

Lilian Jurkevicz Fleuri

**UMA PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE
CORPUS PARALELO BILÍNGUE E DE PEQUENA DIMENSÃO**

Tese submetida ao Programa de Pós-graduação em Estudos da Tradução da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Estudos da Tradução.

Orientador: Profa. Dr^a. Maria Lúcia B. Vasconcellos

Coorientador: Prof. Dr. Lincoln Fernandes

Florianópolis
2013

Ficha de identificação da obra elaborada pelo autor
através do Programa de Geração Automática da Biblioteca Universitária
da UFSC.

Fleuri, Lilian

Uma Proposta Metodológica para Compilação de Corpus Paralelo
Bílingue e de Pequena Dimensão / Lilian Fleuri ; orientadora,
Maria Lúcia B. Vasconcellos ; coorientador, Lincoln Fernandes. -
Florianópolis, SC, 2013. 425p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro
de Comunicação e Expressão. Programa de Pós-graduação em
Estudos da Tradução.

Inclui referências

1. Estudos da Tradução. 2. Estudos da Tradução Baseados em
Corpus. 3. Metodologia de Corpus. 4. Programação. I. B.
Vasconcellos, Maria Lúcia. II. Fernandes, Lincoln. III.
Universidade Federal de Santa Catarina. Programa de Pós-
Graduação em Estudos da Tradução. IV. Título.

Lilian Jurkevicz Fleuri

**PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE CORPUS
PARALELO BILÍNGUE DE PEQUENA DIMENSÃO**

Esta Tese foi julgada adequada para obtenção do Título de Doutor, e aprovada em sua forma final pelo Programa Pós-graduação em Estudos da Tradução.

Florianópolis, 09 de dezembro de 2013.

Profa. Dra. Andréia Guerini
Coordenadora do Curso

Banca Examinadora:

Prof^a Dr^a Maria Lúcia B. Vasconcellos / Orientadora
Universidade Federal de Santa Catarina

Prof. Dr. Lincoln Fernandes / Orientador
Universidade Federal de Santa Catarina

Prof. Dr. Marco Rocha
Universidade Federal de Santa Catarina

Prof. Dr. Marcos Morgado
Universidade Federal de Santa Catarina

Profa. Dra Adriana S. Pagano
Universidade Federal de Minas Gerais

Prof. Dr. Marcos Feitosa
Universidade Federal de João del Rei

Prof. Dr. Marie-Hélenè C. Torres / Suplente
Universidade Federal de Santa Catarina

Este trabalho é dedicado aos cinco pilares que sustentam minha vida: eu, minha família, meu marido, meus amigos e meu trabalho e estudos.

AGRADECIMENTOS

Faz alguns anos que concluí que nossa vida funciona fundada em uma base composta por cinco pilares, sobre os quais se constituem nossas identidades, igualmente importantes que são representados por: nossa vida familiar, vida social, vida amorosa, vida profissional e vida pessoal. A partir dessa conclusão, passei a equilibrar esses cinco pilares dando igual tempo e valor a minha relação comigo mesma, com minha família, com meus amigos, com meu marido e com meu trabalho. Tentando praticar essa filosofia, consegui tocar minha vida tranquila e feliz nessa trajetória de cinco anos que compôs meu doutorado. Portanto, sou grata a mim mesma, por respeitar meus limites e acreditar que sou capaz e seguir caminhando apesar dos obstáculos, os quais só me desafiaram a crescer. Sou grata a minha família—as minhas inspirações (pais, irmão, avós e tios) e minhas motivações (filhos e sobrinhos), que me apresentaram e apresentão ao mundo, ensinando-me valores, virtudes e guiando-me. Aos meus amigos – “psicólogos de botequim”, com quem contei em muitos momentos divertidos ou difíceis de minha vida, permitindo-me simplesmente ser. Ao meu marido, que na criação de uma intimidade constrói comigo a aventura de uma vida a dois e em família. Ao meu trabalho e estudos, que me abriram um mundo de desafios e de questionamentos, levando-me a ver que o inatingível sim se alcança. Presentes nessa rede de relações, agradeço em especial:

Minha orientadora e amiga **Prof^ª Dr^ª Maria Lúcia B. Vasconcellos** – companheira acadêmica que, durante dez anos de minha vida, formou-me com sua sabedoria, humildade e entusiasmo. Muito obrigada!

Meus pais, **Reinaldo Fleuri** e **Dionisia J. Fleuri**, por sempre acreditarem em mim e seu incondicional apoio ao longo de toda minha vida! A vocês minha infinita gratidão.

Meu marido, **Victor Carreto**, que me acompanhou, não apenas na construção de nossa família, como também na construção desta tese. Victor foi, durante esse processo, um grande companheiro que através de diálogos me ajudou a refletir sobre a tese, questionando-me e guiando-me a buscar respostas. Muito obrigada por todo o apoio, a torcida, a força e a parceria durante esses cinco maravilhosos anos!

Meus filhos, **Pedro Fleuri Carreto** e **Matias Fleuri Carreto**, que me acompanharam desde o primórdio de suas vidas todos os passos de meu doutorado. Meus meninos, são e serão sempre minhas inspirações e motivações.

Meu irmão, **Leonardo Fleuri**, por simplesmente ser quem é e estar ao meu lado por mais distante ou perto que estivesse.

Minha avó querida, **Carmen Luna Fleuri**, quem admiro e em quem me inspiro, por sua alegria, fé e força.

Meus sogros, **Marcela Pavón** e **Victor Carreto Valadez**, meus cunhados, **Barbara Carreto** e **Rodolfo Cuevas** e sobrinhos **Barbara** e **Maximiliano**, por representarem em apoio uma linda e deliciosa extensão de minha família.

Todos meus **amigos queridos** que tenho espalhados pelo Brasil, México, Estados Unidos, Canadá, França e Austrália. Que por mais distantes que estivessem, levaram-me a estar onde estou. Não cito nomes para não correr o risco de esquecer de alguém, mas quem lê estas palavras sabe exatamente que é a quem me refiro.

A **Capes**, por me propiciar durante quase dois anos de doutorado a bolsa Reuni, que me ajudou não só financeiramente como também a me introduzir à atividade profissional de pós-graduação.

A **University of Queensland** e ao **Departamento de Promoção de Língua Portuguesa** do governo Federal brasileiro, por me oferecer o incentivo profissional na consolidação da minha prática de ensino e de pesquisa acadêmica.

A todas as pessoas a quem enviei e-mails, perguntas, pedi favores e que com muito carinho e prontidão me ajudaram. São muitos nomes nessa lista e odiaria esquecer alguém, então nomeio meus colegas de pesquisa que me enviaram suas dissertações, textos e outros tipos de ajudas. Obrigada **Daniel Alves**, **Marcos Feitosa**, **Osilene Cruz**, **Roberto Carlos Assis**, **Adriana Pagano**, **Célia Magalhães**, **Roberta**, **Arianna** e todos os autores de teses e dissertações aqui mencionados.

A todos os **membros da banca** que leram atenciosamente minha tese e contribuíram com sugestões enriquecedoras.

A todos vocês, meus sinceros agradecimentos.

