

O *corpus* combinado e a pesquisa nos Estudos da tradução baseados em *corpora*

The use of bidirectional parallel *corpus* within *Corpus* based translation studies

Silvana Maria de Jesus¹

RESUMO: Este artigo aborda as relações de tradução de SAY/DIZER em textos ficcionais no par linguístico inglês-português. Adotando uma perspectiva empírica de observação de dados em *corpus* combinado, este trabalho aborda os tipos de *corpora* e suas características, detalhando o *corpus* combinado e os procedimentos metodológicos ao se utilizar de dois programas computacionais – *WordSmith Tools* e SPSS. O *corpus* combinado apresentado é composto de três romances originais em inglês e suas traduções para o português e três romances originais em português e suas traduções para o inglês, e faz parte do CORDIALL (*Corpus* Discursivo para Análises Linguísticas e Literárias) desenvolvido pelos pesquisadores do LETRA (Laboratório Experimental de Tradução) da Faculdade de Letras da UFMG. Os programas *WordSmith Tools* e SPSS (*Statistical Package for the Social Sciences*) são utilizados para a extração de dados quantitativos para a construção de um banco de dados. Os resultados apresentam os equivalentes possíveis de SAY/DIZER, segundo as ocorrências do *corpus*, detectando-se padrões distintos nas relações de tradução, conforme a direção, seja do inglês para o português ou do português para o inglês.

PALAVRAS-CHAVE: Estudos da tradução baseados em *corpora*. *Corpus* combinado. Relações de tradução. Linguística de *Corpus*.

ABSTRACT: This work reports on a study developed at LETRA - Laboratory for Experimentation in Translation, Faculdade de Letras, UFMG. Building on corpus based translation studies, it examines equivalence relations between two verbs in the pair English-Portuguese, SAY/DIZER. Equivalence is looked at from an empirical perspective drawing on data gathered from translated and non-translated fiction in both languages and in both translation directions. Equivalence relations are studied applying Corpus Linguistics methodology for quantitative data analysis, using two software – *WordSmith Tools* and SPSS (*Statistical Package for the Social Sciences*) – and a bidirectional parallel corpus (comparable and parallel) compiled to that end. It investigates the originals and their translations to search and account for possible equivalents. Results point to different patterns of these verbs, relating to their functions and to equivalence relations.

KEYWORDS: Corpus based translation studies. Bidirectional parallel corpus. Equivalence relations. Corpus Linguistics.

¹ Professora Doutora do Curso de Tradução do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU).

1. Introdução

O uso de *corpora* de textos traduzidos em pesquisas de tradução tem como base os trabalhos de Baker (1993, 1995, 1996), que por sua vez, têm como base os trabalhos de Toury (1995) com sua proposta de Estudos descritivos da tradução. Baker introduziu uma terminologia para os tipos de *corpus*, que se tornou bastante difundida, embora tenha recebido críticas e reformulações. Os trabalhos de Baker também são importantes pela abordagem que propõem – a busca de universais de tradução – que se constituiu em importante marco inicial para a investigação dos textos traduzidos.

Este artigo não pretende fazer um histórico sobre a Linguística de *Corpus*, tampouco sobre os Estudos da Tradução baseados em *corpora*, assuntos sobre os quais o leitor encontrará ampla discussão em Berber-Sardinha (2004) e no periódico *Cadernos de Tradução IX* (TAGNIN, 2002), um número especial sobre Tradução e *Corpora*.

O objetivo deste trabalho é trazer uma discussão sobre a terminologia utilizada para os tipos de *corpora* e introduzir um tipo de *corpus* pouco utilizado, o *corpus* combinado; pretende-se ainda mostrar como os dados brutos (*raw data*) do *corpus* oferecem material de análise ao pesquisador, a partir dos quais análises mais refinadas podem ser desenvolvidas, bem como descrever os procedimentos metodológicos para a obtenção de dados, destacando-se os recursos informáticos utilizados, especialmente os software *WordSmith* e SPSS.

Inicia-se, portanto, este artigo, com a discussão sobre os tipos de *corpora* e a definição de *corpus* combinado. Em seguida, descrevem-se os procedimentos metodológicos para a obtenção de dados do *corpus* através de programas computacionais, observando-se como é possível formular perguntas de pesquisa a partir destes dados brutos e como estas perguntas iniciais podem nortear o refinamento da análise. Por fim, para ilustrar a metodologia, desenvolve-se uma pequena análise das relações de tradução entre SAY e DIZER, a partir dos dados de um *corpus* combinado.

2. Os diferentes tipos de *corpora*

Desde os trabalhos seminais de Baker (1993, 1995, 1996), que marcaram o início da interface entre os Estudos da Tradução e a Linguística de *Corpus*, os tipos de *corpora* utilizados nas pesquisas têm se diversificado de forma a contemplar diversos aspectos da descrição e comparação de línguas.

Inicialmente, foram definidos por Baker (1995:230) três tipos: i) *corpora* paralelos – textos originais em língua A e suas traduções em língua B; ii) *corpora* multilíngues – textos não-traduzidos compilados a partir de critérios comuns, como por exemplo, o gênero do discurso, nas línguas A e B; iii) *corpora* comparáveis – textos originalmente produzidos em uma língua A e textos traduzidos para essa mesma língua A, ou seja, um *corpus* monolíngue, formado de textos não-traduzidos e textos traduzidos.

O termo *comparável*, utilizado por Baker para se definir *corpora* de textos traduzidos e não-traduzidos, pode ser usado como um termo geral, no sentido comum de se comparar duas ou mais características, que podem ser textos traduzidos x não-traduzidos, língua A x língua B, ou ainda gêneros diferentes, como textos acadêmicos x textos midiáticos, o que pode gerar certa confusão terminológica. Nas pesquisas linguísticas, fora do âmbito da tradução, o termo é usado para se nomear um *corpus* bilíngue ou multilíngue de textos não-traduzidos em que aspectos de diferentes línguas estão sendo comparados.

Nos Estudos da tradução, tem-se optado pelos termos *corpus comparável bilíngue* (ou multilíngue) e *corpus comparável monolíngue* para se distinguir entre a comparação de textos não-traduzidos em línguas diferentes e textos traduzidos e não-traduzidos em uma mesma língua, respectivamente (Olohan, 2004:34). Hansen e Teich (2001), em uma análise contrastiva do inglês e do alemão, utilizam os termos *textos comparáveis monolíngues*, para o *corpus* de textos traduzidos e não-traduzidos em alemão, *textos paralelos*, para o *corpus* de textos originais em inglês e textos traduzidos em alemão, e *textos comparáveis multilíngues*, para o *corpus* de textos não-traduzidos em inglês e em alemão.

Observam-se, ainda, os tipos de *corpora* citados por Ana Frankenberg-Garcia (2006), considerados pela autora como relevantes para os estudos da tradução: *corpora* comparáveis bilíngues, *corpora* comparáveis monolíngues, *corpora* paralelos unidirecionais e *corpora* paralelos bidirecionais. A nomenclatura dos tipos de *corpora* por si só não define as características e finalidades dos mesmos, sendo então explicitados pela autora.

Corpora comparáveis bilíngues são formados de um *subcorpus* de textos na língua A e um *subcorpus* de textos na língua B, não-traduzidos, não-paralelos, semelhantes em gênero e função e podem ser usados para extração de terminologia e no dia-a-dia de tradutores. *Corpora comparáveis monolíngues* são formados de um *subcorpus* de textos não-traduzidos e um *subcorpus* de textos traduzidos, ambos em língua A, não-paralelos, semelhantes em gênero e função e podem ser usados para se observarem as diferenças entre a linguagem

traduzida e a não-traduzida e para estudos teóricos sobre a tradução. *Corpora paralelos unidirecionais* são formados de um *subcorpus* de textos originais em língua A e um *subcorpus* com os textos traduzidos para a língua B, são paralelos e semelhantes em gênero e função e podem ser usados para a pesquisa de equivalentes tradutórios, como suporte para a criação de dicionários bilíngues e para a tradução automática. *Corpora paralelos bidirecionais* são *corpora* bastante complexos e merecem uma explicação mais detalhada.

Os *corpora paralelos bidirecionais*, segundo Frankenberg-Garcia (2006), são formados de quatro *subcorpora* que, combinados entre si, podem formar outros *corpora*. Os quatro *subcorpora* se compõem em: 1) textos originais na língua A e 2) as suas traduções para a língua B, 3) textos originais na língua B e 4) as suas traduções para a língua A.

A autora explica que estes *corpora* formam dois *corpora* comparáveis monolíngues (1 e 4, traduzidos e não traduzidos na língua A, e 2 e 3, traduzidos e não traduzidos na língua B) e dois *corpora* comparáveis bilíngues (1 e 3, não-traduzidos na língua A e não-traduzidos na língua B, e 2 e 4, traduzidos na língua A e traduzidos na língua B).

