

## **ESTUDO BASEADO EM CORPORA PARALELOS DE TEXTOS MÉDICOS NAS SUBÁREAS DE CARDIOLOGIA E ORTOPEDIA, NA DIREÇÃO PORTUGUÊS/ INGLÊS**

Profa. Dra. Diva Cardoso de Camargo (Universidade Estadual Paulista)  
Paula Tavares Pinto Paiva (Universidade Estadual Paulista; União das Faculdades dos Grandes Lagos - Doutoranda em Estudos Linguísticos)

### **ABSTRACT**

Based on Baker's proposals (1993, 1995, 1996) and by using the software WordSmith Tools, we have been developing a research using two parallel corpora with approximately 170,000 words. Each corpus is composed of medical articles written in Portuguese and their translations into English, taken from, respectively, a scientific journal of Cardiology and of Orthopaedics. In this part of the research, we have extracted type/token ratio in order to examine the variation of specialized terms in both corpora. We have also observed explicitation and simplification, and identified some linguistic characteristics of two translators selected for our analysis.

**Keywords:** Translation, Translation of Cardiology and Orthopaedics Texts, Corpus-based Translation Studies, Corpus Linguistics.

### **RESUMO**

Tomando como arcabouço teórico-metodológico a proposta lançada por Baker (1993, 1995, 1996) e utilizando o programa WordSmith Tools, estamos desenvolvendo uma pesquisa em dois corpora paralelos com aproximadamente 170.000 palavras. Cada corpus é constituído por artigos médicos escritos em português e por suas respectivas traduções para o inglês, extraídos, respectivamente, de uma revista científica de Cardiologia e de outra de Ortopedia. Nesta etapa da pesquisa, geramos a razão forma/item (type/token ratio) e examinamos a variação de termos médicos especializados. Também observamos a ocorrência de traços explicitadores e simplificadores, e identificamos algumas marcas de comportamento linguístico dos dois tradutores selecionados para a análise.

**Palavras-chave:** Tradução, Tradução de textos nas Subáreas de Cardiologia e Ortopedia, Estudos de Tradução baseados em Corpus, Linguística de Corpus.

## Introdução

A área da tradução técnico-científica tem se notabilizado devido a sua contribuição para a rápida divulgação de pesquisas brasileiras em outros países por meio das publicações bilíngues de revistas especializadas, como, por exemplo, as de Medicina. Por sua vez, a tradução de textos médicos para o vernáculo tem trazido resultados positivos para os profissionais da área da saúde, possibilitando-lhes a aquisição de conhecimentos específicos e atualizados.

Embora tenha havido um número crescente de traduções na área médica, existe uma lacuna na literatura sobre os Estudos da Tradução, uma vez que são raras as pesquisas voltadas para esse tipo de texto.

Por essa razão, julgamos ser relevante um estudo que aborde as estratégias e tendências apresentadas pelos tradutores diante desse tipo de texto, bem como a elaboração de um léxico com os termos<sup>1</sup> mais frequentes nas subáreas de Cardiologia e Ortopedia na direção português/inglês.

Além de profissionais da área médica, nossa pesquisa poderá ajudar professores e alunos de tradução que queiram saber mais sobre a tradução de textos médicos, fornecendo-lhes um levantamento de termos médicos mais frequentes em português e seus correspondentes em inglês nas subáreas de Cardiologia e Ortopedia.

### 1. Perspectiva Teórica

Com base nos Estudos Descritivos da Tradução e na Linguística 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 por meio do auxílio de ferramentas automáticas, o que poderá contribuir para uma melhor compreensão do fenômeno tradutório.

Por meio da metodologia fornecida por Baker (1996), podemos analisar características que estão presentes nos textos traduzidos, observando dois dos traços de tradução propostos pela autora: a explicitação e a simplificação.

---

1. Neste estudo, “termo” é entendido como a “designação, por meio de uma unidade linguística, de um conceito definido em uma língua de especialidade” (ISO 1087, 1990, p. 5, *apud* BARROS, 2004).

A explicitação é definida como uma “tendência geral de explicar termos que estariam implícitos nos TPs”<sup>2</sup> (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”<sup>3</sup> (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/item (*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/item mais baixa nos TCs sugeriria um maior índice de repetições empregado pelo tradutor.

Para a construção dos dois corpora, 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.

## 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, o qual conta com 53.220 palavras nos TPs e 54.613 nos TCs. Por sua vez, os artigos do corpus paralelo de Ortopedia foram extraídos da *Revista Brasileira de Ortopedia*, nos anos de 2003 e 2004 e conta com 31.613 palavras nos TPs e 28.870 palavras nos TCs.