Lilian Jurkevicz Fleuri

“A tarefa não é tanto ver aquilo que ninguém viu,
mas pensar o que ninguém ainda pensou sobre
aquilo que todo mundo vê.”
(SCHOPENHAUER, 2008)

RESUMO

Localizando-se no contexto de pesquisa do projeto CORDIAL (Corpus Discursivo para Análises Linguísticas e Literárias) (UFMG) e do Grupo de Pesquisa TRACOR (Grupo de Pesquisa em Tradução e Corpora) (UFSC), esta tese apresenta uma proposta metodológica de compilação de corpus, pautada no perfil e nas necessidades emergentes em 20 dissertações de mestrado produzidas entre 2003 e 2010 em Estudos da Tradução, utilizando simultaneamente o arcabouço metodológico de corpus e teórico da Linguística Sistêmico-Funcional. Nesta tese a Linguística de Corpus oferece os fundamentos teórico-metodológicos, no que tange à Metodologia de Corpus (Barnbrook, 1996; Kennedy, 1998; Bowker, 2001; Mason, 2008) e os Estudos da Tradução Baseados em Corpus (ETBC) oferecem os fundamentos teórico-metodológicos voltados aos Estudos da Tradução (cf. Baker, 1995; Olohan, 2004; Vasconcellos, 2009; Assis, 2012; Feitosa, 2005; Fernandes, 2006). Para o desenvolvimento da pesquisa, são adotados procedimentos metodológicos para: a coleta e a descrição dos métodos das 20 pesquisas analisadas; a criação de uma proposta metodológica, por meio de *Templates* em Processadores de Texto, como o MS Word, e Aplicativos de Planilhas, como o MS Excel; e o teste destes *Templates* em um Estudo Piloto. A análise do perfil metodológico das 20 dissertações identifica que a maioria das dissertações passa pelos seguintes processos de compilação de corpus: (i) preparação do corpus para um alinhamento semiautomático; (ii) alinhamento das unidades desalinhadas; (iii) anotação do corpus e revisão da anotação e (iv) quantificação dos dados. A análise dessas dissertações evidencia, entretanto, algumas inconsistências metodológicas nos processos de compilação do corpus, o que pode comprometer a eficiência e a continuidade da pesquisa, tais como: (a) ausência de padronização metodológica que implica desperdício de tempo de pesquisa na fase de compilação do corpus; (b) excesso de etapas metodológicas para a execução de um único procedimento; (c) demasiadas transições de um programa a outro durante a fase de compilação do corpus e (d) produção de inúmeros documentos. Com base nos resultados dessas análises, propõe-se um método mais eficiente, que implica em: (a) agilizar os processos de compilação de corpora paralelos bilíngues de pequena dimensão, (b) reduzir o número das etapas em cada processo, (c) reduzir o número dos programas utilizados; (d) reduzir o número da geração de diferentes documentos e (e) flexibilizar o modo de anotação do corpus. O modelo metodológico desenvolvido em um *Template* no processador de texto e

outro em um aplicativo de planilha. Esses *Templates* são testados em um Estudo Piloto e seus resultados comparados ao método de compilação adotado em Fleuri (2006), tendo utilizado ambos o mesmo corpus. Os resultados desse Estudo apontam para a produtividade e viabilidade da proposta metodológica, que mostra ser capaz de agilizar o processo (i) de preparação do corpus para o alinhamento, (ii) de correção das unidades desalinhadas no corpus paralelo, (iii) de quantificação dos dados, assim como, ser capaz de flexibilizar o código de anotação do corpus e de organizar a exposição em tabela e gráficos dos elementos anotados e dos dados quantificados. O Estudo Piloto, em comparação a Fleuri (2006), mostrou ter reduzido a $\frac{1}{4}$ o tempo total de compilação, a metade o número de etapas totais executadas, a $\frac{1}{5}$ o número de programas utilizados e a menos da metade o número de documentos gerado.

Palavras-chave: 1. Estudos da Tradução Baseados em Corpus 2. Metodologia de Corpus 3. Programação de Macros 4. Corpus paralelo bilíngue 5. Corpus de pequena dimensão 6. Aplicativos de Planilhas 7. Processadores de Texto

ABSTRACT

Abstract

Within the context of research in the project CORDIAL and TRACOR developed at UFMG and UFSC (Brazil), this PhD thesis presents a methodological proposal for corpus compilation, based on the profiles and the needs of 20 Master's thesis in the interface with Translation Studies, Corpus Linguistics, and Systemic-Functional Linguistics, developed in Brazil between 2003 and 2010. The methodological and theoretical concepts that conduces this thesis are presented by the Corpus-Based Method presented in Corpus Linguistics (Barnbrook, 1996 Kennedy, 1998; Bowker, 2001; Mason, 2008), and the Corpus-Based Translation Studies (cf. Baker, 1995; Olohan, 2004; Vasconcellos., 2009; Assis, 2012; Feitosa, 2005; Fernandes, 2006). The methods followed in this research consist in: collecting and describing the 20 MA thesis mentioned before; studying their methods; creating a fast and economic methodological proposal for corpora compiling, using programming features of Word Processor and Spreadsheet Application; and testing this method in an Pilot Study. The analysis of the methodological profile of these 20 Master thesis identifies that their parallel bilingual small corpus compilation processes are the following: (i) corpus preparation for a semi-automatic alignment; (ii) alignment; (iii) corpus annotation and annotation edition; and (iv) data quantification. Nevertheless, the analysis points to methodological inconsistencies in the processes of corpus compilation, which can impair the investigation itself or the continuation of it in further studies. The inconsistencies concern the: (i) high amount of time spent on the compilation processes; (ii) high number of stages involved in just one process; (iii) high number of transitions between different software; and (iv) high production of documents. Based on these results, this Thesis proposes to solve such inconsistencies by creating an efficiate method of corpus compilation, that aims to: (i) fasten the compilation processes; (ii) reduce the number of stages involved in each process; (iii) reduce the number of software accessed during the corpus compilation; (iv) reduce the number of production of different documents; and plus (v) to turn the corpus annotation more flexible. The proposal is developed on MS Office software (MS Word and MS Excel). Templates with Macros and Formula are created and tested in a Pilot Study, whose results are compared with the corpus compilation results in Fleuri (2006). The methodological proposal reveals to fasten the process of (i) corpus preparation for the alignment; (ii) alignment; (iii) data

quantification; and to make the corpus annotation more flexible and the data display more organized (in tables and graphics). The Pilot Study, comparing to Fleuri (2006), reduced to $\frac{1}{4}$ the total time involved in the corpus compilation; to $\frac{1}{2}$ the total number of stages involved in the corpus compilation; to $\frac{1}{5}$ the total number of transitions among different software and to less than $\frac{1}{2}$ the number of document produced.

Keywords: 1. *Corpus Translation Studies* 2. *Corpu-Based Methodology* 3. *Programming Macros* 4. *Prallel Bilingual Corpus* 5. *Small Corpus* 6. *Spreadsheet Application* 7. *Word Processor*

LISTA DE FIGURAS

Figura 1: Tipologia de Baker (1995) para Corpora em ET	74
Figura 2: Critérios de Classificação para Corpora in ET	75
Figura 3: Ferramentas e utilidades do <i>WordSmith Tools</i> (Scott, 2011).....	91
Figura 4: Ilustração do cruzamento da lista de Assis (2012) e da de Alves e Vasconcellos (no prelo), gerando Lista AAV.	100
Figura 5: Exemplo da Planilha Excel gerada para criação de Gráficos.....	107
Figura 6: Elementos adicionados ao corpus para teste.....	112
Figura 7: Tela do <i>Find & Replace</i> – Formatos e Especiais.....	113
Figura 8: Pasta de Trabalho de Rotulação do TF do <i>Template Corpus Paralelo</i>	122
Figura 9: Tela <i>Template Corpus Paralelo</i> – pasta “Listas e Gráficos”	123
Figura 10: Tela da quinta área de trabalho (pasta) do <i>Template Corpus Paralelo</i>	124
Figura 11: Como ativar os macros do <i>Template Editor</i> no Windows 8.	217
Figura 12: Resultado do texto processado no <i>Template Editor</i>	218
Figura 13: Tela da primeira pasta, pasta de alinhamento, do <i>Template Corpus Paralelo</i>	220
Figura 14: Tela de como passar textos da coluna B para a A (pasta 1 <i>T.C.P</i>)	222
Figura 15: Tela – ativando a formatação da coluna A (texto recém-colado) ...	222
Figura 16: Tela – ativando a formatação da coluna A (texto formatado).....	223
Figura 17: Tela de formatação da coluna A para alinhamento (parte 1)	224
Figura 18: Tela de formatação da coluna A para alinhamento (parte 2)	224
Figura 19: Tela – busca por desalinhamento.....	225
Figura 20: Tela – ativando a função de <i>Juntar</i> do T. C.Paralelo (parte 1)	226
Figura 21: Tela – ativando a função de <i>Juntar</i> do <i>Template Corpus Paralelo</i> (parte 2)	226
Figura 22: Tela – ativando a função de <i>Separar</i> do <i>Template Corpus Paralelo</i> (parte 1)	227
Figura 23: Tela – ativando a função de <i>Juntar</i> do T. Corpus Paralelo (parte 2)	228
Figura 24: Tela – ativando a função de <i>Juntar</i> do <i>Template Corpus Paralelo</i> (parte final)	229
Figura 25: Tela – inserindo linha em branco (parte 1).....	230
Figura 26: Tela - inserindo linha em branco (parte 2).....	230
Figura 27: Tela - inserindo linha em branco (parte 3).....	231
Figura 28: Tela - inserindo linha em branco (parte 4).....	231
Figura 29: Tela <i>Template Corpus Paralelo</i> – formatação final, colorindo linhas em branco	232
Figura 30: Tela T. Corpus Paralelo– pasta “Rotulação TF”	233
Figura 31: Tela T. Corpus Paralelo– Classificação dos elementos analisados.	235
Figura 32: Tela <i>Template Corpus Paralelo</i> – gerando gráficos (parte 1)	235
Figura 33: Tela <i>Template Corpus Paralelo</i> – gerando gráficos (parte 2)	236
Figura 34: Tela <i>Template Corpus Paralelo</i> – pasta “Listas e Gráficos”	237