Este tipo de *corpus* é denominado por Vela e Hansen-Schirra (2006) de *corpus* combinado paralelo-comparável (*combined parallel-comparable corpus*), ou simplesmente, *corpus combinado* e apesar de sua complexidade tem-se mostrado mais produtivo para as pesquisas em tradução. Hansen (2002:20) utiliza o termo *corpus* comparável e paralelo (*comparable and parallel corpus*) e descreve dois tipos de *corpus* combinado: o bilíngue e o multilíngue (mais de duas línguas).

Um dos estudos pioneiros na utilização de um *corpus* combinado, embora o autor não tenha utilizado esta denominação, é o de Johansson (1998). O autor utiliza um *corpus* formado de textos originais em inglês e suas traduções para o norueguês e textos originais em norueguês e suas traduções para o inglês. Johansson sugere a utilização dos seguintes termos para a denominação dos diferentes tipos de *corpora*: i) *corpora* comparáveis bilíngues/multilíngues ou *corpora* de textos originais comparáveis, ii) *corpora* de tradução ou *corpora* de textos originais e suas traduções, podendo ser bilíngue (original na língua A e tradução na língua B) ou multilíngue (original na língua A e traduções nas línguas B, C, D), e iii) *corpora* comparáveis monolíngues ou *corpora* de textos originais e traduzidos na mesma língua.

Observando-se as classificações apresentadas, nota-se que os *corpora*, que geralmente são formados por distintos *subcorpora*, são classificados segundo três aspectos: i) a(s) língua(s) envolvida(s), ii) o status do texto, e iii) a direcionalidade.

Em relação à língua, os *corpora* podem ser monolíngues (somente textos em língua A), bilíngues (textos em língua A e B), ou multilíngues (textos em língua, A, B, C... n).

O status classifica o texto na sua relação com outros textos, podendo ser considerado como original, tradução e texto não-traduzido. Destaca-se aqui que este status não se refere ao texto em si, mas à sua posição em relação a outro(s) texto(s) do *corpus*. Um texto será considerado como original se ele estiver em relação de tradução com outro(s) texto(s), formando assim um *corpus* paralelo bilíngue (ou multilíngue) com original e tradução (ou traduções). Este mesmo texto pode ser classificado como texto não-traduzido se ele estiver sendo analisado em relação a um texto na mesma língua que é um texto traduzido, formando assim, um *corpus* comparável monolíngue com texto(s) não-traduzido(s) em língua A e texto(s) traduzido(s) em língua A.

O termo *original* é visto por alguns teóricos, principalmente aqueles voltados para a interface entre os Estudos da tradução e os Estudos culturais, como inadequado, por pressupor aspectos relativos às questões de hierarquia. Entretanto, opta-se por utilizar os termos *original/tradução* para se destacar que são textos que estão em relação de tradução, sendo os termos *texto traduzido/texto não-traduzido* utilizados para se referir aos textos que não estão em relação de tradução entre si.

Aspectos discursivos também podem ser apontados no status, como tipo textual, registro ou gênero textual. O *corpus* pode ser compilado de forma a contrastar as características de um gênero textual em duas línguas ou de gêneros textuais diferentes em uma mesma língua.

O aspecto direcionalidade indica se a análise é feita apenas na direção da língua A para a língua B ou se ela considera também a direção inversa, ou seja, da língua B para a língua A. Pode, portanto, apresentar-se como unidirecional, quando, por exemplo, compara-se uma dada característica em textos originais na língua A e em suas traduções na língua B; ou bidirecional, quando esta análise é ampliada com a comparação desta mesma característica em textos originais na língua B e em textos traduzidos para a língua A, ou seja, quando se utiliza um *corpus* combinado.

Estes tipos básicos podem ser combinados conforme as necessidades de pesquisa e, como não há ainda uma terminologia consagrada para cada aspecto, cabe ao pesquisador explicitar as diversas características do *corpus* utilizado.

Outros aspectos relacionados à tipologia de *corpus*, como modo (falado ou escrito), tempo (sincrônico ou diacrônico) e outros, são abordados por Berber Sardinha (2004:20). A título de ilustração de diferentes tipos de *corpora*, podem-se observar alguns *corpora* em desenvolvimento nos Estudos da tradução, como por exemplo, o TEC, o CORDIALL e o CroCo.² Remete-se ainda o leitor ao trabalho de Dayrell (2005:48) para um levantamento de *corpora* existentes em português.

3. O *corpus* combinado

O *corpus* combinado apresentado nesse artigo foi usado por Jesus (2008), sendo parte do Projeto CORDIALL (*Corpus* discursivo para análises linguísticas e literárias, desenvolvido pelos pesquisadores do LETRA – Laboratório Experimental de Tradução da Faculdade de Letras da UFMG)³. É composto por doze romances, somando um total de 1.237.970 palavras. Caracteriza-se como um *corpus* combinado que, como foi dito, é composto de vários *corpora* básicos, cujas combinações permitem a composição de vários *corpora* de análise. Os quatro *corpora* básicos desta pesquisa são⁴:

- 1- *Corpus* PO - para português original ou não-traduzido
- 2- *Corpus* IT - para inglês traduzido
- 3- *Corpus* IO - para inglês original ou não-traduzido
- 4- *Corpus* PT - para português traduzido

Neste *corpus* os *subcorpora* 1-2 e 3-4 são paralelos, ou seja, no *subcorpus* 2 temos as traduções dos textos que compõem o *subcorpus* 1 e no *subcorpus* 4 as traduções dos textos que compõem o *subcorpus* 3. Em outra combinação, os *subcorpora* 1-4 e 2-3 são comparáveis monolíngues, ou seja, são formados de textos traduzidos e não-traduzidos em português e em

² TEC (<http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>), CORDIALL (<http://letra.letras.ufmg.br/letra/index.xml>) e CroCo (http://fr46.uni-saarland.de/croco/index_en.html).

³ Site do LETRA - <http://letra.letras.ufmg.br/letra/>

⁴ As siglas IO, IT, PO, PT foram utilizadas por Tagnin (2002a:203) na construção do corpus para o Projeto terminológico para tradutores. A opção de utilizá-las na presente pesquisa deve-se ao fato de que as siglas facilitam a visualização do tipo de corpus, destacando-se ainda que os textos não-traduzidos utilizados no corpus desta pesquisa são também os textos originais.

inglês, respectivamente. Há ainda um *corpus* comparável bilíngue formado por 1-3, com textos não-traduzidos em português e inglês.

Os textos foram utilizados anteriormente por pesquisadores do LETRA, oferecendo, portanto, a vantagem de já se encontrarem digitalizados. A compilação do *corpus* obedeceu aos seguintes critérios: romances disponíveis no CORDIAL, no par linguístico português-ínglês, publicados no século XX. Foram selecionados três romances para cada um dos *subcorpora*. O Quadro 1 apresenta os dados bibliográficos de cada romance e os *corpora* que eles compõem.⁵

Quadro 1 – Dados bibliográficos dos romances do *corpus* combinado

Corpora	Título	Sigla	Autor/Tradutor	1ª edição	Edição utilizada
IO	<i>Point counter point</i>	PCP	Aldous Huxley	1928	1994
	<i>Interview with the vampire</i>	IWV	Anne Rice	1976	1997
	<i>Beloved</i>	BEL	Toni Morrison	1987	1998
IT	<i>Macunaíma</i>	MAC	E A Goodland	1984	1984
	<i>Gabriela, clove and cinammon</i>	GAB	James L Taylor e William Grossman	1962	1962
	<i>The hour of the star</i>	THS	Giovanni Pontiero	1986	1992
PO	<i>Macunaíma: o herói sem caráter</i>	MHS	Mário de Andrade	1928	1980
	<i>Gabriela, cravo e canela</i>	GCC	Jorge Amado	1958	1958
	<i>A hora da estrela</i>	AHE	Clarice Lispector	1977	1998
PT	<i>Contraponto</i>	CPO	Érico Veríssimo	1934	1971
	<i>Entrevista com o vampiro</i>	ECV	Clarice Lispector	1977	1996
	<i>Amada</i>	AMA	Evelyn Kay Massaro	1987	1987

⁵ Para maiores detalhes sobre os corpora, remete-se o leitor aos trabalhos de Assis (2004) para os romances *Beloved* e *Amada*; Jesus (2004) para os romances *Point counter point* e *Contraponto*; Bueno (2005) para os romances *Macunaíma, o herói sem caráter* e *Macunaíma*; Cançado (2005) para os romances *Interview with the vampire* e *Entrevista com o vampiro*; Morinaka (2005) para os romances *Gabriela, cravo e canela* e *Gabriela, clove and cinammon*; e Rodrigues (2005) para os romances *A hora da estrela* e *The hour of the star*.