Para a escolha dos *tradutores*, o critério foi o de selecionarmos profissionais com larga experiência nas subáreas analisadas. Com base nesse critério, escolhemos os artigos de Cardiologia traduzidos pela Dra. Stella Maris C. Gandour, que tem formação em Medicina e mestrado em

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.

tradução pela PUC do Rio de Janeiro. Os artigos de Ortopedia foram traduzidos pelo Dr. Jacques Vissoky, que também atua tanto na área médica quanto na de tradução. Importante ressaltarmos que o Sr. Vissoky foi contemplado com o “Prêmio União Latina de Tradução Científica e Técnica”, de 2003, na Bienal do Rio de Janeiro.

Quanto aos *procedimentos* adotados para este estudo, os artigos de Cardiologia 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. Os artigos da revista de Ortopedia passaram pelo mesmo processo descrito primeiramente, uma vez que não se encontram disponíveis na Internet até o presente momento.

A seguir, os textos foram processados pelo programa WordSmith Tools versão 3.0. Utilizamos, em primeiro lugar, 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/item 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 ambiguidades 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 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, ou nóculo (Berber Sardinha, 2004: 105).

### 3. Discussão e resultados

Para a apresentação, primeiramente, dos dez termos médicos mais frequentes nos TPs e TCs de cardiologia, elaboramos a tabela 1 ilustrada na próxima página, com o número de ocorrências e sua porcentagem.

Como podemos observar, alguns dos termos selecionados aparecem com uma frequência alta em ambos os subcorpora. Este é o caso, por exemplo, 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 com frequência alta 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.

Tabela 1: Termos mais frequentes no Corpus Paralelo de Cardiologia

Corpus em Português (TPs)			Corpus em Inglês (TCs)		
Word	WordSmith Tools Freq.	%	Word	WordSmith Tools 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

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 esse que nos levou a observar as linhas de concordância conforme veremos mais adiante.

A tabela 2 que aparece na próxima página, apresenta os dez termos mais frequentes nos TPs e TCs de Ortopedia. Assim como no corpus de cardiologia, o corpus de Ortopedia também apresenta vários termos de alta frequência no português e no inglês, como em: "enxerto" e *graft* (97 ocorrências nos TPs e nos TCs); "fratura" (73 ocorrências) e *fracture* (58 ocorrências); "ósseo" (68 ocorrências) e *bone* (171 ocorrências); "cotovelo" e *elbow* (55 ocorrências nos TPs e nos TCs); "joelho" (42 ocorrências) e *knee* (58 ocorrências); "tendão" (33 ocorrências) e *tendon* (50 ocorrências). As diferenças em número de ocorrências apontam para a necessidade de observarmos os termos inseridos nas linhas de concordância, de modo similar ao realizado para o corpus de cardiologia. Esse é o caso de termos como o adjetivo "ósseo" (68 ocorrências) que pode também aparecer no feminino

"óssea" (52 incidências). O possível correspondente de "ósseo" seria bone que também pode se referir ao substantivo "osso" e que teve nos TPs 31 ocorrências. Somando-se as palavras "ósseo", "óssea" e "osso", chegaríamos a 161 ocorrências, número próximo ao das incidências de bone (171 ocorrências). Outro termo que nos chamou a atenção foi "patela" (46 vezes) nos TPs e patellar (101 ocorrências) nos TCs, que discutiremos mais adiante ao abordar as linhas de concordância.

Tabela 2: Termos mais frequentes no Corpus Paralelo de Ortopedia

Corpus em Português (TPs)			Corpus em Inglês (TCs)		
Word	WordSmith Tools Freq.	%	Word	WordSmith Tools Freq.	%
ENXERTO	97	0,31	BONE	171	0,59
FRATURA	73	0,23	PATELLAR	101	0,35
ÓSSEO	68	0,22	GRAFT	97	0,34
LESÃO	60	0,19	JOINT	83	0,29
COTOVELO	55	0,17	LIGAMENT	59	0,20
LUXAÇÃO	49	0,15	FRACTURE	58	0,20
ARTICULAR	48	0,15	KNEE	58	0,20
PATELA	46	0,15	ELBOW	55	0,19
JOELHO	42	0,13	TENDON	50	0,17
TENDÃO	33	0,10	ACETABULAR	41	0,14

Outro aspecto para o exame dos TCs é o número de itens e de formas em relação aos TPs. Nos TCs de cardiologia analisados, ocorre maior número de itens (54.613) do que nos TPs (53.220). A maior incidência de itens nos TCs confirma a hipótese da explicitação pela qual os TCs costumam ter mais palavras do que os TPs. Já o número de formas 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/item dos TPs de cardiologia é de 13,40 e a dos TCs de 10,23. A razão forma/item padronizada de 41,28 nos TPs e de 37,21 nos TCs também aponta para uma menor variação de palavras na tradução, 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.