Figura 35: Tela da quinta pasta do <i>Template Corpus Paralelo</i>	238
Figura 36: Tela – alinhamento gerado em Fleuri (2006)	244
Figura 37: Tela Template Corpus Paralelo – alinhamento gerado no estudo piloto.....	245
Figura 38: Tela – tabela para revisão das classificações em Fleuri (2006)	250
Figura 39: Tela Template Corpus Paralelo - tabelas e gráficos gerados no Estudo Piloto	252

LISTA DE QUADROS

Quadro 2: CROSF-14 Metafunção Ideacional.....	88
Quadro 3: Processos de planejamento e execução de uma pesquisa em corpus— do projeto de pesquisa à exposição dos dados.	96
Quadro 7: Planilha de <i>Excel</i> com <i>Lista AAVF</i> para a geração da <i>Lista ESFTBC</i>	103
Quadro 8: as cinco listas desta pesquisa: duas consultadas, a <i>Lista Assis</i> e a <i>Lista Alves e Vasconcellos</i> , e três geradas, a <i>Lista AAV</i> , a <i>AAVF</i> e a <i>ESFTBC</i>	104
Quadro 10: Informações coletadas e descritas em cada dissertação em <i>ESFTBC</i>	105
Quadro 12: combinações de <i>Wildcards</i> para substituições e respectivos <i>shortcuts</i> gravados em <i>Macros</i>	115
Quadro 13: Comandos e seus <i>shortcuts</i> – Pasta Alinhamento do <i>Template</i> <i>Corpus Paralelo</i>	118
Quadro 15: Resumo das características e procedimentos metodológicos em ..	129
Quadro 16: Resumo das características e procedimentos metodológicos em Mauri (2003).....	131
Quadro 17: Resumo das características e procedimentos metodológicos em Jesus (2004).....	133
Quadro 18: Códigos de anotação criados utilizados por Jesus (2004)	135
Quadro 19: Resumo das características e procedimentos metodológicos em Assis (2004).....	137
Quadro 20: Resumo das características e procedimentos metodológicos em Bueno (2005).....	140
Quadro 21: Resumo das características e procedimentos metodológicos em Rodrigues (2005).....	142
Quadro 22: Resumo das características e procedimentos metodológicos em Morinaka (2005).....	144
Quadro 23: Resumo das características e procedimentos metodológicos em Cançado(2005)	148
Quadro 24: Verbos de elocução encontrados no <i>WordList</i> (Cançado, 2005) ..	151
Quadro 25: Quadro de análise em Cançado (2005)	152
Quadro 26: Resumo das características e procedimentos metodológicos em Feitosa (2005).....	153
Quadro 27: Resumo das características e procedimentos metodológicos em Paquilin (2005)	155
Quadro 28: Resumo das características e procedimentos metodológicos em Zuniga (2006).....	157
Quadro 29: Resumo das características e procedimentos metodológicos em Fleuri (2006).....	160
Quadro 30: Resumo das características e procedimentos metodológicos em Zanella (2006)	163

Quadro 31: Resumo das características e procedimentos metodológicos em Alves (2007)	166
Quadro 32: codificação adotada por Alves (2007)	169
Quadro 33: Resumo das características e procedimentos metodológicos em Figueredo (2007)	170
Quadro 34: Resumo das características e procedimentos metodológicos em Araújo (2007)	173
Quadro 35: símbolos de análise adotados por Araújo (2007)	175
Quadro 36: Resumo das características e procedimentos metodológicos em Filgueiras (2007).....	176
Quadro 37: Resumo das características e procedimentos metodológicos em Fernandes (2009)	178
Quadro 38: Resumo das características e procedimentos metodológicos em Pires (2009)	180
Quadro 39: Resumo das características e procedimentos metodológicos em Nunes (2010)	182
Quadro 40: Balanço das Metafunções exploradas nas dissertações em ESFTBC.	185
Quadro 41: Objetos investigados nas dissertações em ESFTBC.	187
Quadro 42: A versão original dos textos que compuseram os corpora nas dissertações em ESFTBC.....	189
Quadro 43: tipo dos corpora investigados nas pesquisas ESFTBC.....	191
Quadro 44: o método de alinhamento nas pesquisas em ESFTBC que utilizaram corpora paralelos bilíngues.	193
Quadro 45: códigos de anotação utilizados nas pesquisas em ESFTBC que anotaram seu corpus.	196
Quadro 46: código criado por Jesus (2004)	199
Quadro 47: Programa utilizado no processamento dos corpora em pesquisas em ESFTBC.	203
Quadro 48: Ferramentas do <i>WST</i> utilizadas no processamento dos corpora na pesquisas em ESFTBC	204
Quadro 49: Quantidade de Ferramentas utilizadas no processamento dos corpora na pesquisas em ESFTBC.....	205
Quadro 50: método e ferramenta do <i>WST</i> para a quantificação dos dados nas pesquisas em ESFTBC	207
Quadro 51: Tempo, o número de etapas e os documentos gerados em Fleuri (2006) —Preparação para alinhamento.....	241
Quadro 52: Tempo, o número de etapas e os documentos gerados no Estudo Piloto—Preparação para alinhamento.....	242
Quadro 53: Comparação Fleuri (2006) / Estudo Piloto no <i>Template Editor</i> do processo de preparação para alinhamento.....	243
Quadro 54: Tempo, o número de etapas e os documentos gerados em Fleuri (2006)—Alinhamento.....	244
Quadro 55: Tempo, o número de etapas e os documentos gerados no Estudo Piloto—Alinhamento.....	246

Quadro 56: Comparação Fleuri (2006) / Estudo Piloto no <i>Template Corpus Paralelo</i> do processo de alinhamento.	246
Quadro 57: Tempo, o número de etapas e os documentos gerados em Fleuri (2006)—Alinhamento.	248
Quadro 58: Tempo, o número de etapas e os documentos gerados no Estudo Piloto—Classificação dos objetos analisados.	249
Quadro 59: Comparação Fleuri (2006) / Estudo Piloto no <i>Template Corpus Paralelo</i> do processo de anotação de corpus e classificação dos objetos analisados.	249
Quadro 60: Tempo, o número de etapas e os documentos gerados em Fleuri (2006)—Revisão de Etiquetas.	251
Quadro 61: Tempo, o número de etapas e os documentos gerados no Estudo Piloto— Revisão de Etiquetas.	253
Quadro 62: Comparação Fleuri (2006) / Estudo Piloto no <i>Template Corpus Paralelo</i> do processo de revisão da anotação do corpus e da classificação dos objetos analisados.	253
Quadro 63: Comparação Fleuri (2006) / Estudo Piloto no <i>Template Corpus Paralelo</i> do processo de quantificação dos dados e de organização de sua exposição.	254
Quadro 64: Total das etapas nos métodos de compilação em Fleuri (2006).	255
Quadro 65: Total das etapas nos métodos de compilação no Estudo Piloto.	256
Quadro 66: as cinco listas desta pesquisa: duas consultadas, a Lista Assis e a Lista Alves e Vasconcellos, e três geradas, a Lista AAV, a AAVF e a ESFTBC.	261
Quadro 67: Relação ano/produção de 20 dissertações em ESFTBC no Brasil entre 2003 e 2010.	264

