

A TRADUÇÃO DO ITEM LEXICAL *EVIDENCE*: UMA ANÁLISE COM BASE EM *CORPUS*

Norma Andrade da Silva
Universidade Federal de Santa Catarina
normasil@gmail.com

RESUMO

Este artigo visa ao estudo das escolhas lexicais feitas pelos tradutores do item lexical *evidence* com base na Lingüística de *Corpus*. Para tanto, foram analisados excertos de textos literários que contemplam o par lingüístico inglês (língua-fonte) e português (língua-alvo) extraídos da Linguateca, cujas organizadoras são Ana Frankenberg-Garcia e Diana Santos. A Linguateca propicia pesquisas com base em recursos na área do processamento computacional da língua portuguesa. Um desses recursos é o COMPARA, um *corpus* eletrônico paralelo, extensível, disponível gratuitamente na Web, e que auxilia o trabalho de tradutores, de lexicógrafos e de professores de língua estrangeira.

PALAVRAS-CHAVE: *Evidence*. Lingüística de *Corpus*. Linguateca. COMPARA.

ABSTRACT

This article aims at the analysis of the lexical choices made by translators of the lexical item *evidence* based on *Corpus* Linguistics. Excerpts of literary texts (translated from English to Portuguese) were analysed. They were extracted from Linguateca, whose organizers are Ana Frankenberg-Garcia and Diana Santos. It provides resources based on computational processing of Portuguese. One of these resources is COMPARA, an electronic parallel *corpus*, extensible, free in Web, that helps the work of translators, lexicographers and teachers of foreign languages.

KEYWORDS: *Evidence*. *Corpus* Linguistics. Linguateca. COMPARA.

Introdução

O surgimento do computador deu início a uma revolução nos estudos lingüísticos, procurando respostas às perguntas sobre como a língua é organizada, como deve ser estudada e ensinada. Com os avanços obtidos pelos pesquisadores na área da Lingüística, pode-se afirmar que o computador pessoal “com memória poderosa e capacidade de armazenamento, começa a desempenhar, nas ciências humanas, o papel transformador que o telescópio teve na física e nas ciências exatas” (BERBER SARDINHA, 2004, p.xvii). Kenny (2001, p. xiii) compara o computador, além da comparação já feita com o telescópio e com o microscópio, a um caleidoscópio. “Kaleidoscopes allow us to view patterns, and to change those patterns at will”.

Segundo Borba (2002, p. 410), *corpus* é “conjunto de dados sobre determinado assunto”. Conforme o dicionário Novo Aurélio Século XXI (1999 p. 561), *corpus* é “conjunto de documentos, dados e informações sobre determinada matéria”. O Dicionário Houaiss (2001, p. 844) registra a seguinte acepção para o termo *corpus*: “4 LING conjunto de enunciados numa determinada língua, ger. colhidos de atos reais da fala, que servem como material para análise lingüística”. Já *corpus* eletrônico é todo *corpus* coletado e armazenado em arquivo na memória de computador (este tanto pode ser pessoal como de empresas). O primeiro *corpus* eletrônico, o *Brown Standard Corpus*¹, foi lançado em 1964, e continha um milhão de palavras. Naquela época, havia uma dificuldade muito grande para se informatizar todos esses textos, pois a tecnologia ainda não estava tão avançada como nos dias atuais. Hoje o computador é utilizado para coletar, armazenar, organizar e analisar as ocorrências de determinados itens lexicais. São aproveitados textos retirados tanto da *Internet* quanto de outras fontes como livros, jornais, revistas, correspondências. De acordo com Rocha,

os organizadores de um *corpus* devem preocupar-se em incluir outros tipos de textos, tais como artigos científicos, livros, revistas especializadas, correspondência empresarial, publicidade, textos acadêmicos e linguagem jurídica, além de, idealmente, material menos formal, como redações escolares e, sobretudo, língua falada (2003, p. 147).

A Lingüística de *Corpus*, no conceito de Berber Sardinha (2004), “ocupa-se da coleta e da exploração de corpora, ou conjunto de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador”. Ele afirma ainda que, embora os corpora sejam de inegável valor para os Estudos da Tradução, não se atingiu a importância que McEnery e Wilson haviam previsto na década de 1990.

Atualmente, a Lingüística de *Corpus* tem seus estudos mais avançados sendo feitos na Europa, notadamente na Grã-Bretanha, Noruega, Itália e Portugal. Em algumas universidades, há analistas que se dedicam às pesquisas baseadas em *corpora* computadorizados para descrever e registrar os aspectos mais variados da linguagem.

