

ANÁLISE DE UM CORPUS PARALELO CONSTITUÍDO DE TEXTOS NA SUBÁREA DE CARDIOLOGIA

Profa. Dra. Diva Cardoso de CAMARGO (Universidade Estadual Paulista)

Paula Tavares Pinto PAIVA (Universidade Estadual Paulista; Centro Universitário do Norte Paulista)

ABSTRACT: We developed an analysis on a parallel corpus of texts taken from a bilingual journal of Cardiology. Based on Baker's proposals (1993, 1995, 1996) and using the software WordSmith tools, we examined words from medical terminology and also observed features of simplification and explicitation.

KEYWORDS: Translation Studies; Corpus Linguistics; Medical Translation; Simplification; Explicitation.

0. Introdução

O número de revistas bilíngües na área da saúde tem aumentado nos últimos anos tornando esse campo de estudos importante tanto para tradutores quanto para pesquisadores em tradução. Dentro desse contexto, iniciamos uma análise de um corpus composto por textos médicos, traduzidos do português para o inglês na subárea de cardiologia. Esta investigação toma como base o arcabouço teórico-metodológico proposto por Mona Baker (1993, 1995, 1996), destinado a pesquisas em tradução baseadas em corpora em formato eletrônico, e também as investigações de Berber Sardinha (2004), voltadas para a lingüística de corpus.

1. Perspectiva Teórica

Com base nos estudos descritivos da tradução e na lingüística de corpus, Baker (1993, 1995, 1996) lança uma proposta em que considera a tradução como um fenômeno em si, direcionando as pesquisas para a linguagem dos textos de chegada (TCs), uma vez que as traduções costumam apresentar traços característicos que, em geral, não são encontrados nos textos de partida (TPs). A autora sugere que grandes corpora permitirão que o objeto de estudo da tradução seja explorado

numa escala muito maior e por meio do auxílio de ferramentas automáticas, que poderão contribuir para uma melhor compreensão do fenômeno tradutório.

Por meio da metodologia fornecida por Baker, podemos analisar princípios que governam o comportamento tradutório. Em nosso trabalho, observamos dois dos traços característicos de tradução apresentados por Baker (1996): a explicitação e a simplificação.

A explicitação é definida como uma “tendência geral de explicar termos que estariam implícitos nos TPs”¹ (BAKER 1996:180). Essa tendência pode ser encontrada em ocorrências como, por exemplo, o tamanho maior do TC em relação ao TP. Outro fator a ser considerado na observação de traços explicitadores são evidências lexicais, expressas por meio de conjunções e locuções conjuntivas explicativas e conclusivas, as quais apareceriam mais nos TCs que nos TPs como, por exemplo: *cause, reason, due to, lead to, because, therefore, consequently* etc.

A simplificação é definida como “a tendência de tornar mais simples a linguagem usada na tradução”² (BAKER 1996: 180) para facilitar a compreensão do leitor da língua de chegada. A simplificação também pode ser observada nos TCs quando ocorrer em quebras de sentenças mais longas nos TPs. Mudanças na pontuação nos TCs também são características de simplificação. Por sua vez, a razão forma/ocorrência (*type/token ratio*) pode mostrar estratégias empregadas de forma consciente ou inconsciente, que poderiam ser identificadas pelo analista como traços de simplificação, as quais seriam decorrentes de uma menor variação de vocábulos nos TCs. Segundo Baker (1996), uma razão forma/ocorrência mais baixa nos TCs sugeriria um maior índice de repetições empregado pelo tradutor.

De acordo com Berber Sardinha (2004:26), nosso corpus de cardiologia, nessa etapa inicial da investigação, é de tamanho pequeno, com aproximadamente 50.000 palavras. É importante ressaltarmos que esse corpus é apenas uma parte de um corpus médio que está sendo compilado para a nossa pesquisa de mestrado. Para a construção do corpus, procuramos seguir os requisitos que Berber Sardinha (2004:19) destaca como sendo primordiais na construção de um corpus computadorizado: autenticidade, conteúdo escolhido criteriosamente, e representatividade.

Para esta etapa da pesquisa, levantamos a lista de frequência dos termos médicos mais usados ao longo de todos os TPs e TCs que constituem o corpus, bem como observamos a razão forma/ocorrência.

2. Método

Quanto ao *material* utilizado para a construção do corpus paralelo de cardiologia, coletamos quinze artigos publicados pela revista *Arquivos Brasileiros de Cardiologia*, entre os anos de 2000 e 2004. Para a escolha do *tradutor*, o critério foi o de selecionarmos um profissional com larga experiência na subárea analisada. Com base nesse critério, escolhemos os artigos traduzidos pela Dra. Stella Maris C. Gandour, que tem formação em medicina e mestrado em tradução pela PUC do Rio de Janeiro.