LISTA DE GRÁFICOS

(ii) Gerar gráficos a partir desta tabela. Gráfico 1: Balanço das pesquisas da <i>Lista AAVF</i> (em ESFTBC e ESFT).....	104
Gráfico 2: Balanço das Metafunções exploradas nas dissertações em ESFTBC	186
Gráfico 3: Objetos investigados nas dissertações em ESFTBC.	187
Gráfico 4: A versão original dos textos que compuseram os corpora nas dissertações em ESFTBC.....	189
Gráfico 5: tipo dos corpora investigados nas pesquisas ESFTBC	192
Gráfico 6: o método de alinhamento nas pesquisas em ESFTBC que utilizaram corpora paralelos bilíngues.	194
Gráfico 7: códigos de anotação utilizados nas pesquisas em ESFTBC que anotaram seu corpus.	196
Gráfico 8: Programa utilizado no processamento dos corpora em pesquisas em ESFTBC.	203
Gráfico 9: Ferramentas do <i>WST</i> utilizadas no processamento dos corpora na pesquisas em ESFTBC	205
Gráfico 10: Quantidade de Ferramentas utilizadas no processamento dos corpora na pesquisas em ESFTBC.....	206
Gráfico 11: método e ferramenta do <i>WST</i> para a quantificação dos dados nas pesquisas em ESFTBC	207
Gráfico 12: Comparação do total das etapas metodológicas entre o Estudo Piloto e Fleuri (2006).	256
Gráfico 13: Relação ano/produção de 20 dissertações em ESFTBC no Brasil entre 2003 e 2010.....	263
Gráfico 14: Comparação entre o Estudo Piloto e Fleuri (2006).	266

LISTA DE EQUAÇÕES

Equação 1: Exemplo do primeiro teste do <i>Template Editor</i>	115
Equação 2: Fórmulas para se buscar desalinhamento entre textos em Português e Inglês—Pasta Alinhamento no <i>Template Corpus Paralelo</i>	120
Equação 3: Fórmula que transfere o nome da etiqueta ao lado do elemento etiquetado, nas pastas ‘Rotulação TF’ e ‘Rotulação TA’	123
Equação 4: Fórmulas que quantificam os nódulos “ <i>translator</i> ” e “tradutor”, na pasta ‘Dados’	124
Equação 5: Fórmulas que quantificam os elementos analisados, na pasta ‘Dados’	125
Equação 6: Eliminar parágrafos com o <i>Find & Replace</i>	216
Equação 7: Eliminar quebras manuais de página com o <i>Find & Replace</i>	216

LISTA DE ABREVIATURAS E SIGLAS

AAV - Assis, Alves e Vasconcellos
 AAVF – Assis, Alves, Vasconcellos e Fleuri
 ABRAPT – Associação Brasileira de Pesquisadores em Tradução
 ANPOLL – Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística
 BNC – British Nacional Corpus
 CALL – Computer Assisted Language Learning
 CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
 CNPq – Conselho Nacional de Conhecimento Científico e Tecnológico
 COPA-LIJ - Corpus Paralelo de Literatura Infanto-Juvenil
 COPA-MET - Corpus Paralelo de Meta-Discurso em Tradução
 COPA-MUM - Corpus Paralelo de Multimodalidade
 COPA-RAC - Corpus Paralelo de Resumos Acadêmicos
 COPA-TRAD - Corpus Paralelo de Tradução e quatro subcorpora
 CORDIALL - Corpus Discursivo para Análises Linguísticas e Literárias
 CORPRAT - Corpus Processual para Análises Tradutórias
 CROSF – Código de Rotulação Sistemico Funcional
 DAMSL - Dialog Act Markup in Several Layers
 DD – Discurso Direto
 DDL – Discurso Direto Livre
 DHI – Delayed Human Intervention
 EDT – Estudos Descritivos da Tradução
 EHI – Early Human Intervention
 ESFT – Estudos Sistemico-Funcionais da Tradução
 ESFTBC – Estudos Sistemico-Funcionais da Tradução Baseados em Corpus
 ET –Estudos da Tradução
 ETBC – Estudos da Tradução Baseados em Corpus
 EUA – Estados Unidos da América
 FALE - Faculdade de Letras
 GBP – British Pounds
 GST – Gramática Sistemico-Funcional
 HTML - HyperText Markup Language
 INTSINT - International Transcription System for INTonation
 ToBI - Tones and Break Indices
 TSM – Tonetic Stress Marks
 LC – Linguística de Corpus
 LSF – Linguística Sistemico-Funcional

MC – Metodologia de Corpus
MAC – Metodologia Assistida por Computador
MLA - Modern Language Association of America
MS-DOS - Microsoft Disk Operating System
NET – Núcleo de Estudos da Tradução
NILC - Núcleo Interinstitucional de Linguística Computacional
OCR – Optical Character Recognition
OTH – Os Tradutores na História
PDF - Portable Document Format
PPGET –Programa de Pós-Graduação em Estudos da Tradução
PPGI - Programa de Pós-Graduação em Inglês
PROCAD - Programa Nacional de Cooperação Acadêmica
PUC-SP – Pontífca Universidade Católica de São Paulo
RTF – Rich Text Format
SGML - Standard Generalized Markup Language
SJRP - São José do Rio Preto
CRPC - Corpus de referência do português contemporâneo
PORTEXT - Banco de Português
ITA – Instituto Tecnológico de Aeronáutica
SPAAC - SPeech Act Annotated Corpus
SPSS - Statistical package for the social sciences
TCC – Trabalho de Conclusão de Curso
TEI - Text Encoding Initiative
TRACOR – Grupo de Pesquisa Tradução e Corpora
TTH – Translators through History
UECE – Universidade do Estado do Ceará
UFMG – Universidade Federal de Minas Gerais
UFRGS – Universidade Federal do Rio Grande do Sul
UFRJ – Universidade Federal do Rio de Janeiro
UFSC – Universidade Federal de Santa Catarina
UMIST - University of Manchester Institute of Science and Technology
UNB – Universidade de Brasília
UNESP – Universidade do Estado de São Paulo
UNICAMP - - Universidade de Campinas
USP - - Universidade de São Paulo
VBA – Visual Basic for Application
XML - Extensible Markup Language

SUMÁRIO

1 INTRODUÇÃO	34
1.1. JUSTIFICATIVA DE PESQUISA	38
1.2. CONTEXTO DA PESQUISA	41
1.3. OBJETIVOS E PERGUNTAS DE PESQUISA	45
1.3.1. Objetivo geral	45
1.3.2. Objetivos específicos	45
1.3.3. Perguntas de pesquisa	46
2 REVISÃO DE LITERATURA	47
2.1. CAMPOS TEÓRICO-METODOLÓGICOS	48
2.1.1. A interface entre os ETBC e os ESFT	48
2.1.2. A Metodologia de Corpus nos ETBC	52
2.1.3. Estudos de método relacionados à Metodologia de Corpus	57
2.1.3.1. Desenvolvendo <i>Software</i> para pesquisa em corpus (Mason, 2008)	58
2.1.3.2. Desenvolvendo <i>WordSmith Tools</i> (Scott, 2008)	59
2.1.3.3. Desideratas para desenho de <i>software</i> linguístico (Garretson, 2008)	62
2.1.3.4. Reflexões sobre Mason (2008), Scott (2008) e Garretson (2008)	65
2.2. METODOLOGIA DE CORPUS EM ETBC: PROJETO DE PESQUISA E PROGRAMAS PARA PROCESSAMENTO DE CORPUS	67
2.2.1. O projeto de pesquisa em ETBC	69
2.2.1.1. Desenho do corpus	70
2.2.1.2. Compilação do corpus	78
2.2.1.3. Anotação de corpus	83
2.2.1.4. Processamento dos dados	89
2.2.1.5. Exposição dos dados quantitativos	90
2.2.2. Programas prontos	91
2.2.2.1. <i>WordSmith Tools</i> (Scott, 2010)	91
2.2.2.2. <i>AntConc</i> e <i>AntPConc</i> (Anthony, 2010)	93
2.2.3. Programas adaptáveis	94
2.2.4. Reflexões sobre o planejamento e execução de pesquisas em corpus .	96
3 METODOLOGIA	99
3.1. PROCEDIMENTOS PARA A COLETA DE PESQUISAS EM ESFTBC.....	99
3.2. PROCEDIMENTOS PARA DESCRIÇÃO DAS DISSERTAÇÕES	105
3.3. PROCEDIMENTOS PARA A CRIAÇÃO DOS TEMPLATES	108
3.3.1. <i>Template Editor</i>	111
3.3.1.1. Montagem de um corpus-teste	111

