

# Uso de *corpora* customizados para aperfeiçoar o texto traduzido

Ana Julia Perrotti-Garcia

Faculdades Montessori de Ibiúna – (FMI); Unibero/ Faculdades Anhanguera;  
Universidade de Franca (UniFran); Mestranda LAEL PUC-SP

Caixa Postal 11746 CEP 05049-970 São Paulo SP

anajulia@corpuslg.org

**Abstract.** *To translate is to make choices. The precision of these choices is directly related to the quality of search sources used by the translator. Using the Internet as a source of information is a simple and affordable method, but the results can lack precision accuracy. Ready corpora seem to be more reliable, although the majority of them are not composed by texts of specialty language. The author proposes the establishment of customized corpora – attending to specific research needs (register, linguistic variants, types of documents, target public, among others). Customized corpora, together with WordSmith tools set, will help translators make the final text more natural and precise, in a quick and cheap way. Customized corpora may be modified, enriched or altered whenever necessary, in order to be used in a new translation project, or even to continue a current project.*

**Keywords.** *Customized corpora; corpus linguistics; translation; WordSmith tools.*

**Resumo.** *Traduzir é fazer opções. A precisão dessas opções está diretamente relacionada com a qualidade das fontes de pesquisa utilizadas pelo tradutor. Usar a rede mundial de computadores (Internet) como fonte de pesquisa, embora seja um método simples e de baixo custo (que vem se difundindo entre a comunidade de tradutores), nem sempre é uma forma segura e garantida de resultados precisos e naturais. A pesquisa em corpora prontos tem mostrado ser uma alternativa mais confiável. Contudo, para textos científicos, praticamente não existem corpora prontos que respondam às principais dúvidas lexicais dos tradutores. Sendo assim, a autora propõe que sejam coletados corpora customizados – adaptados ao público-alvo do texto meta; levando em consideração, durante a coleta, outros aspectos como: registro; campo semântico (e suas subáreas); variações regionais, diafásicas, diacrônicas e dialógicas; entre outros. Assim, de posse de um corpus customizado, lançando mão de um programa computadorizado apropriado (WordSmith tools, mesmo na versão gratuita, para demonstração), o tradutor irá conseguir produzir textos mais naturais, precisos e adequados, sem aumentar muito o tempo de pesquisa, com um custo baixo e acessível.*

**Palavras-chave.** *Corpora customizados; lingüística de corpus; tradução; WordSmith tools.*

## 1. Introdução

Traduzir é fazer opções. A precisão dessas opções está diretamente relacionada com a qualidade das fontes de pesquisa utilizadas pelo tradutor. O avanço qualitativo representado pelo advento dos computadores e da Internet é inegável. E o trabalho do tradutor certamente foi influenciado positivamente pela introdução dos motores de busca, pela pesquisa na rede mundial de computadores e pela facilidade de acesso a um conteúdo vasto e constantemente atualizado. Entretanto, muitas vezes o tradutor precisa de muito mais do que velocidade e variedade de informações. O processo tradutório envolve pesquisas em assuntos técnicos, científicos e, muitas vezes, é preciso pesquisar assuntos quase inéditos, ou apenas abordados em sites muito específicos.

A fim de comparar as opções oferecidas ao tradutor ao pesquisar na Internet e aquelas fornecidas pelos *corpora* customizados<sup>1</sup>, faremos a seguir uma análise empírica dos resultados obtidos com cada uma dessas fontes.

## 2. Internet e motores de busca

Usar a rede mundial de computadores (Internet) como fonte de pesquisa, embora seja um método simples e de baixo custo (que vem se difundindo entre a comunidade de tradutores), nem sempre é uma forma segura e garantida de resultados precisos e naturais. A variedade e a quantidade crescente de textos disponibilizados pela rede mundial de computadores são dois aspectos que certamente devem ser levados em consideração, ao escolher a Internet como fonte para pesquisas, ao realizar trabalhos de tradução. A rapidez com que os resultados são apresentados é extremamente interessante, promovendo uma maior agilidade ao processo de tradução.