No Brasil, há estudos sendo feitos por Adriana Pagano, Stella Tagnin, Tony Berber Sardinha e outros.

Todos esses estudos que estão sendo feitos pelos pesquisadores da Lingüística de *Corpus* têm por objetivo auxiliar os tradutores a encontrar uma forma de resolver muitos problemas de ordem semântica. Também visam desenvolver uma tecnologia confiável para a tradução de máquina.

Como afirma Mona Baker (1995, p. 223),

Corpus-based research has become widely accepted as a factor in improving the performance of machine translation systems, and corpus-based terminology compilation is now the norm rather than the exception. [...] It is becoming increasingly important to take stock of what is happening on this front and to start working towards the development of an explicit and coherent methodology for corpus-based research in this discipline [translations studies].

Linguateca e o COMPARA

Em Portugal, Ana Frankenberg-Garcia é pesquisadora na área de Linguística Aplicada em Lisboa e junto com Diana Santos, coordenadora do projeto Processamento Computacional da Língua Portuguesa (www.linguateca.pt), são responsáveis pela criação do COMPARA, um *corpus* paralelo de inglês e português com acesso gratuito na Web. A criação e a organização da Linguateca têm financiamento da Fundação para a Ciência e Tecnologia (FCT), um órgão do Ministério da Ciência e do Ensino Superior. O modelo de funcionamento da Linguateca é descrito pelo modelo de Informação-Recursos-Avaliação (IRA). Em Recursos, foi criado um acesso pela Internet a um conjunto de recursos para a engenharia da linguagem em português. Os recursos mais utilizados são o COMPARA, o Corpógrafo, o CETEMPúblico e o AC/DC.

Até 1º de abril de 2007, o Brasil foi o país que teve o maior número de acessos acumulados ao *site* da Linguateca (499.298), deixando Portugal em segundo lugar em acessos acumulados (382.557) e, também, suplantando os países dos demais continentes (216.766). Diante desses dados estatísticos, constata-se que, no Brasil, está havendo um crescimento nas pesquisas com base em *corpus*.

O COMPARA é de extrema importância para os analistas de tradução que trabalham com textos em inglês alinhados com suas traduções para o português e vice-versa. Apresenta variantes linguísticas, como o português de Portugal, do Brasil, de Moçambique e de Angola, e o inglês da Grã-Bretanha, dos Estados Unidos e da África do Sul. Há também diversas possibilidades de busca: de palavra ou de expressão, direção de procura (inglês-português ou português-inglês), seleção da variante ou combinação de variantes, seleção de textos por data de publicação, ou ainda por autor ou tradutor.

O *corpus* está em constante expansão, havendo, no momento, mais de um milhão de palavras. O seu objetivo é o de incorporar um número cada vez maior de textos e suas respectivas traduções e disponibilizá-los pela *Internet* a todos os interessados nos estudos de

lexicografia ou de tradução. Serve para comparar os textos da língua-fonte com os da língua-alvo por meio de pesquisas automáticas. Ao se ter acesso ao COMPARA, há duas opções para se fazer a pesquisa por meio da interface DISPARA². Uma das opções é a Pesquisa Simples, que apresenta os resultados sobre o item lexical consultado em todos os textos; a outra é a Pesquisa Avançada, que permite refinar a busca por meio da seleção das partes do COMPARA que se pretende utilizar, como a seleção da língua, a restrição ou não das variantes lingüísticas e a condução de originais para traduções ou de traduções para originais.

Um exemplo do uso desses dois tipos de pesquisa foi feito com o item lexical *evidence*, do inglês para o português. Na Pesquisa Simples, foram registradas cinquenta e quatro ocorrências de *evidence*, e na Pesquisa Avançada, obteve-se trinta e três ocorrências para o mesmo termo. Nessa busca mais refinada, fez-se o direcionamento de originais para traduções e não houve restrição das variantes lingüísticas.

Após a pesquisa ser feita, as concordâncias aparecem em duas colunas, com o texto original à esquerda e sua tradução à direita. Os excertos estão alinhados lado a lado e as concordâncias aparecem em negrito no texto original, permitindo ao analista de tradução uma investigação mais acurada.