3.3.1.2. Estudo dos recursos de edição automática— <i>Find & Replace</i> e a linguagem de <i>Wildcards</i>	113
3.3.1.3. Criação dos códigos de <i>Wildcard</i>	114
3.3.1.4. Realização de testes e fixação da sequência de códigos em <i>Wildcards</i>	114
3.3.2. <i>Template Corpus Paralelo</i>	116
3.3.2.1. Definição do desenho da planilha e de suas pastas de trabalho ..	117
3.3.2.2. Estudo dos programas prontos para réplica de suas ferramentas	117
3.3.2.3. Criação dos <i>Macros</i> e Fórmulas.....	118
3.3.2.4. Teste e reelaboração dos <i>Macros</i>	120
3.3.2.5. Criação de fórmulas, tabelas e gráficos.....	121
3.3.2.6. Edição do desenho das planilhas	125
3.4. PROCEDIMENTOS DE EXECUÇÃO DO ESTUDO PILOTO.....	126
4 ANÁLISE.....	128
4.1. AS DISSERTAÇÕES EM ESFTBC E SEUS MÉTODOS	128
4.1.1. Osilene Cruz (2003)	129
4.1.2. Cristina Mauri (2003).....	131
4.1.3. Silvana de Jesus (2004).....	133
4.1.4. Roberto Assis (2004).....	137
4.1.5. Letícia Bueno (2005).....	140
4.1.6. Roberta Rodrigues (2005)	142
4.1.7. Eliza Morinaka (2005).....	144
4.1.8. Tassiani Cançado (2005)	148
4.1.9. Marcos Feitosa (2005).....	153
4.1.10. Viviane Paquilin (2005)	155
4.1.11. Gleimara Zuniga (2006)	157
4.1.12. Lilian Fleuri (2006).....	160
4.1.13. Ariana Zanella (2006).....	163
4.1.14. Daniel Alves (2007)	166
4.1.15. Giacomo Figueredo (2007).....	170

4.1.16. Cristiano Araújo (2007).....	173
4.1.17. Joanna Filgueiras (2007)	176
4.1.18. Alinne Fernandes (2009).....	178
4.1.19. Thiago Pires (2009)	180
4.1.20. Leonardo Nunes (2010).....	182
4.2. ANÁLISE GERAL DOS MÉTODOS DAS DISSERTAÇÕES EM ESFTBC.....	184
4.2.1. Objetivo da pesquisa: a Metafunção	185
4.2.2. Objeto de análise.....	186
4.2.3. O corpus: versão original dos textos.....	188
4.2.4. O Corpus: tipo.....	190
4.2.5. Método de Compilação: alinhamento.....	193
4.2.6. Código de anotação	196
4.2.7. Método de Processamento: programa adotado	202
4.2.8. Método de Processamento: ferramentas do WST.....	204
4.2.9. Método de Processamento: quantificação dos dados	206
4.2.10. Conclusão: perfil geral das dissertações em ESFTBC	208
4.3. DESENVOLVIMENTO DIACRÔNICO DA METODOLOGIA DAS DISSERTAÇÕES EM ESFTBC.....	209
5. PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE CORPUS.....	212
5.1. <i>TEMPLATE EDITOR</i> : PREPARAÇÃO DO CORPUS PRÉ- ALINHAMENTO	216
5.2. <i>TEMPLATE CORPUS PARALELO</i> : ALINHAMENTO, CLASSIFICAÇÃO DOS OBJETOS ANALISADOS, QUANTIFICAÇÃO E EXPOSIÇÃO DOS DADOS	219
5.2.1. Pasta de Alinhamento	220
5.2.2. Pasta de Rotulação TF e a Pasta de Rotulação TA	233
5.2.3. Pasta de Listas e Gráficos.....	236
5.2.4. Pasta de Dados	237
6 ESTUDO PILOTO: TESTE DOS TEMPLATES.....	239
6.1. PREPARAÇÃO PARA O ALINHAMENTO	240
6.2. EXECUÇÃO DO ALINHAMENTO	243
6.3. ANOTAÇÃO DO CORPUS / CLASSIFICAÇÃO DOS OBJETOS DE ANÁLISE	247

6.4.	REVISÃO DAS ETIQUETAS	250
6.5.	QUANTIFICAÇÃO DOS DADOS E O MODO DE EXPOSIÇÃO	253
6.6.	REFLEXÕES SOBRE O ESTUDO PILOTO	255
7	CONSIDERAÇÕES FINAIS	258
8	BIBLIOGRAFIA.....	270
9	ANEXOS.....	286
	Anexo 1: Lista Assis (2012)	286
	Anexo 2: Lista Alves e Vasconcellos (no prelo).....	287
	Anexo 3: Resumos de teses e dissertações coletadas.....	288
	Anexo 4: Orientações de Dra. Maria Lúcia Vasconcellos até 2011.....	336
	Anexo 5: Orientações de Dra. Adriana Pagano até 2011.....	353
	Anexo 6: Orientações de Célia Magalhães até 2011.....	373
	Anexo 7: Emails trocados com Ariana Zanella	393
	Anexo 8: Perguntas no Fórum MS Office	395
	Anexo 9: Macros, fórmulas e shortcuts <i>Template Corpus Paralelo</i>	411
	Anexo 10: Macros e Shortcuts <i>Template MS Work</i>	419

1 INTRODUÇÃO

Uma nova forma de fazer pesquisa em Estudos da Tradução (ET) iniciou-se no Brasil em 1997 com a publicação da tese de Maria Lúcia Vasconcellos (VASCONCELLOS, 2004) e (VASCONCELLOS, 2009), que utilizou o arcabouço teórico da Linguística Sistêmico-Funcional (LSF) para se analisar textualizações e suas novas construções em retextualizações¹. O projeto CORDIALL² (Corpus Discursivo para Análises Linguísticas e Literárias), criado em 1999, incorporou a Metodologia de Corpus aos Estudos da Tradução em análises descritivas sistêmico-funcionais. A primeira dissertação nessa interface foi defendida por Osilene Cruz em 2003, orientada por Adriana S. Pagano da Universidade Federal de Minas Gerais. A partir de então, a interface entre os Estudos da Tradução (ET), a Linguística Sistêmico-Funcional (LSF) e a Linguística de Corpus (LC) ou Metodologia de Corpus (MC)³ começa a se consolidar como uma tendência de se fazer pesquisa no Brasil, principalmente no âmbito do projeto CORDIALL. Nesta tese, para fins de distinguir dos trabalhos em ETBC e ESFT⁴, utilizo o termo *Estudos Sistêmico-Funcionais da Tradução Baseados em Corpus* (ESFTBC), indicando trabalhos realizados na interface simultânea entre ET, LC e LSF. Identifica-se até 2010 no Brasil a publicação de 24 pesquisas (20 dissertações de mestrado e 4 teses de doutorado) que operam nos ESFTBC, desenvolvidas principalmente em duas universidades

¹ Segundo Fleuri (2006, p.13): “Através da leitura de qualquer texto original, denominado nesta dissertação de textualização, constroem-se imagens e ideias de um determinado evento a partir do modo como esse evento é textualizado (cf. HALLIDAY; MATHIESSEN, 2004, p. 168-305). Ao se traduzir a textualização para uma outra língua, essas imagem e/ou ideias (o perfil ideacional) ganham uma especificidade nova. Por essa razão, adotou-se o termo trazido inicialmente por Coulthard, (1987) e Costa (1992) — “retextualização” — e assim se concebeu que, ao se traduzir um texto para outra língua, há um novo processo de textualização e neste processo novas imagens de um mesmo evento podem ser (re)construídas”.

² O projeto CORDIALL será explicado na seção 1.2.

³ Em alguns casos o termo ‘abordagem baseada em corpus’ e ‘metodologia baseada em corpus’ são utilizados nesta tese com mesmo significado.

⁴ ETBC - Estudos da Tradução Baseados em Corpus—que estabelecem uma interface exclusiva entre os ET e a LC
 ESFT - Estudos Sistêmico-Funcionais da Tradução —que estabelecem uma interface exclusiva entre ET e LSF

brasileiras: a Universidade Federal de Santa Catarina (UFSC) e a Universidade Federal de Minas Gerais (UFMG).

