar o tradutor a encontrar equivalentes para termos técnicos em diversas áreas.

Em relação aos corpora usados, tanto no do CorTec quanto no da Copa do Mundo, os resultados demonstraram que: a) de uma forma ou de outra, foi possível estabelecer os equivalentes, o que indica que os corpora são de fato comparáveis, contendo os mesmos assuntos/temas, nas mesmas áreas; b) nos casos em que isso foi mais difícil, como no caso da Copa do Mundo, o fato deveu-se justamente à constituição do corpus, que apresentava um desequilíbrio no número de palavras em cada língua. A partir dessa constatação, o ideal seria o corpus em português ser aumentado para obter melhor balanceamento. Infelizmente, nesse caso, não será fácil encontrar novos textos em português sobre esse tema, uma vez que a Copa se encerrou em julho de 2006 e grande parte do material não está mais disponível na internet, mas geralmente esse é um procedimento possível e recomendável.

Quanto à metodologia, verificou-se que: a) em alguns casos já foi possível estabelecer eventuais candidatos a equivalentes a partir das listas de frequência; b) quando isso não se deu, recorremos às linhas de concordância para verificar os colocados e seus contextos de ocorrência. Como no site do CorTec não existe a possibilidade de ordenar as concordâncias por ordem alfabética dos colocados, baixamos para nossa máquina, quando necessário, as concordâncias geradas por esse programa e passamos a usá-las com o WordSmith Tools, que permite essa manipulação; c) a Lingüística de Corpus, como é característica sua, permitiu também descobrir fatos que não eram o objetivo direto do estudo, como foi o caso de *gol contra* e *overtime*, que apareceram durante uma busca por *score a goal*, despertaram nossa curiosidade e levaram-nos a buscar seu equivalente.

Acreditamos que os diversos procedimentos metodológicos apresentados demonstraram sua eficácia no estudo dos fenômenos lingüísticos em geral, e lexicais em particular, tão

essenciais para os tradutores – e terminólogos bilíngües, conforme lembramos no início deste trabalho –, principalmente no tocante à identificação de equivalentes em duas línguas, seja para um trabalho de tradução, seja para a construção de glossários ou dicionários bilíngües.

Referências bibliográficas

BOWKER, L. Exploring the Potential of Corpora for Raising Language Awareness in Student Translations, **Language Awareness**, Vol. 8. No. 3&4, p. 160-173, 1999.

BOWKER, L. **Computer-aided Translation Technology**, Ottawa: University of Ottawa Press, 2002.

CASTANHO, R. M. C. & GINEZI, L. L. Compilação do corpus comparável inglês-português de **Hipertensão** do Cortec. http://www.fflch.usp.br/dlm/comet/consulta_cortec.html, 2005.

FONSECA, L. C. Compilação do corpus comparável inglês-português de **Instrumentos Contratuais** do Cortec. http://www.fflch.usp.br/dlm/comet/consulta_cortec.html, 2005.

FRANKENBERG-GARCIA, A. Using a parallel corpus to examine English and Portuguese translations. Apresentado no congresso internacional **Translation (Studies): a crossroads of disciplines**, Faculdade de Letras, Universidade de Lisboa, 14-15 novembro 2002, 2002^a.

FROMM, G. Compilação do corpus comparável inglês-português de **Informática** do Cortec. http://www.fflch.usp.br/dlm/comet/consulta_cortec.html, 2005.

FRANKENBERG-GARCIA, A. COMPARA, language learning and translation training. In MAIA, B, HALLER, J. & ULRYCH, M. (eds.) **Training the Language Service Provider for the New Millennium**. Porto: FLUP, p. 187-198, 2002b.

FRANKENBERG-GARCIA, A. Using a parallel corpus in translation practice and research. In Actas da *Contrapor 2006, 1^a Conferência de Tradução Portuguesa*, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, p. 142-148, 2006.

HOUAISS, A. (ed.) **Dicionário Inglês-Português – Nova Edição**, Rio de Janeiro/São Paulo: Record.

LAVIOSA, S. **Computer-based Translation Studies – Theory, Findings, Applications**, Amsterdam/New York: Rodopi, 2002.

MARTINS, J. C. Compilação do corpus comparável inglês-português de **Ecoturismo** do Cortec. http://www.fflch.usp.br/dlm/comet/consulta_cortec.html, 2005.

MORAES, H. F. R. & TEIXEIRA, E. D. Os advérbios na Culinária – um ingrediente importante para a tradução adequada das receitas. 2006. Manuscrito.

OLOHAN, M, **Introducing Corpora in Translation Studies**, London/New York: Routledge, 2004.

RODRIGUES, C. C. Tradução e diferença: uma proposta de desconstrução da noção de equivalência em Catford, Nida, Lefevere e Toury, Tese de doutorado apresentada à Universidade Estadual de Campinas, UNICAMP, Brasil, 1998.

SCOTT. M. **WordSmith Tools**, Oxford: Oxford University Press, 1996.

TAGNIN, S.E.O. **O Jeito que a Gente Diz**, São Paulo: Disal, 2005.

TEIXEIRA, E. D. Compilação do corpus comparável inglês-português de **Culinária** do Cortec. http://www.fflch.usp.br/dlm/comet/consulta_cortec.html, 2005.

ZANETTIN, F. DIY Corpora: The WWW and the Translator. In MAIA, B, HALLER, J. & ULRICH, M. (eds.) **Training the Language Service Provider for the New Millennium**. Porto: FLUP, p. 239-248, 2002.