Quanto aos *procedimentos* adotados para este estudo, os artigos que foram publicados até o ano 2001 foram escaneados, limpos e salvos em forma de texto sem formatação (txt). Já os artigos de 2002 a 2004 foram diretamente salvos em txt porque a revista passou a ser disponibilizada na Internet. A seguir, os textos foram processados pelo programa WordSmith Tools versão 3.0. Primeiramente, utilizamos o aplicativo WordList que nos forneceu uma lista de frequência de palavras, da qual extraímos as dez mais recorrentes nos TPs e nos TCs. O próximo passo foi calcular a razão forma/ocorrência a fim de observarmos a variação lexical dos TPs e dos TCs. Tal variação nos auxilia nas análises do traço de simplificação e, de certa forma, também de explicitação. Em seguida, passamos à observação de aspectos referentes a mudanças na pontuação e a resoluções de ambigüidades nos TCs em relação aos TPs.

Como passo seguinte, fizemos uma comparação com alguns termos dos TPs e os possíveis correspondentes nos TCs; para tanto, utilizamos o aplicativo Concord que, ao gerar todas as linhas de concordância em que as palavras aparecem, permitiu-nos a observação das palavras em seu contexto tanto nos TPs quanto nos TCs. Por contexto, entende-se “o texto ao redor” da “palavra de busca” (BERBER SARDINHA, 2004: 105).

3. Discussão e resultados

Para a apresentação dos dez termos médicos mais frequentes nos TPs e TCs de cardiologia, elaboramos a tabela a seguir, com o número de ocorrências e sua porcentagem.

Corpus em Português (TPs)				Corpus em Inglês (TCs)			
WordSmith Tools				WordSmith Tools			
Word	Freq.	%		Word	Freq.	%	
ARTERIAL	213	0,40		VENTRICULAR	218	0,40	
HIPERTENSÃO	173	0,33		HEART	211	0,39	
MITRAL	132	0,25		CORONARY	198	0,36	
VENTRÍCULO	121	0,23		HYPERTENSION	164	0,30	
VENTRICULAR	118	0,22		PULMONARY	157	0,29	
ATRIAL	114	0,21		MITRAL	150	0,27	
PRESSÃO	91	0,17		PRESSURE	148	0,27	
SISTÓLICA	83	0,16		ARTERY	142	0,26	
ARTÉRIA	82	0,15		VALVE	127	0,23	
ATRESIA	82	0,15		ATRIAL	123	0,23	

Como podemos observar na tabela acima, alguns dos termos selecionados aparecem com uma frequência alta em ambas as tabelas. Este é o caso do termo “hipertensão” que apresenta 173 ocorrências nos TPs e *hypertension* com 164 ocorrências nos TCs. Outros termos como “ventricular”, “atrial”, “pressão” e “artéria” também aparecem com os seus respectivos correspondentes *ventricular*, *atrial*, *pressure* e *artery* entre os dez termos médicos de alta frequência nos TPs e nos TCs. No entanto, vale notar que todos esses termos ocorrem em número maior nos

TCs. Essa evidência poderia estar apontando para o traço de simplificação por meio da repetição de palavras nos textos traduzidos.

Outro caso a ser comentado é o de termos que contêm um dos seus elementos com função de adjunto adnominal, o qual apresenta diferenças significativas na frequência ao ser empregado como locução adjetiva, como por exemplo: “exame do coração”. Mesmo assim, no caso da palavra “coração” que registrou 54 ocorrências nos TPs, não mostrou uma frequência semelhante à da palavra *heart* (211 ocorrências) nos TCs, fato este que nos levou a observar as linhas de concordância conforme veremos mais adiante.

Outro aspecto para o exame dos TCs é o número de ocorrências e de formas em relação aos TPs. Nos TCs analisados, ocorre maior número de palavras/tokens (54.613) do que nos TPs (53.220). A maior incidência de palavras nos TCs confirma a hipótese da explicitação pela qual os TCs costumam ter mais palavras do que os TPs. Já o número de formas/types nos TCs é menor (5.589) que nos TPs (7.134), o que evidencia maior número de repetições e menor variação de vocábulos na tradução.

A razão forma/ocorrência de 13,40 nos TPs e de 10,23 nos TCs, também aponta para essa menor variação de palavras, confirmando o princípio de simplificação. Poderíamos inferir que a tradutora usou de estratégias, de modo consciente ou inconsciente, a fim de facilitar a compreensão do texto para os leitores de língua inglesa.

Outra evidência de traço de simplificação pode ser observada pela mudança na pontuação dos TCs. Pudemos identificar, em nosso subcorpus de tradução, a quebra de alguns parágrafos longos dos TPs deixando mais curtas as sentenças dos TCs como nos fragmentos a seguir:

Contudo, influenciaram de maneira diversa a elastância sistólica final que foi de $9,2 \pm 1,2\text{mmHg/mm}$ na elevação “transitória” e de $21,0 \pm 7,3\text{mmHg/mm}$ na elevação “sustentada” ($p < 0,05$). [grifo nosso]

However, they acted differently in end systolic pressure-diameter relation. It was greater in the SAH than TAH protocol, $21.0 \pm 7.3\text{mmHg/mm}$ vs. $9.2 \pm 1.2\text{mmHg/mm}$ ($p < 0.05$).