Nesta tese descrevo e analiso os procedimentos metodológicos das vinte dissertações⁵ em ESFTBC publicadas entre 2003 e 2010 no Brasil. Nesta análise observo o perfil e as necessidades metodológicas dessas dissertações em relação aos processos de compilação de corpus. A partir desta identificação, proponho um método de compilação de corpus, passível de ser adotado, reproduzido e adaptado por pesquisas em ETBC.

Tal método não apenas visa a oferecer uma alternativa de se trabalhar eletronicamente com corpora bilíngues de pequena dimensão, como também, visa a tornar acessível economicamente a manipulação automática de corpus e de dados – visto que todos os softwares de processamento de corpus comumente utilizados apresentam um custo relativamente elevado. Por privilegiar a análise metodológica e por propor um método, esta é uma pesquisa essencialmente de cunho metodológico, considerando método “um procedimento regular, explícito e passível de ser repetido” (Bunge apud Moresi, 2003, p.12-15). A menção ao cunho metodológico desta tese, significa também dizer que a investigação se apoia no referencial teórico-metodológico dos ETBC, ou seja, do Método de Corpus voltados a estudos de traduções (cf. Baker, 1995; Olohan, 2004; Vasconcellos, 2009; Assis, 2012⁶; Feitosa, 2005; Fernandes, 2006).

O método proposto busca apresentar uma possível alternativa de lidar com as dificuldades apontadas nas dissertações durante o processo de compilação do corpus relacionados ao: tempo gasto, ao elevado número de etapas executadas em cada processo, às transições entre programas diferentes e à geração de inúmeros documentos. A proposta metodológica se fundamenta na programação de *Macros* e de *Fórmulas* em Processadores de Textos (i.e. *MS Word*) e Aplicativos de Planilhas (i.e. *MS Excel*). Tal programação é gerada em dois *Templates*⁷: um chamado *Template Editor*, criado nesta tese no *MS Word*, e outro chamado *Template Corpus Paralelo*, criado no *MS Excel*. Ambos os *Templates* desempenham a função de automatizar:

⁵ Por limitação de tempo de pesquisa, principalmente, as 4 teses das 24 pesquisas identificadas não são analisadas. Dá-se preferência às dissertações exclusivamente por serem mais numerosas e, portanto, mais representativas.

⁶Fonte online:

<http://seer.bce.unb.br/index.php/traduzires/article/view/6655/5371>

⁷ Disponíveis para download gratuito no site:

<https://sites.google.com/site/lilianjfleuri/>

- (i) A correção e a formatação dos textos a serem alinhados (no *Template Editor*);
- (ii) O alinhamento e a correção de pares desalinhados em um corpus paralelo (no *Template Corpus Paralelo*),
- (iii) A organização da rotulação dos termos analisados (no *Template Corpus Paralelo*);
- (iv) A quantificação simultânea dos dados (no *Template Corpus Paralelo*);
- (v) A organização de uma tabela que auxilia no processo de revisão dos elementos rotulados, arranjando automática e paralelamente os termos e os rótulos correspondentes (no *Template Corpus Paralelo*).

Tal método é testado em um Estudo Piloto com um corpus criado a partir do utilizado por Fleuri (2006). Além de testar o funcionamento dos *Templates*, o Estudo Piloto compara seus resultados aos métodos utilizados por Fleuri (2006) em relação ao tempo, ao número de etapas e de transições entre programas e à quantidade de documentos gerados durante o processo de formatação de textos, alinhamento do corpus, rotulação/correção das etiquetas e quantificação dos dados.

Esta tese apresenta os resultados da pesquisa desenvolvidos em sete capítulos: (i) a introdução, (ii) o referencial teórico, (iii) a metodologia, (iv) a análise, (v) a proposta metodológica, (vi) o estudo piloto e (vii) as considerações finais.

Esta **introdução** justifica e contextualiza a presente pesquisa e expõe os objetivos e as perguntas que a guiaram. A seção 1.1 expõe a justificativa da pesquisa; a 1.2 apresenta o contexto deste estudo, explicando o projeto CORDIALL, assim como o PROCAD e o TRACOR, que situam esta tese na tradição de pesquisas desenvolvidas em ESFTBC; a 1.3 apresenta os objetivos e as perguntas de pesquisas, explicitando os objetivos geral e específicos, bem como as perguntas que guiaram o desenvolver da pesquisa.

O **referencial teórico** é organizado em duas partes. A primeira apresenta um debate teórico sobre a Metodologia de Corpuse os ETBC. A segunda parte apresenta um debate sobre o planejamento e execução de pesquisas em ETBC, explorando os elementos envolvidos na execução de pesquisas em metodologia de corpus, tais como o projeto, o desenho do corpus, as etapas de compilação e os programas de processamento de corpus.

No **capítulo de metodologia** se descrevem os procedimentos de análise. Para poder realizar a análise de 20 dissertações em ESFTBC foi necessário primeiramente coletar essas dissertações, em seguida

descrevê-las buscando traçar seu perfil metodológico, para se planejar e elaborar a proposta metodológica e finalmente testar tal proposta. Da mesma forma, o capítulo de metodologia se configura em quatro etapas que visam a explicar o modo como se deu cada uma dessas etapas de análise e da proposta metodológica. Em 3.1 explicam-se os procedimentos de coleta do objeto de estudo desta tese, ou seja, as dissertações em ESFTBC publicadas até 2010 no Brasil. Em 3.2 expõem-se os procedimentos de descrição e de análise destas dissertações. Em 3.3 descrevem-se os procedimentos de criação dos *Template Editor* e *Template Corpus Paralelo*. Em 3.4 explicam-se os procedimentos de execução do estudo piloto.

O **capítulo de análise** se configura em três seções. Na seção 4.1 se apresentam e se descrevem as dissertações brasileiras em ESFTBC em relação às informações gerais, ao objetivo geral da pesquisa, ao objeto de análise, ao corpus utilizado na pesquisa, ao método de compilação desse corpus e de seu processamento e ao código de anotação adotado. Na seção 4.2 traça-se o perfil metodológico das pesquisas em ESFTBC. Neste momento se observa a proporção, exposta em gráficos, entre as pesquisas em relação à metafunção adotada para a análise do corpus, ao objeto de análise escolhido, ao desenho do corpus, ao modo de alinhamento dos corpora paralelos bilíngues, ao modo de anotação do corpus, ao programa e às ferramentas adotadas no processamento do corpus e ao modo de proceder com relação à quantificação dos dados. Na seção 4.3 se analisa o desenvolvimento diacrônico em relação aos métodos adotados nas dissertações analisadas.

O **capítulo de proposta metodológica** apresenta os *Templates* propostos nesta tese, com explicação detalhada do funcionamento do *Template Editor* e do *Template Corpus Paralelo* para se proceder com a correção automática de textos (no *Template Editor*) e para se alinhar textos em paralelo, assim como para rotular elementos textuais e quantificá-los automaticamente, organizando-os simultaneamente em listas e gráficos (no *Template Corpus Paralelo*).

O **capítulo de estudo piloto** testa os *Templates* propostos e compara os resultados dos testes com os procedimentos adotados por Fleuri (2006) durante a compilação de corpus. Este capítulo é composto por seis seções. As cinco primeiras seções correspondem aos procedimentos testados no estudo piloto e observado em Fleuri (2006), sendo a última seção dedicada às considerações finais de comparação entre ambos os estudos. As seções são: 6.1 Preparação para o alinhamento; 6.2 Execução do Alinhamento; 6.3 Anotação do corpus / Classificação dos objetos de análise; 6.4 Revisão das etiquetas; 6.5

Quantificação dos dados e o modo de exposição e 6.6 Reflexões sobre o Estudo Piloto.

O capítulo de considerações finais encerra a tese resgatando as conclusões de cada capítulo e respondendo pontualmente às perguntas de pesquisa, apontando futuras perspectivas de investigação nesta área.

Nesta tese opto por: (i) traduzir as citações originalmente escritas em inglês, sendo todas as traduções realizadas por mim, (ii) manter alguns termos computacionais em inglês tais como *Wildcards*, *shortcuts*, *Template* etc., destacando-os em itálico; e (iii) expor fórmulas e as programações em *VBA* e em *Macros* em anexo.