Por outro lado, o tradutor precisa ponderar muito bem, para não ser seduzido por essa grande quantidade de textos disponíveis que, algumas vezes, em sentido inverso, acaba comprometendo a qualidade dos resultados das pesquisas – pois oferece opções heterogêneas, com mistura de registros, variantes lingüísticas, público alvo. Além disso, sem menções claras ao autor, local e data da redação e da publicação implicando na ausência de comprometimento com a qualidade e a precisão de alguns textos.

Partiremos da seguinte situação verídica e muito freqüente no dia-a-dia do profissional que atua com textos da área de saúde: um tradutor precisa fazer uma tradução português – inglês de um resumo (*abstract*) de um artigo médico para ser publicado em uma revista científica que tem, entre outras exigências, como requisito o uso de inglês britânico. Para simplificar nossa explanação, faremos a análise de um termo, entre os muitos que foram pesquisados durante o processo de tradução. Como o artigo referia-se a enxertos ósseos, faremos a análise de um termo composto extremamente representativo do tema: sobrevivência do enxerto. Para a tradução, o termo sugerido seria *graft survival*, e este foi então pesquisado pelo motor *Google*. Ao usar esse sistema de busca, a avaliação dos resultados obtidos deve sempre ser feita com bom-senso e com rigor científico, pois da escolha feita irá depender diretamente a

qualidade do texto traduzido. Assim, foi feita a pesquisa do termo *graft survival* na Internet, pelo buscador *Google*, e de posse dos resultados obtidos (figura 1), passaremos à análise dos resultados.



Figura 1.

Primeiros resultados obtidos para o termo *graft survival*

(Internet, buscador Google).

A análise da Figura 1 revela que a pesquisa pelo termo *graft survival*, feita em 23 de junho de 2007, produziu 644 mil resultados (*hits*). O número parece bastante convincente, para servir de fiador de uso, se fizermos um julgamento estritamente quantitativo. Entretanto, convém lembrar que os textos apresentados ao pesquisar na Internet, seja qual for o motor de busca ou sistema de pesquisa utilizado, podem ser extremamente valiosos, mas também podem ser inúteis, enganosos e mal redigidos, uma vez que não há qualquer tipo de seleção ou pré-requisito para a publicação de textos na *web*.

Passemos, então, à análise dos resultados da busca. Por questões de restrição de espaço e de tempo (e até porque, há uma crença, anedótica e empírica, mas que alguns tradutores costumam considerar com fundamentada, que afirma que “os primeiros resultados de uma pesquisa na Internet são os de melhor qualidade”), analisaremos os primeiros quatro resultados obtidos:

**U.S Transplant:** *Website* redigido em inglês americano, e textos dirigidos a pacientes (os textos são apresentados em linguagem leiga e os termos mais técnicos, quando citados, são explicados ou facilitados).

**2003 OPTN/SRTR Annual Report** Reprodução do relatório anual de um órgão estadunidense responsável pela localização de doadores e pela rede de transplantes naquele país (OPTN/SRTR). Aqui podemos notar a heterogeneidade dos textos

apresentados na Internet, pois este resultado e o anterior são procedentes do mesmo *website*, no entanto têm público-alvo diverso, são gêneros textuais diferentes e, a menos que o tradutor faça uma leitura mais aprofundada (indo além das poucas linhas apresentadas pelo buscador), muitas dessas informações podem passar despercebidas.

**NEJM:** A revista médica *New England Journal of Medicine*, uma das mais tradicionais publicações da área da saúde, certamente é uma fonte de textos bem redigidos, confiáveis e que podem ser aproveitados como base de pesquisa para tradutores. Entretanto, é conveniente lembrar que neste texto temos um vocabulário dirigido aos médicos, o que não pode ser esquecido, principalmente se o texto a ser traduzido tiver como destino leitores não médicos. Além disso, por exigência da própria revista, a variante adotada nos textos é o inglês americano.

**Hindu on Net:** este *hit* traz uma entrevista concedida por um médico japonês a um repórter indiano. Sendo assim, temos no mínimo, influências regionais bem marcantes. Mesmo partindo do pressuposto que o médico entrevistado seja (o que deve realmente ser) fluente em inglês, é importante ponderar quanto à conveniência de usar este material como guia para a produção de um texto em inglês, uma vez que muito já se falou sobre a influência do falante não nativo sobre a qualidade e a naturalidade do texto produzido.