Estes *subcorpora* podem ser combinados, segundo os objetivos de análise, formando outros *corpora* para a investigação de aspectos específicos:

- i) *corpus paralelo IO-PT* – formado de textos originais em inglês (IO) e suas traduções para o português (PT)
- ii) *corpus paralelo PO-IT* – formado de textos originais em português (PO) e suas traduções para o inglês (IT)
- iii) *corpus comparável monolíngue IO-IT* – formado de textos não-traduzidos em inglês (IO) e textos traduzidos em inglês (IT)
- iv) *corpus comparável monolíngue PO-PT* – formado de textos não-traduzidos em português (PO) e textos traduzidos em português (PT)
- v) *corpus comparável bilíngue IO-PO* – formado de textos não-traduzidos em inglês (IO) e textos não-traduzidos em português (PO)

Em relação aos três aspectos que são observados em um *corpus*, quais sejam, língua(s), status e direcionalidade, este *corpus* caracteriza-se como bilíngue, composto por textos do par linguístico português-inglês; combinado, ou seja, contendo textos em ambas as relações – traduzidos/não-traduzidos e traduzidos/originais; e bidirecional, visto que a análise pode ser feita tanto na direção do português para o inglês quanto do inglês para o português. A próxima seção apresenta os procedimentos metodológicos para obtenção dos dados com o uso de programas computacionais.

4. Procedimentos metodológicos

O trabalho com *corpus* exige do pesquisador o uso de ferramentas computacionais para a obtenção de dados quantitativos. Esta seção introduz dois programas que podem ser usados para análises linguísticas, e os procedimentos metodológicos utilizados, ilustrando como as perguntas de pesquisa vão sendo desenvolvidas concomitantemente com a análise do *corpus*.

4.1 O programa *WordSmith Tools* e os dados brutos

Após a compilação do *corpus* da pesquisa, o tratamento do *corpus* é feito com a combinação de procedimentos manuais por parte do pesquisador e automáticos com a utilização de programas computacionais.

O software *WordSmith Tools*⁶ tem sido bastante utilizado em pesquisas com *corpus*. Para maiores detalhes sobre o programa remeto o leitor a Jesus (2004), Berber Sardinha (2004), Alves e Morinaka (2004) e ao site do programa⁷.

Com a ferramenta WordList é possível obter os seguintes dados do *corpus*: i) o número total de vocábulos (*types*) e o número total de ocorrências (*tokens*) do *corpus* e também de cada romance e/ou *subcorpus*; e ii) o número de ocorrências das formas de SAY/DIZER no *corpus*. A partir das formas encontradas na lista de palavras fornecida pela ferramenta Wordlist, buscam-se todas as ocorrências de SAY/DIZER com a ferramenta Concord.

A ferramenta Concord permite a busca de todas as formas de SAY/DIZER colocando-se como nóculo, que é a palavra de busca, a forma *say*/said* para SAY e *disse*/diz*/diria*/diga*/digo/dito/ditas/dir-se-ia* para DIZER. A definição das formas a serem utilizadas como nóculo foi feita através da observação das listas de palavras obtidas com a ferramenta WordList. O asterisco (*) indica ao programa que se deve buscar todas as palavras que iniciem pelas letras antes do asterisco, assim, no caso de *say** o programa busca também as formas *saying* e *says*. No caso de DIZER, quando se usa *disse** buscam também as formas *disseram*, *dissemos* e assim por diante. Neste processo, o programa busca também palavras que não são formas de SAY/DIZER, como por exemplo, *dizimados* para a forma *diz**, sendo necessário checar as linhas de concordância e eliminar estes casos. Para se eliminar as linhas, utiliza-se o comando deletar e o botão ZAP.

A Tabela 1 apresenta os dados de cada *subcorpora* do *corpus* combinado.

⁶ Uma outra opção é o programa AntConc, que não possui as mesmas funcionalidades, mas é de uso gratuito http://www.laurenceanthony.net/antconc_index.html.

⁷ Mike Scott's Web <http://www.lexically.net/wordsmith/>.

Tabela 1 – Dados dos quatro *corpora* básicos do *corpus* combinado

	IO	IT	PO	PT
Types	29.392	22.594	26.336	40.434
Tokens	402.525	234.599	203.504	397.342
Total de ocorrências de SAY/DIZER	2.723	776	707	2.385
% de SAY/DIZER x tokens	0.67	0.33	0.34	0.60

O tamanho do *corpus* é um aspecto bastante discutido na sua compilação (Olohan, 2004:45), com uma tendência, nos Estudos da tradução baseados em *corpora*, para o uso de *corpora* de mesmo tamanho (em relação ao número de textos ou de *tokens*). Entretanto, o *corpus* combinado aqui apresentado é formado de romances inteiros, originais e traduções, o que dificulta a padronização do tamanho. Além disso, serão analisadas 150 ocorrências de cada *subcorpus*, não sendo, portanto, relevante a questão do tamanho do *corpus*.

Os dados brutos já permitem algumas considerações para análise:

i) *corpus* comparável bilíngue IO-PO

Há 2.723 ocorrências de SAY em IO e 707 ocorrências de DIZER em PO.

Logo, a ocorrência de SAY em textos não-traduzidos (IO) é significativamente *maior* do que a ocorrência de DIZER em textos não-traduzidos (PO); SAY representa 0.67% de *tokens*, ao passo que DIZER representa 0.34%;

ii) *corpus* paralelo IO-PT

Há 2723 ocorrências de SAY em IO e 2.385 ocorrências de DIZER em PT.

Logo, a ocorrência de SAY nos textos originais é *maior* do que a de DIZER nos textos traduzidos. SAY equivale a 0.67% de *tokens* dos textos originais, enquanto que nas traduções a ocorrência de DIZER é de 0.60%.

iii) *corpus* paralelo PO-IT

Há 707 ocorrências de DIZER em PO e 776 ocorrências de DIZER em IT.

Logo, a ocorrência de DIZER nos textos originais é *semelhante* à ocorrência de SAY nos textos traduzidos. DIZER equivale a 0.34% de *tokens* dos textos originais enquanto que nas traduções a ocorrência de SAY é de 0.33%. Nota-se que a diferença é mínima em termos percentuais, ao passo que, em termos de ocorrências, o texto traduzido tem mais ocorrências

de SAY do que os originais têm de DIZER. Nota-se que a ocorrência de SAY é maior do que a ocorrência de DIZER tanto no *corpus* comparável bilíngue (IO-PO) quanto nos *corpora* paralelos (PO-IT e IO-PT);

iv) *corpus* comparável monolíngue (IO-IT)

Há 2.723 ocorrências de SAY nos textos não-traduzidos (IO) e 776 ocorrências de SAY nos textos traduzidos (IT). SAY equivale a 0.67% de *tokens* dos textos não-traduzidos, enquanto que nas traduções a ocorrência é de 0.33%. Nota-se, portanto, que a ocorrência de SAY nas traduções é significativamente *menor* do que a ocorrência em textos não-traduzidos.

v) *corpus* comparável monolíngue (PO-PT)

Há 707 ocorrências de DIZER nos textos não-traduzidos (PO) e 2.385 ocorrências de DIZER nos textos traduzidos (PT). DIZER equivale a 0.34% de *tokens* dos textos não-traduzidos, enquanto que nas traduções a ocorrência é de 0.60%. Nota-se, portanto, que a ocorrência de DIZER nas traduções é significativamente *maior* do que a ocorrência em textos não-traduzidos.

Estas considerações iniciais, obtidas através da observação dos dados brutos do *corpus*, permitem a elaboração de perguntas de pesquisas que orientarão a análise qualitativa do *corpus*. Cada um dos *subcorpora* do *corpus* combinado permite a análise de distintos aspectos das relações entre os textos que, somadas, contribuem para uma compreensão mais abrangente das relações de tradução. Estas perguntas baseiam-se no modelo de análise proposto por Teich (2003):

- 1- Qual a relação entre as orações verbais realizadas por SAY em textos ficcionais não-traduzidos em inglês e orações verbais realizadas por DIZER em textos ficcionais não-traduzidos em português?
- 2- As orações verbais realizadas por SAY são frequentemente traduzidas por orações verbais realizadas por DIZER e vice-versa? Qual a probabilidade de equivalência entre SAY/DIZER? Quais os outros itens lexicais em relação tradutória com SAY/DIZER? É possível estabelecer padrões que condicionem as relações de tradução?

- 3- Qual a relação entre as orações verbais realizadas por SAY/DIZER em textos não-traduzidos e as orações verbais realizadas por SAY/DIZER em textos traduzidos? Os textos traduzidos apresentam os mesmos padrões dos textos não-traduzidos ou estes padrões são condicionados pela relação dos textos traduzidos com os textos originais?

Para a investigação de cada pergunta, focaliza-se uma combinação diferente do *corpus*: 1- relações entre textos não-traduzidos em português e inglês (IO-PO); 2- relações entre textos traduzidos e textos não-traduzidos em inglês (IO-IT); 3- relações entre textos traduzidos e textos não-traduzidos em português (PO-PT); 4- relações entre textos originais em inglês e suas traduções para o português (IO-PT); e 5- relações entre textos originais em português e suas traduções para o inglês (PO-IT).