1.1. JUSTIFICATIVA DE PESQUISA

A justificativa desta investigação está intrinsecamente relacionada à falta de um modelo metodológico eficiente de compilação de corpus paralelo bilíngue em pesquisas em ESFTBC e à escassez de pesquisas em ETBC dedicadas a desenvolver ferramentas para analisar corpora.

A fase de compilação implica tanto o escaneamento e digitalização dos textos impressos, quanto a correção ortográfica de leitura do *OCR* (*Optical Character Recognizer*), a anotação do corpus, a preparação do corpus para seu alinhamento automático e o alinhamento do corpus paralelo bilíngue. Nota-se que a fase de compilação das pesquisas em ESFTBC, apesar de responder com sucesso a seus objetivos, têm ocorrido de modo intuitivo, com base em tentativas, erros e acertos, levando a um desperdício de tempo de pesquisa e a uma desorganização de documentações. Além disso, a fase de compilação tem demandado o uso de programas de análise lexical pagos, como o *WordSmith Tools* (Scott, 2009), limitando ou impossibilitando a pesquisa em caso de orçamentos curtos. Como a compilação de corpus tem ocorrido de modo espontâneo e armazenado conforme a organização pessoal de cada pesquisador, torna-se difícil o acesso a tais corpora por outros investigadores, sendo difícil ou impossível processar o corpus ou então resgatar os dados da pesquisa. Como exemplo dessa elaboração intuitiva de um método de compilação de corpus tomo minha pesquisa de mestrado (Fleuri, 2006) e minha tentativa de resgate de dados de três pesquisas em ESFTBC que utilizaram o mesmo objeto de análise: Fleuri (2006), Zuniga (2006) e Gysel (2009).

A elaboração da metodologia de minha dissertação⁸ demandou aprender a lidar com ferramentas da LC e da Metodologia de Corpus,

⁸ Este item é mais explorado no Capítulo 6 desta tese.

como programas de análise lexical tais como o *WordSmith Tools* (Scott, 1999) e a lidar com o sistema de codificação para rotulação sistêmico-funcional (CROSF) (Feitosa, 2005), para então poder manipular os dados quantitativos emergentes da análise textual. A elaboração da metodologia, a aprendizagem de métodos e de programas e a compilação do corpus demandaram muito tempo de pesquisa, aproximadamente um ano. A compilação implicou inúmeras transições entre o programa do *WordSmith Tools* (*WST*) e o *MS Word*, gerou inúmeros documentos em Documento *Word* (utilizado para edição do corpus) e *Plain Text* (utilizado para processamento do corpus pelo *WST*) e demandou uma série de edições e reedições: (i) dos textos recém-escaneados (como a correção dos erros de escaneamento principalmente dos termos que seriam etiquetados como “Tradutor” e “Translator”); (ii) do corpus antes do alinhamento (como a eliminação de todos os pontos não marcadores de final de sentença, como era o caso dos pontos em abreviações); (iii) do corpus em alinhamento e já alinhado (como a separação das unidades de alinhamento dos itens analisados do resto das unidades de alinhamento); (iv) do processo de anotação (como a inserção de cada rótulo dentro dos complexos oracionais que apresentavam o item lexical “tradutor” ou “translator”) e (v) dos rótulos já anotados no corpus (como a elaboração de uma tabela com os rótulos e os elementos rotulados para se revisar possíveis erros de anotação). O fato de se perder muito tempo de pesquisa no processo de compilação levou-me a limitar: (i) o tamanho do corpus (compilei apenas cinco dos nove capítulos da versão original e da tradução) e (ii) a descrição textual (detendo-se a apenas a análise da microestrutura, sem associá-la à macroestrutura – conforme pretendia inicialmente fazer, seguindo Tymoczko (1998). Ambas as limitações foram apontadas como futuros estudos.

Apesar de o material compilado e os dados analisados em minha dissertação (Fleuri, 2006) terem cumprido com a proposta da pesquisa, a disponibilização e acessibilidade desse corpus e dados a futuros pesquisadores não puderam ser garantidas, dificultando sua continuação por outros investigadores. Para ilustrar esse caso, busquei resgatar os dados de três pesquisas em ESFTBC—Fleuri (2006), Zuniga (2006) e Gysel (2009). Apesar de apresentarem objetivos semelhantes – analisar a construção do perfil ideacional da entidade textual “tradutor” e “translator”, cada uma dessas pesquisas adotou métodos diferentes de compilação e de exposição dos dados e dos corpora, dificultando, ou até mesmo impossibilitando, a coleta de dados e o resgate do corpus etiquetado. Mesmo sendo coerentes com suas propostas, cumprindo com o que se propunham a realizar, as três pesquisas não dialogaram

metodologicamente: (i) no modo de rotular o item analisado, impossibilitando que se buscasse automaticamente cada tipo de etiqueta, (ii) no modo de expor os dados e resultados, dificultando a soma direta e rápida de cada tipo de Participante e (iii) na forma de apresentar o corpus e os dados analisados nos anexos ou no corpo do texto, impossibilitando o processamento do corpus em programa se análise lexical. Adentrar nas pesquisas de Gysel (2009), Fleuri (2006) e Zuniga (2006) apenas confirmou a importância dessa proposta, para tornar possível o acesso e manipulação de dados por futuros pesquisadores.

Além de pretender colaborar com futuras pesquisas em ESFTBC, a presente tese almeja contribuir com pesquisas em ETBC⁹. Esta contribuição justifica-se ao se oferecer uma forma pouco explorada (cf. Mason, 2008) de lidar com corpus paralelo em programas domésticos, como Suítes de Aplicativos para Escritório, no caso desta pesquisa, o *MS Office*. Como o computador e suas ferramentas são elementos intrínsecos à Metodologia de Corpus (cf. Kennedy, 1998), as formas de se realizar pesquisa envolvendo esta metodologia vai além da análise e descrição textual/linguística/discursiva de corpus, ela abre campo também para grupos de pesquisadores que se dedicam a desenvolver ferramentas para analisar corpora (cf. Kennedy, 1998, p.9; Mason, 2008; Scott, 2008;

⁹ Ao longo do meu doutorado, foram apresentadas três comunicações em congressos sobre uma proposta de sistematização metodológica para pesquisas em ESFTBC: o primeiro foi no *VI Congresso da ALSFAL* (<http://www.uece.br/posla/index.php/noticias/14-lista-de-noticias/110-vi-congresso-da-alsfal-sera-em-fortaleza->) em 2010, em Fortaleza (Fleuri, 2010); o segundo foi uma palestra no *I Seminário de Pesquisas em Tradução e Corpora*, na Universidade Federal de Santa Catarina em 2011 (<http://noticias.ufsc.br/2011/05/25/i-seminario-de-pesquisa-do-grupo-traducao-corpora/>) (Fleuri & Vasconcellos, 2011); o terceiro no *International Conference on Translation and Cross-Cultural Communication*, em 2011 na Universidade de Queensland (Brisbane/Austrália) (<http://www.slccs.uq.edu.au/index.html?page=156662>) (Fleuri, 2011). A recepção e as respostas foram positivas. A audiência apontou principalmente para a necessidade de uma dinamização do processo de compilação do corpus⁹ em ETBC e LC e reafirmou a dificuldade que se tem na criação de um método envolvendo formas de edição dos textos, criação de códigos de rotulação e, principalmente, de diálogos e continuação com e entre pesquisas anteriores. Segundo a resposta da plateia observei que esses problemas estavam igualmente presentes não apenas na área dos ESFTBC, como também nas áreas dos ETBC e da LC.

Garretson, 2008). Saber que dentre as escolhas dos tipos de programa a se adotar na compilação e no processamento do corpus há também a possibilidade de se criar programas que respondam às necessidades da sua pesquisa pode ajudar muito o pesquisador em ETBC a planejar seu estudo em corpus(cf. Seção 2.2.1 desta tese).