No subcorpus de TCs de Ortopedia, ao contrário do que ocorre com o subcorpus de cardiologia, houve um menor número de itens nos TCs (28.870) que nos TPs (31.613) assim como ocorre

também um menor número de formas nos TCs (3.566) que nos TPs (4.585). A razão forma/item dos TPs foi de 14,50 e dos TCs de 12,35. Já a razão forma/item padronizada dos TPs foi de 41,17 e a dos TCs de 39,22. Este corpus não confirma o princípio de explicitação, uma vez que houve uma redução do número de itens e de formas nos TCs.

Para ilustrar esse traço, destacamos, no exemplo (1) do corpus de Ortopedia em que podemos notar a diferença na extensão do segmento do TP e a do TC:

- (1) Conclusões: 1) A reconstrução do ligamento patelofemoral medial (LPFM) com a restauração da funcionalidade dessa estrutura anatômica mostrou-se um bom método de tratamento de luxação recidivante da patela. 2) A rápida recuperação pós-operatória, garantida pela baixa morbidade da técnica empregada, permitiu aos pacientes operados retorno antecipado às suas atividades profissionais, reduzindo os custos sociais inerentes aos procedimentos cirúrgicos usualmente empregados no tratamento dessa patologia. 3) O seguimento de mais de cinco anos mostrou que os resultados da reconstrução da patela sustentaram-se adequadamente em face da inexorável pressão do tempo.

*Conclusions: Medial patellofemoral ligament reconstruction was very instrumental to improve unstable extension mechanism, and to prevent lack of confidence, gradual joint deterioration, and disabling pain during dislocation episodes, in a minimal 5 years follow-up.*

Já o menor número de formas dos TCs em relação aos TPs ratifica o princípio de simplificação, uma vez que há mais repetição nos TCs, sugerindo que o tradutor procurou tornar a leitura mais fácil no texto traduzido.

A pontuação usada pelos tradutores também mostra a presença de traços de simplificação, confirmando a hipótese sugerida por Baker (1996) de que, normalmente, usa-se uma pontuação mais "forte" no TC, ou seja, uma vírgula transforma-se em ponto e vírgula no TC, ou ainda um ponto e vírgula ou dois pontos do TP passam a ser ponto final no TC, a fim de quebrar sentenças mais longas dos TCs. O exemplo (2) a seguir, extraído do corpus de cardiologia, apresenta mudanças na pontuação:

- (2) Em todos os animais estudados foram aplicados dois protocolos experimentais: hipertensão arterial transitória e hipertensão arterial sustentada representados esquematicamente na figura 1. [grifo nosso]

*In all animal studied, two experimental protocols were applied. The transient and sustained arterial hypertension protocols are schematically represented in figure 1.* [grifo nosso]

Como podemos notar nos trechos grifados por nós, houve uma substituição de dois pontos no TP por um ponto final no TC, dando-se a retomada de informação na sentença seguinte.

Como outro exemplo de mudança de pontuação no corpus de Ortopedia, temos o exemplo (3):

- (3) Descrevemos a rotina para exame, expondo quadro objetivo, para respostas quanto ao exame neurológico antes da redução; exame vascular; classificação radiográfica da luxação em posterior, póstero-lateral, anterior, lateral ou rotatória; tipo de anestesia para redução (bloqueio plexular, geral e sem anestesia); a radiografia pós-redução deve identificar se a articulação está congruente, espaço articular no perfil em milímetros, e se há fratura do capitulo na superfície articular póstero-lateral. [grifo nosso]

*Examination Routine describes the objective clinical picture, assessing neurological status before reduction, vascular examination, radiographic classification of the dislocation as posterior, posterolateral, anterior, lateral, or rotary. Kind of anesthetic employed for reduction (plexus blockade, general anesthetic, no anesthesia), post-reduction X-ray assessing joint congruity, and joint space in millimeters in the lateral view, and the presence of capitellum fracture at the posterolateral joint surface were all assessed. [grifo nosso].*

Nesse segmento do TP, o autor usou ponto e vírgula; já o tradutor empregou o ponto final, tendo dividido o período longo do TP em duas orações no TC. É interessante ressaltar que os textos de cardiologia escritos originalmente em português são, na sua maior parte, mais elaborados, contendo orações complexas e extensas. Talvez seja esse o motivo que levou a tradutora dos textos de cardiologia a modificar a pontuação dos TCs, numa tentativa, consciente ou inconsciente, de facilitar a leitura na língua de chegada.

Notamos que no subcorpus de TCs de Ortopedia, os períodos, em geral, são curtos, não tendo havido, desse modo, a necessidade de o tradutor recorrer a quebras de orações.