4.2 O programa SPSS e o banco de dados

A partir dos dados brutos e das primeiras considerações que eles permitem, é possível uma análise mais refinada dos dados. No total, há cerca de 6500 ocorrências de SAY/DIZER no *corpus*, que podem ser localizadas utilizando-se conjuntamente as ferramentas *WordList* e *Concord* do software *WordSmith Tools*, como foi explicado. Optou-se por selecionar cerca de 10% das ocorrências e, pensando-se em uma amostragem com o mesmo número de ocorrências de cada romance, optou-se por selecionar as primeiras 50 ocorrências de cada romance, totalizando 599 ocorrências, visto que o número total de ocorrências em um dos romances é 49.

Dado o foco do estudo, os processos verbais, foram eliminadas 30 ocorrências, 04 de SAY e 26 de DIZER, de orações em que SAY/DIZER realizam processos simbólicos e processos relacionais, fugindo, portanto, do escopo da análise.

Cada uma das 599 ocorrências foi analisada na linha de concordância fornecida pela ferramenta *Concord* e selecionaram-se apenas as ocorrências em que SAY/DIZER realizam processos verbais. Após este recorte, restaram 569 linhas de ocorrências, sendo 296 ocorrências de SAY (149 em textos não-traduzidos (IO) e 147 em textos traduzidos (IT)) e 273 ocorrências de DIZER (132 em textos não-traduzidos (PO) e 141 em textos traduzidos (PT)).

Estes dados já apontam o fato de que o papel de SAY na realização de orações verbais parece ser mais frequente do que o de DIZER, visto que foram eliminados apenas 04 casos de SAY e 26 casos de DIZER.

O próximo passo foi analisar cada ocorrência em relação às categorias focalizadas. O início da análise foi feito através de anotação⁸ manual do *corpus*, que consiste em o pesquisador estabelecer códigos para as categorias analisadas e inserir estes códigos no *corpus* dentro de etiquetas (denominam-se etiquetas os parênteses angulares < e >, dentro dos quais se insere o código no *corpus*). Posteriormente, as etiquetas são contabilizadas automaticamente com o programa *WordSmith Tools* para a obtenção de dados, possibilitando ao pesquisador desenvolver sua análise qualitativa a partir dos dados quantitativos. Entretanto, este processo é bastante limitado para análise de grande quantidade de dados, principalmente quando se utilizam várias categorias. Optou-se, portanto, pelo uso de um programa estatístico, o SPSS (*Statistical package for the social sciences*).

O programa SPSS trabalha com casos ou ocorrências, variáveis e categorias, formando um banco de dados em uma planilha semelhante às utilizadas pelo programa Excel. O SPSS apresenta muitas facilidades: a construção do banco de dados é simples; é possível classificar separadamente cada aspecto observado e depois cruzar os dados; as ferramentas de geração de tabelas e gráficos são simples e com muitos recursos de produção e edição; estão disponíveis diferentes formas de visualização dos dados. Dentre suas limitações pode-se apontar: custo muito alto do programa; dificuldade para se trabalhar com uma variável que possua mais de cinco categorias; não apresenta nenhuma interface com o texto, visto que o programa não foi desenvolvido especificamente para linguistas; necessidade de conhecimentos de estatística para análises mais elaboradas.

O programa destaca-se pela facilidade de construção de um banco de dados que possibilita o cruzamento de variáveis, a extração de dados de frequência e a geração de tabelas e gráficos, sem aprofundamento em testes estatísticos. Olohan (2004:86) aponta que vários linguistas consideram que os dados de frequência podem ser suficientes para a análise e que os testes de significância não são sempre necessários. A autora cita uma observação de Halliday sobre esse aspecto, em que o autor diz que “dados brutos de frequências geralmente são suficientes para se aceitar a afirmativa do pesquisador de que alguma característica é

⁸ Confira Feitosa (2005) para um estudo mais detalhado sobre anotação de corpora.

relevante no texto e que se deve verificá-la”⁹. Desta forma, o programa foi utilizado para o processamento automático de frequência das variáveis e categorias analisadas com a geração de tabelas e gráficos¹⁰.

Inicialmente, o pesquisador estabelece as variáveis e categorias de análise que irão fazer parte do banco de dados do SPSS. Recomenda-se a construção de um único banco de dados, portanto, as variáveis e categorias devem englobar todas as ocorrências do *corpus*. Algumas variáveis podem existir apenas para facilitar a organização do banco de dados, como, por exemplo, a variável OCORRÊNCIAS, que não possui categorias e em que cada número corresponde a uma das 50 linhas de concordância selecionada para investigação em cada romance. Outras variáveis servem para facilitar a seleção de parte dos dados dentro do banco de dados, por exemplo, a variável STATUS cujas categorias são inglês original (IO), português original (PO), inglês traduzido (IT), e português traduzido (PT). O programa permite a seleção de parte dos dados, segundo as diferentes categorias, para a geração de dados, o que permite ao pesquisador, no exemplo da variável STATUS, selecionar os dados em relação às ocorrências de português em geral, ou somente de português traduzido ou somente de português original, segundo os objetivos da pesquisa.

Após vários testes, foram estabelecidas nove variáveis para a análise com suas respectivas categorias, descritas a seguir.

1) Variável: OCORRÊNCIAS. Sem categorias

Foram selecionadas 300 ocorrências de SAY e 299 ocorrências de DIZER que se constituíram das 50 primeiras ocorrências de cada romance, sendo que um deles possui apenas 49 ocorrências de DIZER. O SPSS conta cada ocorrência como um caso, ocupando uma linha do banco de dados e numerando-as de 1 a 599. Depois de eliminados os 30 casos em que SAY/DIZER não realizam processo verbal, restaram 569 casos. Na variável OCORRENCIA cada uma foi numerada segundo as linhas de concordância selecionadas com o programa *WordSmith*, ou seja, de 1 a 50 pela ordem de ocorrência.

⁹ Minha tradução de: “a rough indication of frequencies is often just what is needed: enough to suggest why we should accept the analyst’s assertion that some feature is prominent in the text, and allow us to check his statements” (Halliday apud Olohan, 2004:86).

¹⁰ Para exemplos de pesquisas linguísticas com o uso do SPSS, veja a tese de Rothe-Neves (2002) na área de tradução e a dissertação de Oliveira (2006) na área de fonologia da língua portuguesa.

2) Variável: LEXICO. Categorias: SAY e DIZER

Cada ocorrência foi classificada como SAY ou DIZER, conforme faça parte do *corpus* em português ou do *corpus* em inglês.

3) Variável: CORPUS. Categorias: os doze romances do *corpus* combinado

Foi assinalada a origem de cada ocorrência, segundo os doze romances que compõem o *corpus* combinado. Essa classificação permite que se façam análises separadas de cada romance, procedimento necessário, por exemplo, para a localização de linhas de concordância a serem utilizadas como exemplos. Atribuiu-se uma sigla para cada um, conforme descrito na Tabela 1.

PCP - Point counter point; IWV - Interview with the vampire; BEL – Beloved;
MAC – Macunaíma; GAB – Gabriela, clove and cinammon; THS - The hour of the star;
MHS – Macunaíma: o herói sem caráter; GCC – Gabriela, cravo e canela;
AHE - A hora da estrela; CPO – Contraponto; ECV - Entrevista com vampiro;
AMA - Amada

4- Variável: FORMA. Categorias: as formas de SAY/DIZER

Foi analisada a frequência das diferentes formas de SAY/DIZER. Enquanto SAY possui apenas quatro formas, *say*, *says*, *said*, *saying*, DIZER possui várias, sendo as mais frequentes, *disse*, *dizer*, *dizia*.

5- Variável: MODOPROJ. Categorias: relato, citação, verbiagem

Nesta variável, cada oração verbal foi analisada segundo o modo de projeção: relato ou citação. Os casos classificados como verbiagem indicam que não ocorreu a projeção.

6- Variável: MODOEXP. Categorias: congruente e metafórico

Cada oração verbal foi classificada segundo o modo de expressão: congruente ou metafórico. As orações de verbiagem não foram analisadas em relação aos modos de expressão.

7- Variável: TRADUÇÃO. Categorias: traduções de SAY/DIZER no *corpus* paralelo

Para a classificação desta variável foi necessário realizar antes o alinhamento do *corpus* paralelo. Alinhar significa colocar um trecho do texto original seguido do trecho do texto traduzido correspondente (Lawson, 2001). Este trabalho foi feito com o auxílio de bolsistas do LETRA e utilizaram-se dois recursos diferentes: a ferramenta *Viewer and Aligner* do programa *WordSmith* e o processador de texto *Word*. O resultado não é muito distinto. O *Viewer* faz o alinhamento automático, por sentenças ou parágrafos, mas apresenta problemas que precisam ser corrigidos manualmente pelo pesquisador; o texto alinhado pode ser exportado para ser manuseado no *Word*. No *Word*, o alinhamento é feito manualmente dividindo-se as sentenças ou parágrafos em duas colunas de uma tabela, uma para o texto original e outra para o texto traduzido.