O *corpus* COMPARA ainda se encontra num estágio inicial. Há planos para a expansão de outros gêneros de textos. Por enquanto, há somente textos literários contemporâneos e antigos. Dessa forma, é possível compararem-se as traduções, ainda que feitas em épocas diferentes, de um mesmo texto original em inglês ou em português. Um exemplo são as traduções do romance *Iracema*, de José de Alencar. Uma delas foi feita em 1866³, e a outra, em 2000⁴.

As Tabelas 1 e 2, a seguir, comprovam o crescimento do conteúdo do COMPARA.

| COMPARA | Língua | Língua | Total |
|---------------|------------|---------|--------|
| Novembro 2001 | Portuguesa | Inglésa | |
| Originais | 7 | 3 | 10 |
| Traduções | 4 | 8 | 12 |
| Palavras | 187093 | 193548 | 380641 |

Tabela 1: Conteúdo do COMPARA em novembro de 2001.

| COMPARA Março 2005 | Língua Portuguesa | Língua Inglesa | Total |
|-----------------------|---------------------------------|---------------------------------|---------------------------------|
| Originais | 33 | 22 | 55 |
| Traduções | 24 | 34 | 58 |
| Palavras | 1096647 (originais e traduções) | 1183966 (originais e traduções) | 2280613 (originais e traduções) |

Tabela 2: Conteúdo do COMPARA em março de 2005.

Como ficou comprovado pelos números das tabelas acima, constatou-se que houve um aumento significativo tanto no número de palavras de língua inglesa quanto no de língua portuguesa. O aumento do número de palavras da língua portuguesa de novembro de 2001 para março de 2005 foi de 5,86 vezes; já o número de palavras de língua inglesa aumentou 6,11 vezes. Portanto, constata-se que houve um aumento do número de palavras de 5,99 vezes.

Um dos escritores, cujos fragmentos de textos estão disponíveis no COMPARA, é o britânico David Lodge, nascido em South London, em 28 de janeiro de 1935. Foi Professor Emérito de Literatura Inglesa na Universidade de Birmingham. O autor, crítico e novelista, ganhou alguns prêmios por suas obras, dentre elas *Changing Places*, *Nice Work* e *Therapy*. Algumas delas foram adaptadas para seriados de televisão. Obras de Lodge constantes do COMPARA: *Changing Places* (1988), *How Far Can You Go* (1980), *Paradise News* (1991), *Therapy* (1995) e *Nice Work* (1988). *Therapy* (*Terapia*) e *Chaging Places* (*Invertendo os Papéis*) foram traduzidas para o português do Brasil por Lídia Cavalcante-Luther, pela Editora Scipione.

A tradução do item lexical evidence

De acordo com o dicionário monolíngüe inglês *Webster's* (1985, p. 430), o termo *evidence* tem as seguintes acepções:

1 a: an outward sign: INDICATION **b:** something that furnishes proof: TESTIMONY; *specif:* something legally submitted to a tribunal to ascertain the truth of a matter **2:** one who bears witness; *esp:* one who voluntarily confesses a crime and testifies for the prosecution against his accomplices [...].

No *Moderno Dicionário Michaelis Inglês* (2000, p. 239), encontram-se as seguintes acepções para o termo traduzido para a língua portuguesa:

n 1 evidência, prova, indício [...] **2** sinal, indicação, mostra [...] **3** testemunho, depoimento de testemunha principal [...] **4** testemunha [...].

Já no *Dicionário de Usos do Português do Brasil- DUP* (2002, p. 658), um dicionário todo baseado em *corpus*, encontraram-se as seguintes acepções para o item lexical **evidência**:

Nf [**Abstrato de estado**] **1** aquilo que é incontestável ou evidente [...] **2** indício; sinal [...] **3** prova [...] **4** certeza [...] **5** características, traços [...] **6** proeminência, saliência [...] **7** destaque [...]

Para se investigar como o item lexical *evidence* foi traduzido para a língua portuguesa (constam trinta traduções de português de Portugal e três do português do Brasil), fez-se uma busca no COMPARA 6.1 (<http://www.linguateca.pt/COMPARA/>). Os resultados da busca mostraram como os tradutores das duas variantes lingüísticas traduziram o substantivo de modos diferentes.

Por se tratar de um exame de apenas um item lexical, todos os comentários sobre as escolhas feitas pelos tradutores seguiram a teoria lingüística de Catford (1980). Segundo ele, ao se traduzir um texto, há a “substituição de material textual numa língua (LF) por material textual equivalente noutra língua (LM)”.