No tocante às linhas de concordância geradas pelo aplicativo Concord, nas quais as palavras aparecem junto ao seu cotexto, pudemos averiguar como os termos levantados são empregados na linguagem médica. A palavra "coração", com 54 incidências, apresentou um número muito abaixo de ocorrências se comparada à palavra heart (com 211 incidências) nos TPs. Na maior parte dos TPs, encontramos a locução adjetiva "do coração", como em "capacidade contrátil do coração", "ápice do coração", "doença isquêmica do coração" etc, ou apenas "coração", como podemos observar nas linhas de concordância apresentadas no exemplo (4) a seguir:

(4) LINHAS DE CONCORDÂNCIA DA PALAVRA "coração" COMO NÓDULO

- 1 em preparações de *coração* isolado e no *coração* in situ,
- 2 a capacidade contrátil do *coração* por razões de ordem
- 3 contornava o ápice do *coração* e ascendia no sulco
- 4 o aparecimento da doença isquêmica do *coração* e
- 5 bem como os estudos radiológicos do *coração*, esôfago

A diferença significativa na frequência da palavra *heart*, que apresentou 211 incidências, deve-se à sua co-ocorrência, principalmente na função de adjunto adnominal em inglês, como em *heart failure*, *heart rate*, *heart attack*, *heart disease* e *heart valve*. Apresentamos, no exemplo (5), uma amostra de linhas de concordância da palavra *heart* no corpus de TCs de cardiologia.

(5) LINHAS DE CONCORDÂNCIA DA PALAVRA *heart* COMO NÓDULO

- 1 an incidence of 12.5%. In none of the *hearts* that we
- 2 and 26,000 deaths occurred due to *heart* failure in the
- 3 the control of cardiac rhythm or *heart* rate in atrial
- 4 countries, where one third of *heart* attacks and one
- 5 the absence of congenital or secondary *heart* disease.

No subcorpus de TCs de Ortopedia, destacamos o termo *patellar*, que ocorreu 101 vezes. Ao observarmos os respectivos TPs, verificamos que o mesmo termo em inglês tem como correspondente o adjetivo "patelar", como em "tendinite patelar" e "componente patelar" traduzidos respectivamente por *patellar tendinites* e *patellar component*. Também ocorre como locução adjetiva "da patela" em termos como "pólo inferior da patela" no TP e lower patellar pole no TC. Para ilustrarmos tal fato, apresentamos as linhas de concordância nos exemplos (6), (7), (8) e (9) a seguir:

(6) LINHAS DE CONCORDÂNCIA DA PALAVRA "patelar" COMO NÓDULO

- 1 dos a tratamento cirúrgico da tendinite *patelar* no Hospital Madre
- 2 quanto à colocação ou não do componente *patelar*, bem como

(7) LINHA DE CONCORDÂNCIA DA PALAVRA "patela" COMO NÓDULO

- 1 com corticosteróides no pólo inferior da *patela*, sem melhora

(8) LINHAS DE CONCORDÂNCIA DA PALAVRA *patellar* COMO NÓDULO

- 1 Also known as jumpersÆ knee, *patellar* tendinitis usually afflicts
- 2 sure the remaining bone stock under the *patellar* component, the
- 3 such as drilling, shaving, and lower *patellar* pole resection.

## 4. Conclusão

O levantamento de termos médicos mais frequentes neste estudo revela a organização das palavras dentro de 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 e Ortopedia. Quanto aos traços de tradução propostos por Baker (1996), as observações feitas até o presente momento, com base nos princípios da Linguística de Corpus e o auxílio de ferramentas eletrônicas, evidenciaram traços de simplificação, por meio de mudanças de pontuação e quebras de sentenças longas. A comparação de textos traduzidos por dois profissionais em duas subáreas médicas também aponta algumas características da tradução na área médica bem como revela o uso de padrões específicos de cada tradutor. Ao observarmos a razão forma/item nos corpora das duas subáreas, pudemos notar a presença de traços de simplificação nos TCs analisados, destacando-se os TCs de Cardiologia com um número significativamente maior de palavras em relação aos TPs.

Como não houve confirmação da hipótese de explicitação sugerida por Baker de que os textos traduzidos seriam mais longos do que os respectivos textos originais no tocante ao corpus de Ortopedia, pretendemos observar, na continuação desta pesquisa, se há influência de um determinado TC desviante no conjunto de textos desse corpus, ou se há a ocorrência de TCs com menor extensão devido a idiosincrasias do respectivo tradutor. Para tanto, realizaremos uma análise detalhada em cada TC e TP separadamente a fim de verificarmos o motivo de tal resultado.

Além disso, pretendemos aumentar os corpora de estudo por meio da compilação de corpora comparáveis, compostos por artigos originalmente escritos em português e em inglês, para que possamos observar as características da linguagem de especialidade utilizada pelos autores nativos de cada língua e compará-las às características presentes na linguagem empregada pelos tradutores profissionais.

## Referências

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.

BARROS, L. A. **Curso básico de terminologia**. São Paulo: USP, 2004.

\_\_\_\_\_. 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.