Então, cada equivalente encontrado passa a ser uma categoria, sendo que esta variável tem tantas categorias quantas forem os possíveis equivalentes, agrupados sob a forma do infinitivo, que funciona como lema. Por exemplo, os itens lexicais mais frequentes correspondentes a SAY no *corpus* são DIZER, FALAR e RESPONDER. E os mais frequentes correspondentes a DIZER são SAY, TELL e EXPLAIN. Os dados observados nos *corpora* alinhados foram lançados diretamente no SPSS.

8- Variável: RELAÇÕES. Categorias: prototípico, típico, atípico, omissão e não-verbal

Os itens lexicais que aparecem no *corpus* em relação de tradução com SAY/DIZER, foram agrupam nestas categorias, segundo o tipo de verbo dentro dos verbos que podem realizar processo verbal.

9- Variável: STATUS. Categorias: inglês/português e original/traduzido

Cada ocorrência foi classificada em relação aos quatro *corpora* básicos: inglês original (IO), português original (PO), inglês traduzido (IT) e português traduzido (PT).

Todas as 569 ocorrências foram classificadas em relação a estas nove variáveis, formando assim o banco de dados da pesquisa. O SPSS permite que se realize uma análise parcial, excluindo-se uma ou mais variáveis e/ou categorias, facilitando a exploração do banco de dados para diferentes tipos de análise. O programa gera tabelas e gráficos de dados, com extensas possibilidades de edição e apresentação.

A Figura 1 mostra uma janela do SPSS com as nove variáveis de análise e uma amostra parcial das ocorrências que compõem o banco de dados.

	occorenc	lexico	corpus	forma	modoproj	modoexp	relacoes	tradução	status	var							
1	1	SAY	PCP	said	citação	congrue	típicos	responder	IO								
2	2	SAY	PCP	say	citação	congrue	prototípi	dizer	IO								
3	3	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
4	4	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
5	5	SAY	PCP	said	verbiage	NA	prototípi	dizer	IO								
6	6	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
7	7	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
8	8	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
9	9	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
10	10	SAY	PCP	said	citação	congrue	típicos	perguntar	IO								
11	11	SAY	PCP	said	verbiage	NA	típicos	falar	IO								
12	12	SAY	PCP	say	citação	congrue	prototípi	dizer	IO								
13	13	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
14	14	SAY	PCP	say	verbiage	NA	NA	despedida	IO								
15	15	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
16	16	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
17	17	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
18	18	SAY	PCP	said	citação	congrue	NA	omissão	IO								
19	19	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
20	20	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
21	21	SAY	PCP	said	citação	congrue	NA	omissão	IO								
22	22	SAY	PCP	said	citação	congrue	NA	omissão	IO								
23	23	SAY	PCP	say	verbiage	NA	prototípi	dizer	IO								
24	24	SAY	PCP	said	citação	congrue	NA	omissão	IO								
25	25	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
26	26	SAY	PCP	say	citação	congrue	prototípi	dizer	IO								
27	27	SAY	PCP	say	verbiage	NA	prototípi	dizer	IO								
28	28	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
29	29	SAY	PCP	said	relato	metafóri	prototípi	dizer	IO								
30	30	SAY	PCP	saying	citação	congrue	típicos	repetir	IO								
31	31	SAY	PCP	saying	verbiage	NA	prototípi	dizer	IO								
32	32	SAY	PCP	said	citação	congrue	típicos	pedir	IO								
33	33	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
34	34	SAY	PCP	said	citação	congrue	típicos	responder	IO								
35	35	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
36	36	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
37	37	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
38	38	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
39	39	SAY	PCP	said	citação	congrue	típicos	perguntar	IO								
40	40	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								
41	41	SAY	PCP	said	citação	congrue	típicos	convidar	IO								
42	42	SAY	PCP	said	citação	congrue	típicos	perguntar	IO								
43	43	SAY	PCP	said	verbiage	NA	prototípi	dizer	IO								
44	44	SAY	PCP	said	citação	congrue	prototípi	dizer	IO								

Figura 1 – Janela do programa SPSS com visão parcial do banco de dados

A Tabela 2 é um exemplo de tabela produzida pelo SPSS e mostra as ocorrências de SAY/DIZER nos romances do *corpus*, abrangendo os quatro *corpora* básicos do *corpus* combinado. O *corpus* de textos originais em inglês (IO) com 149 ocorrências, o *corpus* de textos originais em português (PO) com 132 ocorrências, o *corpus* de textos traduzidos em inglês (IT) com 147 ocorrências e o *corpus* de textos traduzidos em português (PT) com 141 ocorrências, totalizando 569 casos ou ocorrências de SAY/DIZER. Nota-se que as tabelas produzidas pelo SPSS não seguem os padrões da ABNT.

Tabela 2 – Ocorrências de SAY/DIZER no *corpus* combinado (exemplo de tabela produzida pelo SPSS)

Romances do corpus * Texto original ou tradução Crosstabulation

Count		Texto original ou tradução				Total
		ING nao-trad	POR nao-trad	ING trad	POR trad	
Romances do corpus	PCP	50				50
	IWV	49				49
	BEL	50				50
	MAC			47		47
	GAB			50		50
	THS			50		50
	MHS		40			40
	GCC		47			47
	AHE		45			45
	CPO				49	49
	ECV				45	45
	AMA				47	47
	Total		149	132	147	141

A Tabela 2 está no formato produzido pelo SPSS, mas o programa tem amplos recursos para edição de tabelas. Pode-se, por exemplo, mudar o título, excluir colunas, alterar a fonte, inserir notas, alterar o tamanho e outras edições e formatações, conforme exemplificado na Tabela 3.

Tabela 3 – Exemplo de tabela editada com os recursos do SPSS

Ocorrências de SAY/DIZER nos romances do corpus combinado ^a

Count		Texto original ou tradução			
		ING nao-trad	POR nao-trad	ING trad	POR trad
Romances do corpus	PCP	50			
	IWV	49			
	BEL	50			
	MAC			47	
	GAB			50	
	THS			50	
	MHS		40		
	GCC		47		
	AHE		45		
	CPO				49
	ECV				45
	AMA				47
	Total		149	132	147

a. Corpus paralelo e comparável bidirecional no par linguístico inglês-português

Além do manual do programa (em inglês), é possível encontrar manuais em português com o passo-a-passo para a utilização dos recursos do programa.

Todas estas variáveis foram detalhadamente analisadas em Jesus (2008). A próxima seção apresenta, parcialmente, a análise de uma das variáveis, as relações de tradução, para ilustrar a metodologia utilizada.

5. As relações de tradução entre SAY-DIZER

O conceito leigo de equivalência geralmente está associado à ideia de que para um determinado item em dada língua A há um item equivalente na língua B. Entretanto, os estudos da tradução desde muito apontam que, para um determinado item em uma língua A, existem *vários* itens equivalentes na língua B ou, nos termos de Halliday (1992:16), “um item X na língua do original tem um grupo de itens equivalentes em potencial - A, B, C, D, E, F - na língua da tradução”¹¹. Portanto, pode-se dizer que existem diferentes *relações de tradução* entre um item do texto original e seus vários possíveis equivalentes no texto traduzido.

A análise das relações de tradução de SAY/DIZER tem o objetivo de responder as seguintes perguntas: i) Dado que, em dicionários bilíngues, DIZER é o equivalente prototípico de SAY, com qual frequência SAY é traduzido por DIZER? ii) Dado que há ocorrências de SAY que *não* são traduzidas por DIZER, pois não há apenas um equivalente, quais seriam os possíveis equivalentes de SAY além de DIZER? E qual seria a frequência de ocorrência destes equivalentes? As mesmas perguntas são colocadas em relação a DIZER, pensando-se na direção do português para o inglês.

Considerando-se SAY/DIZER como verbos que realizam Processo verbal, é natural que SAY/DIZER estejam em relações de tradução com outros verbos que também realizam processo verbal. A Gramática sistêmico-funcional (GSF, doravante) (HALLIDAY e MATTHIESSEN 2004) apresenta dois grupos de verbos que realizam processo verbal, segundo o modo de citação e o modo de relato.

Quadro 2 - Verbos que realizam Processo em orações verbais no modo de citação

	Proposição	Proposta
(1) Membro geral	SAY	SAY
(2) Verbos específicos	aa) Declarações: TELL (receptor), REMARK, POINT OUT, ANNOUNCE.	(+a) Ofertas: SUGGEST, OFFER; VOW, REPORT, PROMISE, AGREE.
a) oferta		
b) demanda	b) Perguntas: ASK, DEMAND,	b) Comandos: CALL, ORDER,

¹¹ Minha tradução de: You are aware that an item X in the source language has a potential equivalence of items A, B, C, D, E, F in the target language (Halliday, 1992:16).

	Proposição	Proposta
	INQUIRE, QUERY.	REQUEST, TELL, PROPOSE, DECIDE; URGE, PLEAD, WARN.
(3) Verbos com características adicionais	REPLY, EXPLAIN, PROTEST, CONTINUE, ADD, INTERRUPT, WARN.	[consultar (2) acima]
a) circunstanciais		
b) especificadores de modo ou conotação	INSIST, COMPLAIN, CRY, SHOUT, BOAST, MURMUR, STAMMER.	[em grande parte como as proposições] BLARE, THUNDER, MOAN, YELL, FUSS.