Assim, podemos observar, em uma primeira análise rápida, que os quatro primeiros resultados da pesquisa não apontam para um horizonte muito animador: textos em inglês americano, lado a lado com textos produzidos por não nativos, variedade de público alvo, de registro e de tipo textual. Seria conveniente que cada tradutor, antes de eleger a *web* como sua fonte única ou principal de informações lingüísticas, refletisse um pouco mais sobre a qualidade e a adequação dos textos que circulam nesse oceano imenso de opções quase intangíveis.

### 3. *Corpus* customizado

A pesquisa em *corpora* prontos<sup>2</sup> (por exemplo, apenas para citar alguns dos mais relevantes, *BNC* e *Collins Cobuild*, para pesquisas de inglês, e Banco de Português e *LacioWeb*, para pesquisas de português) tem mostrado ser uma alternativa muito mais confiável. Os textos são pré-selecionados, vindo em sua maioria de fontes fidedignas e que cuja origem e data de produção e coleta costumam ser explicitadas para o pesquisador. Contudo, para textos científicos, praticamente não existem *corpora* prontos que respondam às principais dúvidas lexicais dos tradutores. Sendo assim, para esta análise comparativa, coletamos um *corpus* customizado – adaptado ao público-alvo do texto meta (especialistas em gastroenterologia, neste caso) – levando em consideração, outros aspectos como: registro; campo semântico (e suas subáreas); variações regionais (apenas foram coletados textos em inglês britânico), diafásicas (coletamos textos de artigos médicos, publicados em revistas especializadas e conceituadas), diacrônicas (somente foram coletados textos produzidos na última década) e dialógicas; entre outros (Perrotti-Garcia, 2007). Assim, de posse de um *corpus* customizado, lançando mão de um programa computadorizado apropriado (*WordSmith tools*, na versão 4.0, para demonstração), repetimos a busca, desta vez usando apenas o termo *survival* (um procedimento que, como sabemos, se fosse feito pelo buscador *Google* produziria um resultado tão heterogêneo, numeroso e impreciso que seria impossível de ser

aproveitado da maneira que fizemos aqui). Os resultados da pesquisa feita pela ferramenta *Concordance* do programa *WordSmith* são apresentados na Figura 2.

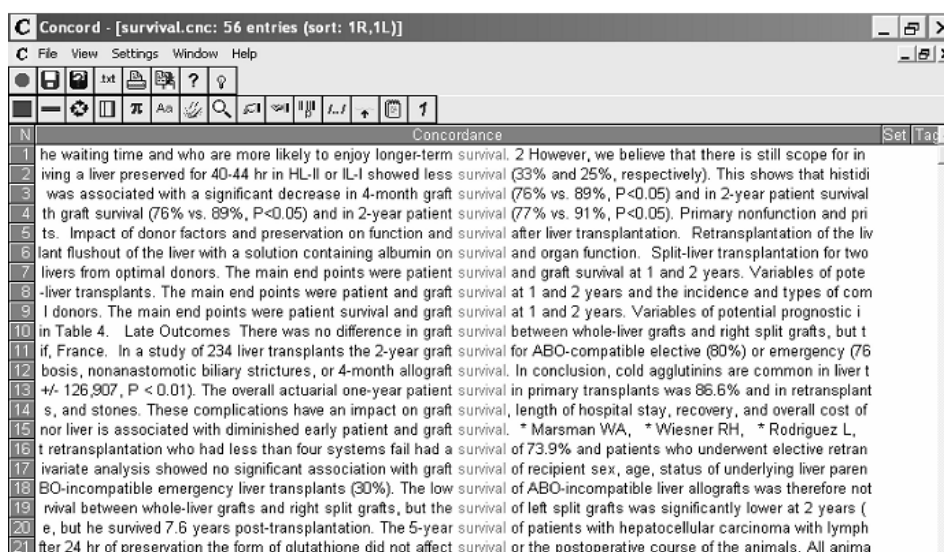


Figura 2. Primeiros resultados obtidos para o termo *survival*

(Programa *WordSmith*, ferramenta *Concordance*, a partir de *corpus* customizado).