Como podemos notar, a tradutora substituiu o pronome relativo “que” do TP por um ponto final no TC, diminuindo o comprimento das sentenças a fim de tornar o texto mais claro para o leitor da tradução.

Por outro lado, em determinados trechos dos TCs, pudemos examinar traços de explicitação como, por exemplo, em:

A importância da dieta também foi demonstrada no estudo de Lyon-de prevenção secundária, onde a alocação a uma dieta mediterrânea foi capaz de reduzir aproximadamente 70% a incidência de novos infartos. [grifo nosso]

The importance of the diet has also been shown in Lyon Diet Heart Study of secondary prevention, where the use of a Mediterranean diet managed to reduce the incidence of new infarctions by approximately 70%.

No trecho acima, a tradutora tornou explícita uma informação que estava implícita no TP. Ela esclarece para o leitor do TC o que é o “estudo de Lyon”. Quando perguntamos o porquê dessa explicação para a tradutora, ela nos informou que esse estudo já era bem conhecido na literatura médica e que preferiu repetir o seu nome.

No tocante às linhas de concordância geradas pelo aplicativo Concord, nas quais essas palavras aparecem junto ao seu contexto, pudemos averiguar como os termos levantados inserem-se na linguagem médica. A diferença significativa na frequência da palavra *heart* deve-se à sua co-ocorrência, principalmente na função de adjunto adnominal em inglês como em *heartbeats*, *heart frequency*, *heart rate* e *heart failure*. Apresentamos, abaixo, uma amostra com dez linhas de concordância das palavras “coração” e *heart* nos corpora em português e em inglês.

Linhas de concordância da palavra “coração”

- 1 acrescentamos ao tipo balanceado, os *corações* nos quais, apesar da artéria
- 2 a pode afetar a capacidade contrátil do *coração* por razões de ordem bioquímicas
- 3 contornava o ápice do *coração* e ascendia no sulco interventricular
- 4 o aparecimento da doença isquêmica do *coração* e doenças
- 5 bem como os estudos radiológicos do *coração*, esôfago e cólons
- 6 na face diafragmática do *coração*, utilizando como referência
- 7 entre o sulco coronariano e o ápice do *coração* em terços superior,
- 8 as coronárias no suprimento arterial do *coração*. MÉTODOS:
- 9 hágsicos crônicos e sua repercussão no *coração*: estudo clínico e o
- 10 em preparações de *coração* isolado e no coração in situ, concluíram

Linhas de concordância da palavra *heart*

- 1 an incidence of 12.5%. In none of the *hearts* that we studied did the
- 2 and 26,000 deaths occurred due to *heart* failure in the year 2000.

3 8.1% to 44.1%) had evolved with chronic *heart* failure, and 35 (23.7%
4 and 67 had evidence of congestive *heart* failure or cardiomegaly, or
5 in a 10% formol solution. These *hearts* came from the Legal Medical
6 a standardized division of the *heart* into areas, the classification of
7 the absence of congenital or secondary *heart* disease. Four hundred
8 according to the Kiliip-Kimball *heart* failure classification was treated
9 years ago, mortality due to ischemic *heart* disease was greater in the
10 almost all patients hospitalized due to *heart* failure are hospitalized

4. Conclusão

O levantamento dos termos médicos mais frequentes neste estudo mostra como as palavras se organizam em sintagmas. A maioria dos termos levantados não tem significado independente, mas seus elementos se inter-relacionam gerando especificidades próprias de acordo com cada situação dentro do contexto da cardiologia.

Quanto aos traços de tradução propostos por Baker (1996), as observações feitas até o presente momento, com base nos princípios da lingüística de corpus e o auxílio de ferramentas eletrônicas, evidenciaram traços de simplificação, por meio de mudanças de pontuação e quebras de sentenças longas. Além disso, a razão forma/ocorrência confirma a presença de simplificação nos TCs analisados. Nossas observações também permitiram identificar traços de explicitação nas traduções que, em geral, tornam mais fácil a leitura dos TCs.

Notas:

1. *I take “explicitation” to mean that there is an overall tendency to spell things out rather than leave them implicit in the translation.*
2. *We can tentatively define “simplification” as the tendency to simplify the language used in translation.*

REFERÊNCIAS BIBLIOGRÁFICAS

BAKER, M. Corpus linguistics and translation studies: implications and applications. In: *Text and technology: in honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 1993. p.233-250.

_____ Corpora in translation studies: an overview and some suggestions for future research. *Target*, v. 7, n. 2, 1995. p. 223-243.

_____ Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, Harold(Ed). *Terminology, LSP and translation studies in language engineering in honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins Publishing Co., 1996. p. 177-186.

BERBER SARDINHA, *Linguística de Corpus*, SP: Manole, 2004.

MATERIAL USADO NO SUBCORPUS DE CARDIOLOGIA:

Arquivos Brasileiros de Cardiologia. Disponível em :
<http://www.scielo.br>