Fonte: Halliday e Matthiessen (2004:448). Tradução de Alves (2006:29)

A GSF aponta SAY como o membro geral da classe de verbos que realizam processos verbais em citações (Halliday e Matthiessen, 2004:448), conforme apresentado no Quadro 2. Destaca-se seu papel como o membro não-marcado deste grupo de verbos (p. 252). Thompson (1994:34-35) considera SAY como o verbo de elocução básico, utilizado quando não se quer indicar nenhuma finalidade ou modo do dizer, portanto, um verbo neutro. Considera ainda, como neutros, os verbos: *tell, ask, write, speak, talk*.

Halliday e Matthiessen (2004:448) consideram ainda que outros verbos “que não são verbos de dizer”, podem, especialmente na narrativa ficcional, realizar Processos verbais. Os autores destacam os verbos que tipicamente realizam Processos comportamentais (como *sob, snort, breathe*) e que quando realizam processos verbais expressam atitudes, emoções e gestos que acompanham o ato de fala.

A classificação dos diferentes tipos de verbos que realizam processos verbais é relevante para a análise das relações de tradução de SAY/DIZER, visto que aqueles são os mais prováveis candidatos ao quadro de equivalentes possíveis destes verbos.

Para esta análise, utilizaram-se os *corpora* paralelos do *corpus* combinado, compostos de textos originais em inglês (IO) e suas traduções para o português (PT) e de textos originais em português (PO) e suas traduções para o inglês (IT). A análise é bidirecional, ou seja, observa as relações de tradução tanto na direção do português para o inglês, quanto na direção do inglês para o português.

Foram selecionadas as primeiras 50 ocorrências de SAY/DIZER de cada romance. Estas ocorrências foram analisadas, excluindo-se as orações em que SAY/DIZER não realizam processo verbal, como mencionado. Portanto, analisam-se 149 orações verbais realizadas por SAY e 132 orações verbais realizada por DIZER.

Inicialmente, foram observadas as relações de tradução de SAY/DIZER, considerando-se o estrato léxico-gramatical, ou seja, as relações de equivalência de SAY (em IO) com DIZER e outros itens lexicais do português (em PT) e, por sua vez, as relações de equivalência de DIZER (em PO) com SAY e outros itens lexicais do inglês (em IT).

Ao invés de se considerar isoladamente cada item em relação de tradução com SAY/DIZER, decidiu-se por agrupá-los em categorias semânticas, na perspectiva experiencial de realização de Processos verbais, segundo a classificação da GSF para os diferentes tipos de verbos: i) membro geral, ii) verbos específicos e iii) verbos que, tipicamente, não são verbos de dizer, mas podem realizar esse significado.

Observa-se que, embora não existam descrições do português baseadas na LSF, Neves (2000:48) apresenta uma classificação dos verbos *discendi*, que se aproxima da proposta da GSF. A autora classifica os verbos de elocução como neutros (*dizer e falar*), verbos que indicam algum aspecto do dizer, como modo ou cronologia (*repetir, gritar, responder*) e verbos que podem funcionar como introdutores do discurso (*consolar, sorrir*). Menciona-se ainda a classificação de Garcia (1986:131), que considera como verbos de elocução os verbos *discendi* (*dizer, perguntar, responder, exclamar*) e os verbos *sentiendi* (*gemer, suspirar, lamentar-se, queixar-se*), que introduzem o discurso e realizam emoções, atitudes, comportamentos. Utilizou-se, ainda, para a análise dos verbos em português, o *Dicionário Houaiss da língua portuguesa*¹², observando-se as significações possíveis para os verbos e suas relações com o universo do dizer. Portanto, utiliza-se uma classificação das relações de tradução de SAY/DIZER em três tipos: i) prototípicos, ii) típicos e iii) atípicos.

Considera-se como prototípico o verbo que é o membro geral da classe, ou seja, SAY/DIZER. E como verbos típicos todos os outros verbos que realizam orações verbais nos modos de citação e relato. Como verbos atípicos consideram-se os verbos que tipicamente realizam *outro* tipo de Processo (mentais, materiais e outros), mas que nos textos ficcionais geralmente realizam Processos verbais ou que eventualmente podem realizar este tipo de Processo.

Assim, consideram-se três tipos de relações de tradução de SAY e DIZER: i) SAY é traduzido por DIZER, e vice-versa, em que a relação se dá com o verbo prototípico; ii) SAY/DIZER são traduzidos por outros verbos típicos de processos verbais; iii) SAY/DIZER

¹² Dicionário online, disponível para assinantes em <http://educacao.uol.com.br/dicionarios/>.

são traduzidos por verbos que não são verbos de dizer, mas que podem realizar processo verbal, em que a relação se dá com o verbo atípico.

Ocorrem ainda, no *corpus*, casos em que SAY/DIZER são traduzidos por um substantivo, ou seja, a relação de tradução não se dá entre verbos. Ou ainda, casos em que a oração verbal realizada por SAY/DIZER é omitida.

Portanto, as relações de tradução de SAY/DIZER encontradas no *corpus* analisado são classificadas em cinco tipos: prototípico, típico, atípico, não-verbal e omissão. Esta classificação é feita a partir dos *corpora* paralelos: i) SAY/DIZER aparecem como equivalentes, ou seja, SAY é traduzido por DIZER e DIZER corresponde a SAY, casos classificados como prototípicos; ii) SAY/DIZER são traduzidos por outros verbos que tipicamente realizam Processo verbal como *falar, perguntar, responder, pedir, tell, speak, remark, point out*, casos classificados como típicos; iii) SAY/DIZER são traduzidos por verbos que tipicamente realizam outros tipos de processos, como *lembrar, fazer, mean, think*, que são classificados como atípicos; iv) casos em que SAY/DIZER são traduzidos por um substantivo, e v) casos de omissão.

Inicialmente, foram observados os itens lexicais em relação de tradução com SAY no *corpus*, apresentados na Tabela 4.

Tabela 4 – Itens lexicais em relação de tradução com SAY

	Frequency	Valid Percent
Valid dizer ^a	98	65,8
omissão	21	14,1
falar	12	8,1
responder	5	3,4
perguntar	3	2,0
não-verbal	2	1,3
comentar	2	1,3
pedir	1	,7
convidar	1	,7
repetir	1	,7
concordar	1	,7
fazer	1	,7
retrucar	1	,7
Total	149	100,0

a. Corpus paralelo inglês-português (IO-PT)

O *corpus* confirma a expectativa de que DIZER seja o equivalente mais frequente de SAY; a relação com o verbo prototípico ocorre em 65.8% dos casos. Nota-se que a segunda

relação mais frequente é a de omissão, que ocorre em 14.1%. O terceiro item mais frequente foi *falar*, um dos verbos típicos, que ocorre em 8.1% dos casos. Há dois casos (1.3%) de relação não-verbal. Agrupando-se estes itens lexicais, segundo as categorias de análise, tem-se outra perspectiva das relações de tradução, apresentadas na Tabela 5.

Tabela 5 – Relações de tradução de SAY segundo as categorias

	Frequency	Valid Percent
Valid prototípico ^a	97	65,1
típicos	28	18,8
atípicos	1	,7
nao-verbal	2	1,3
omissão	21	14,1
Total	149	100,0

a. Corpus paralelo inglês-português (IO-PT)

O equivalente prototípico ocorre em 65.1% dos casos, enquanto que verbos típicos de processos verbais ocorrem em 18.8%. Como foi dito, a omissão ocorre em 14.1% dos casos, há duas ocorrências de relação não-verbal (1.3%) e uma ocorrência de verbo atípico (0.7%). Exemplos do *corpus* são apresentados no Quadro 3.

Quadro 3 – Exemplos do *corpus* das relações de tradução de SAY em IO-PT

Prototípico

04 "You forgetting how little it is," **said** her mother. (BEL)

-- Você está se esquecendo de que ela é muito pequenina, **disse** a mãe. (AMA)

Típico

45 "But I mean we want to get married."

"You just said so. And I **said** all right." (BEL)

--O que estou querendo dizer é que vamos nos casar.

-- Já me disse isso. E eu **falei** que está tudo bem. (AMA)

Atípico

46 "Were you?" **said** Lady Edward smiling and looking from one to the other. (PCP)

- Estavam? - **fez** Lady Edward, sorridente, olhando ora para uma ora para outra. (CPO)

Omissão

17 "We and Denver," she **said**.

"That all right by you? (BEL)

-- Eu e Denver.