Mesmo para o observador que avalia a figura 2 pela primeira vez, já é possível notar algumas diferenças marcantes entre os resultados obtidos com a ferramenta *Concordance*: a palavra de busca (aqui chamada de nóculo) aparece centralizada, grafada de cor diferente, e com um número fixo de palavras a esquerda e à direita. A esse conjunto, chamamos “linhas de concordância”. A ferramenta *Concordance* mostra as linhas de concordância existentes com a palavra de busca selecionada. Além disso, é possível aumentar ou diminuir o número de palavras à direita e à esquerda do nóculo, bastando para isso um simples ajuste no programa ou um clique em um botão determinado.

Além das diferenças estruturais nos resultados obtidos, podemos notar diferenças qualitativas importantíssimas: como os textos pesquisados foram pré-selecionados pelo próprio tradutor, seguindo critérios rígidos, não há termos suspeitos, todos os artigos foram redigidos em inglês britânico e o público alvo é sempre o médico especialista. Deste modo, notamos que as diversas linhas de concordância fornecem muito mais do que uma mera confirmação do uso de um termo pesquisado. Podemos depreender o uso de preposições, padrões colocacionais, termos compostos e, obviamente, vários exemplos autênticos de uso do termo pesquisado *graft survival*, além de termos como *patient survival*, *retransplantation* e *liver transplant*, que certamente serão úteis durante o processo de tradução.

#### 4. Internet X *corpus* customizado - o que concluímos?

Ao fazermos a comparação entre os resultados obtidos pelos motores de busca, na Internet, e pelo programa *WordSmith*, em *corpora* customizados, percebemos que há diferenças marcantes, não apenas do ponto de vista estrutural (formatação, apresentação

do resultados, organização das linhas de concordância), mas também diferenças na qualidade dos resultados obtidos. Como o *corpus* customizado é composto exclusivamente por textos coletados em fontes consideradas fidedignas, seguindo parâmetros relacionados aos aspectos textuais, que estarão em harmonia com o texto de chegada, os resultados são absolutamente precisos e adequados para ir ao encontro das necessidades de pesquisa do tradutor. Ainda muito precisa ser analisado sobre este tema tão fascinante: os *corpora* customizados como fonte de pesquisa para tradutores e o assunto certamente ainda será explorado futuramente. Assim, o tradutor irá conseguir produzir textos mais naturais, precisos e adequados, sem aumentar muito o tempo de pesquisa, com um custo baixo e acessível.

## Notas

<sup>1</sup> Para obter maiores informações sobre a coleta e a montagem de *corpora* customizados, recomendamos a leitura de Perrotti-Garcia (2005)

<sup>2</sup> Para obter os endereços de acesso aos *corpora* citados, consulte o item **6. Anexo(s)**.

## Referências

2006 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: Transplant Data 1996-2005. Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation, Rockville, MD.

BERBER SARDINHA, Tony. *Linguística de Corpus*. Manole, São Paulo, 2004.

PERROTTI-GARCIA, Ana Julia. *O Uso de Corpus Customizado como Fonte de Pesquisa para Tradutores*. Confluências Revista de Tradução Científica e Técnica, Lisboa, v. 3 p. 62-79. 2005. Disponível em: <<http://www.confluencias.net/>> Acesso em 28 set. 2007

\_\_\_\_\_. *Customised corpora – a source of information for translators*, In: Institute of Translation and Interpreting Conference. 2007. Anais eletrônicos. Disponível em: <[http://www.iti-conference.org.uk/conference\\_programme.html](http://www.iti-conference.org.uk/conference_programme.html)> Acesso em 28 set. 2007

## 6. Anexo

### Endereços eletrônicos dos *corpora* e programas citados no artigo:

Banco de Português: <<http://www2.lael.pucsp.br/corpora/bp/conc/index.html>>

British National Corpus (BNC): <<http://www.natcorp.ox.ac.uk/>>

Collins Cobuild: <<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>>

Lacio Web <<http://www.nilc.icmc.usp.br/lacioweb/>>

WordSmith tools: <<http://www.lexically.net/wordsmith/>>