Em alguns dos excertos analisados, pode-se observar que os tradutores fizeram essa substituição, como acontece no exemplo 1, do português de Portugal (PP), e no exemplo 2, do português do Brasil (PB), para a novela *Therapy*. O termo *evidence*, que está sendo analisado, foi traduzido como *vestígios* e *evidência*. Os tradutores mantiveram a mesma classe gramatical, substantivo. Todos os níveis da LF foram substituídos por níveis da LM.

| | | |
|-----------------------|--|--|
| <u>EBDL1T1</u> (242): | <i>There's no evidence of torn meniscus or arthritic degeneration of the joint.</i> | Não há <i>vestígios</i> de entorse do menisco ou de degeneração artrítica na articulação. (PP/1) |
| <u>EBDL1T2</u> (242): | <i>There's no evidence of torn meniscus or arthritic degeneration of the joint.</i> | Não há <i>evidência</i> de meniscos dilacerados ou degeneração artrítica da junta. (PB/2) |

No exemplo 3, retirado de *Nice Work (Um almoço nunca é de graça)*, com uma tradução apenas para o português de Portugal, percebe-se que a tradutora Maria Carlota Pracana optou por uma mudança de classe gramatical: *evidence*, que na LF pertence à classe do substantivo, foi traduzido como verbo *provar* na LM.

| | | |
|-------------|--|---|
| EBDL2(531): | <i>As evidence for this assertion she will point to the fact that the rise of the novel (the literary genre of ‘character’ par excellence) in the eighteenth century coincided with the rise of capitalism; that the triumph of the novel over all other literary genres in the nineteenth century coincided with the triumph of capitalism; and that the modernist and postmodernist deconstruction of the classic novel in the twentieth century has coincided with the terminal crisis of capitalism.</i> | Para <i>provar</i> esta afirmação, realça os factos de o aparecimento do romance (o género literário da «personagem» <i>par excellence</i>), no século XVIII, ter coincido com o advento do capitalismo; de o triunfo do romance sobre todos os outros géneros literários, no século XIX, ter coincido com o triunfo do capitalismo; e de a desconstrução modernista e pós-modernista do romance clássico, no século XX, ter coincido com a crise final do capitalismo. (PP/3) |
|-------------|--|---|

Nos exemplos 4 e 5, retirados de *Changing Places*, que na tradução portuguesa, feita por Helena Cardoso, teve o título de *A Troca*, e na brasileira, traduzida por Lidia Cavalcante-Luther, teve o título como *Invertendo os Papéis*, observa-se que no primeiro caso o termo *evidence* foi traduzido como *prova*, e, no segundo, houve uma mudança de substantivo para o verbo *evidenciar*.

| | | |
|---------------|---|--|
| EBDL3T1(180): | <i>He was in fact a keen sportsman, and the walls of his room bore plentiful evidence of his markmanship in the form of silently snarling stuffed animals.</i> | Era de facto um desportista hábil e as paredes do quarto ostentavam a <i>prova</i> da sua perícia sob a forma de animais embalsamados, que rosnavam em silêncio. (PP/4) |
| EBDL3T2(180): | <i>He was in fact a keen sportsman, and the walls of his room bore plentiful evidence of his markmanship in the form of silently snarling stuffed animals.</i> | Ele era de fato um esportista aplicado e as paredes de sua sala <i>evidenciavam</i> à exaustão sua pontaria através dos vários animais empalhados com iminentes rosnados silenciosos. (PB/5) |

Ainda em *Changing Places* (exemplos 6 e 7), as duas tradutoras optaram por traduzir o substantivo pelos verbos *constatar* (6) e *ver* (7).

| | | |
|---------------|--|---|
| EBDL3T1(540): | <i>Morris had been born and brought up in New York, but he had no intention of returning there, in fact he would not repine if he never saw the city again: on the evidence of his last visit it was only a matter of time before the garbage in the streets reached penthouse level and the whole population suffocated.</i> | Morris fora nado e criado em Nova Iorque, mas não desejava voltar para lá e a verdade é que nem se ralava, se nunca mais visse a cidade: pelo que <i>constatara</i> quando da última visita, era só uma questão de tempo até que o lixo nas ruas chegasse aos últimos andares e a população inteira sufocasse. (PP/6) |
| EBDL3T2(540): | <i>Morris had been born and brought up in New York, but he had no intention of returning there, in fact he would not repine if he never saw the city again: on the evidence of his last visit it was only a matter of time before the garbage in the streets reached penthouse level and the whole population suffocated.</i> | Morris nasceu e cresceu em Nova York, mas não tinha a menor intenção de ver aquela cidade de novo: a julgar pelo que <i>viu</i> em sua última visita, ia ser apenas uma questão de tempo para que o lixo das ruas da cidade alcançasse o nível das coberturas e a população inteira morresse asfixiada. (PB/7) |