-- E está tudo bem com você? (AMA)

Não-verbal

14 When the time came to **say** good-bye, he had shaken the skeleton hand. (PCP)

Chegara o momento da despedida: ele apertou na sua a mão esquelética do doente, a mão que jazia inerte sobre a coberta. (CPO)

08 She had not thought to ask him and it bothered her still that it might have been possible-that for twenty minutes, a half hour, **say**, she could have had the whole thing, every word she heard the preacher say at the funeral (and all there was to say, surely) engraved on her baby's headstone: Dearly Beloved. (BEL)

Não pensara em perguntar-lhe e ainda se perturbava com a probabilidade de ter sido possível - que por vinte minutos, **talvez** meia hora, teria obtido a coisa inteira, cada palavra que ouvira o pastor dizer no funeral (e certamente tudo o que havia para ser dito) gravada na lápide de sua filha: Caríssima Amada. (AMA)

Exemplos do *corpus* paralelo: inglês original (IO) e português traduzido (PT)

Em seguida, observaram-se os itens lexicais em relação de equivalência com DIZER, apresentados na Tabela 6.

Tabela 6 – Itens lexicais em relação de tradução com DIZER

	Frequency	Valid Percent
Valid say ^a	61	46,2
tell	14	10,6
omissão	8	6,1
mention	4	3,0
explain	4	3,0
não-verbal	3	2,3
speak	3	2,3
think	3	2,3
mutter	2	1,5
ask	2	1,5
laugh	1	,8
insist	1	,8
believe	1	,8
conclude	1	,8
pronounce	1	,8
divulge	1	,8
inform	1	,8
snap	1	,8
retort	1	,8
observe	1	,8
call	1	,8
remark	1	,8
confide	1	,8
object	1	,8
comment	1	,8
express	1	,8
beg	1	,8
rejoin	1	,8
reply	1	,8
hear	1	,8
add	1	,8
go	1	,8
resume	1	,8
accuse	1	,8
find	1	,8
concede	1	,8
cover	1	,8
imagine	1	,8
Total	132	100,0

a. Corpus paralelo português-inglês (PO-IT)

DIZER se destaca como o equivalente prototípico de SAY, embora não chegue a 50% dos casos (46.2%). O segundo item mais frequente é *tell*, com 10.6%. A porcentagem de omissão de DIZER é de 6.1% e há três ocorrências de relação não-verbal (2.3%). A Tabela 7 mostra as relações de tradução de DIZER, quando se agrupam estes itens segundo as categorias de análise.

Tabela 7 – Relações de tradução de DIZER segundo as categorias

	Frequency	Valid Percent
Valid prototípico ^a	61	46,2
típicos	49	37,1
atípicos	11	8,3
nao-verbal	3	2,3
omissão	8	6,1
Total	132	100,0

a. Corpus paralelo português-inglês (PO-IT)

O equivalente prototípico ocorre em 46.2% dos casos, enquanto que outros verbos típicos de processos verbais ocorrem em 37.1% e verbos atípicos em 8.3%. A omissão ocorre em 6.1% e a relação não-verbal em 2.3%. Exemplos no Quadro 4.

Quadro 4 – Exemplos do *corpus* das relações de tradução de DIZER em PO-IT

Prototípico

17 -- Isso é revoltante - **dizia** o Doutor enquanto o grupo caminhava pela rua sem calçamento, contornando o morro. (GCC)

"This is revolting," **said** the Doctor as the group walked along the unpaved street skirting the hill.

Típico

37 O ticotiquinho ficava azaranzado porque estava padecendo fome e aquele nhenhêhêm-nhenhêhêm azucrinando ele atrás, **diz**-que "Telo decumê!... telo decumê!..." (MHS)

The cowbird swallowed everything and **resumed** its habitual refrain "Boo-hoo! Mama, I'm famished, I'm famished!" (MAC)

Atípico

32 E todos com muito medo foram correndo pra dentro. Então Chuvisco desapeou e **disse** pra Macunaíma: Está vendo? (MHS)

Shower came down and **laughed** at Macunaíma: "Did you see that?" (MAC)

Omissão

12 Em Itamaracá Macunaíma passou um pouco folgado e teve tempo de comer uma dúzia de manga-jasmim que nasceu do corpo de dona Sancha, **dizem**. (MHS)

They ran to the island of Itamaracá, where the hero ate some mangoes, a dozen jasmim mangoes that had sprung from the grave of the late Dona Sancha; (MAC)

Não-verbal

03 O seu rico andor bordado de ouro, levavam-no sobre os ombros orgulhosos os cidadãos mais notáveis, os maiores fazendeiros, vestidos com a bata vermelha da confraria, e **não é pouco dizer**, pois os coronéis do cacau não primavam pela religiosidade... (GCC)

The gold-embroidered litter bearing the image of the saint was carried on the shoulders of the town's most important citizens, the owners of the largest plantations, dressed in the red gowns of the lay brotherhood. **This was significant**, for the cacao colonels ordinarily avoided religious functions. (GAB)

21 Pelópidas de Assunção d'Ávila descendia de uns Ávilas, fidalgos portugueses estabelecidos nas bandas de Ilhéus ainda no tempo das capitâneas. Pelo menos assim o afirmava o Doutor, dizendo-se baseado em documentos de família. Opinião ponderável, de historiador. Descendente desses celebrados Ávilas, cujo solar elevava-se entre Ilhéus e Olivença, hoje negras ruínas ante o mar, cercadas de coqueiros, mas também de uns Assunções plebeus e comerciantes - **diga**-se em sua homenagem, ele cultuava a memória de uns e de outros com o mesmo fervor exaltado. (GCC)

Pelópidas de Assunção d'Ávila (the Doctor) maintained that he was descended from the Portuguese noblemen named Ávila who had settled near what was now Ilhéus during the period of royal land grants. The line of descent could be clearly traced, he found, in family documents. The solid opinion of a historian. He was descended also from certain plebeian, shopkeeping Assunções, and, **to his great credit**, he cherished the memory of these ancestors and of the Ávilas with the same exalted fervor. (GAB)

01 Si o incitavam a falar exclamava:

- Ai! que preguiça!...

e não **dizia** mais nada. (MHS)

If anyone tried to make him speak he would exclaim, "Aw! What a fucking life!" but **nothing more**.

Exemplos do *corpus* paralelo: português original (PO) e inglês traduzido (IT)

Vários aspectos se destacam nas relações de tradução de SAY e DIZER a partir destes dados do *corpus*. Poder-se-ia pensar que SAY é a tradução de DIZER na mesma proporção que DIZER é a tradução de SAY, mas os dados apontam diferenças nas relações de tradução, conforme a direção seja do inglês para o português ou do português para o inglês. Estas diferenças são discutidas em cada uma das relações de tradução observadas para SAY e DIZER.

Relação de tradução I - equivalente prototípico

A ocorrência do equivalente prototípico é maior na direção do inglês para o português (65.1%) do que do português para o inglês (46.2%).

Relação de tradução II – equivalente típico

SAY apresenta relações de tradução com 09 verbos típicos de Processos verbais no português além de DIZER: *falar, responder, perguntar, comentar, pedir, convidar, repetir, concordar e retrucar*. DIZER, por sua vez, aparece em relação de tradução com 26 verbos típicos de Processos verbais além de SAY: *tell, mention, explain, speak, mutter, ask, insist, pronounce, divulge, inform, snap, retort, observe, call, remark, confide, object, comment,*

express, beg, rejoin, reply, add, resume, accuse e concede. Estatisticamente, 18.8% das ocorrências de SAY estão em relação de tradução com verbos típicos e 37.1% das ocorrências de DIZER.

Relação de tradução III – equivalente atípico

SAY apresenta relação de tradução com 01 verbo atípico (*fazer*), ao passo que DIZER apresenta relações de tradução com 09 verbos atípicos: *think, laugh, conclude, believe, hear, go, find, cover e imagine*. Este tipo de relação representa 0.7% das ocorrências com SAY e 8.3% das ocorrências com DIZER.

Relação de tradução IV – equivalente não-verbal

Há 02 casos com SAY (1.3%) e 03 com DIZER (2.3%).

Relação de tradução V – omissão

A omissão ocorre 14.1% com SAY e 6.1% com DIZER.

6. Apontamentos finais

Este artigo analisou as relações de tradução de SAY/DIZER a partir do uso de um *corpus* combinado. Foram observados os equivalentes possíveis destes itens, a partir dos correspondentes encontrados no *corpus*. Apresentou-se, ainda, a metodologia da pesquisa e os primeiros passos da análise, visto que a metodologia não se desenvolve de forma linear, mas vai sendo construída a partir das primeiras etapas da pesquisa em análises piloto. Discutiu-se a terminologia utilizada para classificação dos diferentes tipos de *corpus*, observando-se que o *corpus* combinado oferece uma maior abrangência de possibilidades de pesquisa, permitindo ao pesquisador ampliar as perspectivas de análise. Foram descritas as características de um *corpus* combinado, bem como os procedimentos metodológicos para extração de dados. Foram introduzidos os programas computacionais *WordSmith Tools* e *SPSS* e as variáveis e categorias que podem ser usadas para pesquisa com estes programas. Mostrou-se como iniciar a pesquisa a partir dos dados brutos, que possibilitam ao pesquisador as primeiras observações a partir das quais a análise pode desenvolver-se. Desta forma, o trabalho ilustra a produtividade do uso de *corpus* combinado nos Estudos da tradução baseados em *corpora*.

Referências bibliográficas

ALVES, D. A. S. **Aspectos da representação do discurso em textos traduzidos**: os verbos de elocução neutros. 2006. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

ALVES, D. A. S.; MORINAKA, E. M. **Compilação de procedimentos metodológicos adotados por pesquisadores(as) em Estudos da Tradução e interfaces com as Linguísticas Sistêmico-Funcional e de Corpus**. Disponível em < www.geocities.com/xalaskero/UFMG/Metodologia >. Acesso em < 15 junho 2007 >.

ASSIS, R. C. **A transitividade na representação de Sethe no corpus paralelo “Beloved-Amada”**. 2004. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

BAKER, M. Corpus linguistics and translation studies: implications and applications. In: BAKER et al. (Ed.). **Text and technology**: in honour of John Sinclair. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1993. Cap., p. 233-250.

BAKER, M. Corpora in translation studies: an overview and some suggestions for future research. **Target**, Amsterdam, v. 7, n. 2, p. 223-243, 1995. **crossref** <http://dx.doi.org/10.1075/target.7.2.03bak>

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

BERBER SARDINHA, T. **Linguística de corpus**. Barueri, SP: Manole, 2004.

BUENO, L. T. **Transitividade, coesão e criatividade lexical em “Macunaíma”, de Andrade, e “Macunaíma”, de Goodland**. 2005. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

CANÇADO, T. **Transitividade e representação do discurso no corpus paralelo “Interview with the Vampire/Entrevista com o Vampiro”**. 2005. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

DAYRELL, C. **Investigating lexical patterning in translated Brazilian Portuguese: a corpus-based study**. 2005. 316 p. Tese (Doutorado em Estudos da Tradução). Centre for Translation and Intercultural Studies, The University of Manchester.

FEITOSA, M. P. **Uma proposta de anotação de corpora paralelos com base na linguística sistêmico-funcional**. 2005. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

FRANKENBERG-GARCIA, A. **Corpora e Tradução**. Disponível em < <http://www.linguateca.pt/escolaverao2006/> >. Acesso em < 15 junho 2007 >.

GARCIA, O. M. **Comunicação em prosa moderna**. 13. ed. Rio de Janeiro: Fundação Getúlio Vargas, 1986.

HALLIDAY, M.A.K. Language theory and translation practice. **Rivista internazionale di tecnica della traduzione**, Trieste, n. 1 (pilot issue), p. 15-25, 1992.

HALLIDAY, M.A.K.; MATTHIESSEN, C. M. I. M. **An introduction to functional grammar**. 3rd edition, rev. ampl. London: Arnold, 2004. 689 p.

HANSEN, S.; TEICH, E. Multi-layer analysis of translation *corpora*: methodological issues and practical implications. In: **EUROLAN 2001: Workshop on Multi-layer Corpus-based Analysis, 2001, Proceedings...** Disponível em <<http://www.coli.uni-saarland.de/publications/show.php?year=2001>>, Acesso em < janeiro de 2008 >.

HANSEN, S. **The nature of translated text: an interdisciplinary methodology for the investigation of the specific properties of translations**. 2002. 245 p. Tese (Doutorado em Estudos da Tradução). Department of Applied Linguistics, Saarland University. (Saarbrücken dissertations in computational linguistics and language technology, v. 13).

JESUS, S. M.. **Representação do discurso e tradução: padrões de textualização em corpora paralelo e comparável**. 2004. 128 p. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

JESUS, S. M. **Relações de tradução: SAY e DIZER em corpora de textos ficcionais**. 2008. Tese (Doutorado em Linguística Aplicada). Faculdade de Letras, Universidade Federal de Minas Gerais.

JOHANSSON, S. On the role of corpora in cross-linguistic research. In: JOHANSSON, S; OKSEFJELL, S. (Ed). **Corpora and cross-linguistic research: theory, method, and case studies**. Amsterdam: Rodopi, 1998. Cap. 1, p. 3-24.

JOHANSSON, S.; OKSEFJELL, S. (Ed). **Corpora and cross-linguistic research: theory, method, and case studies**. Amsterdam: Rodopi, 1998. 376 p. (Language and computers: studies in practical linguistics, n. 24)

LAWSON, A. Collecting, aligning and analyzing parallel corpora. In: GHADDESSY, M. et al. (Ed.). **Small corpus studies and ELT. Theory and practice**. Amsterdam: John Benjamins, 2001. Cap., p. 47-67.

MORINAKA, E. M. **“Gabriela, cravo e canela” and its (re)textualization in English: representation through lexical relations**. 2005. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

NEVES, M. H. de M. **A gramática funcional**. São Paulo: Martins Fontes, 1997. (Coleção Texto e Linguagem).

NEVES, M. H. de M. **Gramática de usos do português**. São Paulo: UNESP, 2000: 31-53.

OLIVEIRA, A. J. **Variação em itens lexicais terminados em /l/+vogal na região de Itaúna/MG**. 2006. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

OLOHAN, M. **Introducing corpora in Translation Studies**. London: Routledge, 2004. 220 p.

RODRIGUES, R. R. **A organização temática em “A hora da estrela” e “The hour of the star”**: um estudo de caso. 2005. Dissertação (Mestrado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

ROTHE-NEVES, R. **Características cognitivas e desempenho em tradução: investigação em tempo real**. 2002. Tese (Doutorado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais.

SCOTT, M. **WordSmith Tools**. 1999. Disponível em < <http://www.lexically.net/wordsmith/> > Acesso em: 12 março 2003.

TAGNIN, S. E. O. (Org). **Cadernos de Tradução IX** (Número especial sobre Tradução e Corpora). Florianópolis: UFSC/NUT, 2002, n. 1, 278 p.

TAGNIN, S. E. O. Os *corpora*: instrumentos de auto-ajuda para o tradutor. **Cadernos de Tradução IX**, Tradução e Corpora. Florianópolis, 2002a, n. 1, p. 191-219.

TEICH, E. Towards a model for the description of cross-linguistic divergence and commonality in translation. In: STEINER, E.; YALLOP, C. (Ed.). **Exploring translation and multilingual text production: beyond content**. Berlin, New York: Mouton de Gruyter, 2001. Cap, p. 191-227.

TEICH, E. **Cross-linguistic variation in system and text: a methodology for the investigation of translations and comparable texts**. Berlin: Mouton de Gruyter, 2003. 276p. (Text, translation, computational processing, 5). **crossref** <http://dx.doi.org/10.1515/9783110896541>

THOMPSON, G. **Collins Cobuild English Guides 5: Reporting**. London: HarperCollins Publishers, 1994.

TOURY, G. **Descriptive translation studies and beyond**. Amsterdam: John Benjamins, 1995. 311 p. (Benjamins Translation Library).

VELA, M; HANSEN-SCHIRRA, S. The use of multi-level annotation and alignment for the translator. In: ASLIB Translating and the computer 28 conference, 2006, Londres. **Proceedings....** Disponível em < http://fr46.uni-saarland.de/croco/publication_en.html > Acesso em < 14 junho 2007 >

Corpus analisado

AMADO, J. **Gabriela, cravo e canela**. Rio de Janeiro: Record, 1958.

AMADO, J. **Gabriela, clove and cinnamon**. Trad. Taylor, J. e Grossman . New York: Avon Books, 1962. (Tradução de *Gabriela, cravo e canela*).

ANDRADE, M. **Macunaíma**: o herói sem nenhum caráter. São Paulo e Belo Horizonte: Martins e Itatiaia, 1980.

ANDRADE, M. **Macunaíma**. Tradução de E. A. Goodland. London: Quartet Books, 1984. (Tradução de: *Macunaíma: o herói sem caráter*).

HUXLEY, A. **Point counter point**. London: Flamingo, 1994. (Coleção Modern Classic).

HUXLEY, A. **Contraponto**. Trad. Érico Veríssimo. Porto Alegre: Editora Globo, 1971. (Tradução de: *Point counter point* – Coleção Imortais da Literatura).

LISPECTOR, C. **A hora da estrela**. Rio de Janeiro: Rocco, 1998.

LISPECTOR, C. **The hour of the star**. Trad. Giovanni Pontiero. New York: New Directions Books, 1992. (Tradução de *A hora da estrela*).

MORRISON, T. **Beloved** . New York: Alfred A. Knopf, 1998. 275p.

MORRISON, T. **Amada**. Tradução de Evelyn Kay Massaro. São Paulo: Ed. Best Seller, 1987. 321p. (Tradução de: *Beloved*).

RICE, A. **Interview with the vampire**. New York: Ed. Ballantine Books, 1997.

RICE, A. **Entrevista com o vampiro**. Trad. Clarice Lispector. Rio de Janeiro: Rocco, 1996. (Tradução de: *Interview with the vampire*).

Artigo recebido em: 15.10.2014

Artigo aprovado em: 25.11.2014