No exemplo 8, *Paradise News (Notícias do Paraíso)*, traduzido para o português de Portugal, o tradutor Carlos Grifo Babo manteve a mesma classe gramatical (substantivo).

| | | |
|--------------|---|--|
| EBDL5(1117): | <i>The fact that he and his son are not ordinary tourists makes the distraction doubly irritating, for it is not as if he is gleaning from it any anecdotal evidence relevant to his research.</i> | O facto de este e o respectivo filho não serem turistas comuns torna a perturbação duplamente irritante, pois nem sequer lhe é possível extrair dali qualquer <i>corroboração</i> anedótica com relevância para a sua pesquisa. (PP/8) |
|--------------|---|--|

Conclusão

O objetivo dessa pesquisa foi o de comparar as traduções alinhadas do inglês e do português e observar como o item lexical *evidence* foi traduzido, qual a escolha lexical feita pelos tradutores. Observou-se ainda que para um mesmo texto houve, por parte dos tradutores, escolhas diversas. Acredita-se que essa diversidade de significados para uma mesma palavra seja a falta de dicionários baseados em *corpus* com ocorrências lexicais

comprovadas pelo uso. Como os dicionários apresentam exemplos criados pelos lexicógrafos, e não pelo uso corrente dos termos, tem-se então essa variedade de traduções.

O computador revolucionou os estudos da tradução, pois forneceu e ainda fornece subsídios para que os tradutores possam desempenhar seu trabalho de forma mais segura ao consultar dicionários, monolíngües ou bilíngües, baseados em dados lingüísticos autênticos.

NOTAS

¹ Brown University Standard Corpus of Present-day American English.

² Interface DISPARA foi concebida, em primeiro lugar, para dar acesso ao *corpus* COMPARA.

³ Iracema, the honey lips: a legend of Brazil, tradução feita por Lady Isabel Burton. London, Bickers, 1866.

⁴ Iracema, tradução de Clifford Landers. New York, Oxford Press, 2000.

REFERÊNCIAS

BAKER, Mona. *Corpora in Translation Studies: an overview and some suggestions for future research*. Target 7: 1995.

BERBER SARDINHA, T. *Lingüística de Corpus*. Barueri, São Paulo: Manole, 2004.

BORBA, Francisco S. *Dicionário de Usos do Português do Brasil – DUP*. São Paulo: Ática, 2002.

CATFORD, J. C. *Uma Teoria Lingüística da Tradução*. São Paulo: Cultrix; Campinas: Pontifícia Universidade Católica de Campinas, 1980.

FERREIRA, Aurélio Buarque de Holanda. *Novo Aurélio Século XXI: o dicionário da língua portuguesa*. 3. ed. Rio de Janeiro: Nova Fronteira, 1999.

HOUAISS, Antônio; VILLAR, Mauro de Salles. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva, 2001.

KENNY, Dorothy. *Lexis and Creativity in Translation*. Manchester: St. Jerome Publishing, 2001.

Michaelis: moderno dicionário inglês-português, português-inglês. São Paulo: Companhia Melhoramentos, 2000.

ROCHA, Marco Antonio Esteves da. *O corpus computadorizado em Lexicografia: guia do consumidor*. (In: *Todas as Trilhas: pesquisas e projetos*./Organizado por Bento Carlos Dias da Silva.) Araraquara Laboratório Editorial /FCL /UNESP; São Paulo: Cultura Acadêmica Editora, 2003.

WEBSTER'S Ninth New Collegiate Dictionary. Massachussets: Merriam-Websters Inc., 1985.

www.linguateca.pt. Acesso em 2005.

BIBLIOGRAFIA ON-LINE (sugerida)

<http://www.ifi.unizh.ch/cl/juste/postworkshop/repository/fzanettin.pdf>

<http://www.ifi.unizh.ch/cl/juste/postworkshop/repository/lbowker.pdf>

<http://www.federicozanettin.net/DIYcorpora.htm>

<http://www.comp.lancs.ac.uk/computing/research/ucrel/papers/techpaper/vol2.pdf>

<http://www.es.queensu.ca./achall97/papers/a006.html>

<http://www.art.man.ac.uk/SML/ctis/events/Conference2000/corpus3.htm>

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.