Apesar de ter sido apontada desde 1990 no congresso em Georgetown University nos Estados Unidos (cf. Kennedy, 1998, p.11) a importância de se desenvolver ferramentas para a análise de corpora, observa-se que esta perspectiva de pesquisa tem ainda sido pouco explorada nos ETBC (cf. Olohan, 2004). Nos ESFTBC, observamos que no Brasil há apenas uma pesquisa de cunho metodológico, voltado aos estudos de corpus: a dissertação de Feitosa (2005). Tal investigação cria e propõe um código para a anotação de corpus, oferecendo uma solução ao problema de rotulação que as pesquisas nessa área, principalmente às vinculadas ao projeto CORDIALL, vinham apresentando (cf. Feitosa, 2005). Em relação a propostas de um método de compilação de corpus, não foi encontrada nenhuma pesquisa em LC ou ETBC que mencionasse especificamente a programação de *Macros* e de Fórmulas em Suítes de Aplicativos para Escritório para desempenhar funções tipicamente encontradas em programas prontos de análise lexical (tais como o *WordSmith Tools* e o *AntConc*). Mason (2008) aponta a possibilidade e afirma a escassez de pesquisas neste campo (cf. Seção 2.1.3.1 desta tese).

A justificativa de propor um método de compilação de corpus que seja economicamente acessível se pauta na necessidade de dinamizar o processo de compilação de corpora paralelos bilíngues de pequena dimensão, que dissertações em ESFTBC vêm apresentando até 2010. Visa-se, com tal proposta, a colaborar com futuras pesquisas em ESFTBC, como as desenvolvidas pelo projeto CORDIALL e TRACOR, bem como com o campo teórico-metodológico dos ETBC.

1.2. CONTEXTO DA PESQUISA

A justificativa desta tese pode ser associada aos fatores acadêmicos, relacionados à colaboração a pesquisas no campo dos ETBC e institucionais, relacionado à afiliação desta investigação a uma tradição de pesquisa em ESFTBC. O fator institucional oferece o contexto da presente investigação, ou seja, situa a tese como parte da tradição de investigações envolvidas no projeto CORDIALL (Corpus Discursivo para Análises Linguísticas e Literárias), que abarcou por um determinado momento o PROCAD (Programa Nacional de Cooperação Acadêmica), e

pelo projeto TRACOR (Grupo de Pesquisa em Tradução e Corpora), ao se incentivar e promover pesquisas em ESFTBC e ETBC. Esta seção se dedica a explicar a interface entre os ET, LSF e LC (ESFTBC) e a configuração desses projetos que contextualizam a presente tese.

Pesquisas na interface entre os ET, LSF e LC corroboram com a afirmação de Tymoczko (2002) de que “diferentes campos de disciplinas oferecem diferentes modos de ver, conceber e conhecer fenômenos que não competem, mas que mutuamente se esclarecem e se reforçam” (idem, p.12) e com Halliday (2006) que argumenta que a Linguística Sistêmico-Funcional e a Linguística de Corpus são perspectivas que mutuamente se completam no processo de análise de textos em relação tradutória. A interface entre os Estudos da Tradução, Linguística Sistêmico-Funcional e Linguística de Corpus (ou Metodologia de Corpus) é igualmente justificada por Munday (2002) como sendo campos de pesquisas complementares. Munday afirma que o uso das ferramentas da Linguística de Corpus permite um rápido acesso a itens linguísticos, “resolvendo o problema logístico de lidar com todos os textos” (idem, p.76). Dentro desta conjuntura vem sendo desenvolvida uma tradição de pesquisas na interface simultânea, dentro das abordagens textuais e discursivas da tradução, entre os ET, LSF e LC denominada aqui, para fins de distinção terminológica, de *Estudos Sistêmico-Funcionais da Tradução Baseados em Corpus* (ESFTBC). A Linguística Sistêmico-Funcional informa e auxilia a análise de sistemas multifuncionais em textos (cf. Munday, 2002), enquanto a LC oferece o arcabouço metodológico (cf. Olohan, 2004), para lidar com corpora paralelos e comparáveis (multilíngues ou não), tipicamente presente nos Estudos da Tradução (cf. Baker, 1995 e Fernandes, 2006). O reconhecimento da característica interdisciplinar das três áreas é cada vez mais evidente e as diferentes formas de combinação entre elas vem sendo exploradas cada vez mais na últimas décadas (cf. Assis, 2012; Vasconcellos, 2009). Destas três disciplinas surgem pesquisas que operam em combinações binárias, como as pesquisas em *Sistemas e Corpus*, debatidas por Thompson & Hunston (2006), que combinam a LSF e a LC para estudar elementos linguísticos; em *Estudos da Tradução Baseados em Corpus* (ETBC), combinação inicialmente apontada por Baker (1993/1995) entre a LC e os ET para se estudar elementos típicos de traduções através do método de corpus; e em *Estudos Sistêmico-Funcionais da Tradução* (ESFT), conforme (STEINER e YALLOP, 2001) e Vasconcellos (2009), que combinam a LSF e ET para estudar sistemas multifuncionais de textos em relação tradutória.

Observa-se que esta tradição de pesquisa em ESFTBC se consolidou no Brasil a partir do projeto CORDIALL (Corpus Discursivo para Análises Linguísticas e Literárias), desde 1999. O CORDIALL é um corpus multilíngue (MAURI, 2009), constituído de textos eletrônicos “selecionados através de critérios específicos relacionados a subprojetos de pesquisa implementados pelos pesquisadores do NET [Núcleo de Estudos da Tradução da UFMG] e da UFSC” (CORDIALL, 2010). Atualmente este corpus ultrapassa um milhão de palavras e abarca quatro subcorpora que vêm sendo continuamente atualizados por pesquisas desenvolvidas nas universidades às que está vinculado. Estes subcorpora são (i) corpus paralelo multilíngue; (ii) corpus comparável monolíngue em português brasileiro; (iii) corpus especializado de textos acadêmicos e jornalísticos e (iv) Corpus Processual para Análises Tradutórias, o CORPRAT (Mauri, 2009, p.18). O CORDIALL

Fomenta a realização de trabalhos cujo foco principal é a representação de personagens e a apresentação do discurso em textos ficcionais, ambos os aspectos de interesse para a presente pesquisa, que visa a estabelecer uma correlação entre os elementos linguísticos investigados na representação de personagens e na apresentação do discurso com o ponto de vista narrativo (idem, *ibidem*)

Este projeto é ligado à Universidade Federal de Minas Gerais (UFMG), mais especificamente ao NET (Núcleo de Estudos da Tradução)¹⁰ da FALE (Faculdade de Letras da Universidade Federal de Minas Gerais) e à Universidade Federal de Santa Catarina (UFSC)—PPGET (Programa de Pós-Graduação em Estudos da Tradução) e ao PPGI (Programa de Pós-Graduação em Inglês). Além disso, este projeto é vinculado internacionalmente à Universidad Autònoma de Barcelona e à Universidade de Lisboa.

O PROCAD foi de 2001 a 2005 o projeto responsável pelo vínculo entre as duas universidades brasileiras – UFSC e UFMG:

A concentração do compromisso sistêmico-funcional com Estudos em tradução nestas duas universidades Brasileiras pode ser justificada pelo desenvolvimento de um projeto conjunto mantido pela CAPES que incentiva contatos acadêmicos

¹⁰ Atualmente representado pelo LETRA (Laboratório Experimental de Tradução)

entre pesquisadores de ambas universidades: 'Programa Nacional de Cooperação Acadêmica - PROCAD', vigente desde Julho de 2001 até Dezembro de 2005 (VASCONCELLOS, 2005).

Os acadêmicos vinculados ao PROCAD têm sua investigação em tradução relacionada à estrutura teórica da LSF, com o auxílio da metodologia de corpus. Esse comprometimento teórico-metodológico promovido pelo programa resulta em uma consolidação produtiva de pesquisas tanto na área de ESFT como de ESFTBC. Concluído em 2005, este programa foi bem avaliado, pois logrou

não somente promover a mobilidade entre os pesquisadores e estudantes/pesquisadores-em-formação dos dois programas de pós-graduação (de referência nacional), como também estender a interação para incluir pesquisadores e estudantes de mais um programa, o recém criado PGET/UFSC. Com o apoio oferecido pelo PROCAD, foi possível (i) contribuir para a consolidação dos programas envolvidos; (ii) incrementar a formação de doutores; (iii) estabelecer vínculos de cooperação em termos de coorientação de dissertações e teses; (iv) desenhar e implementar projetos conjuntos; (v) maximizar os recursos humanos e de infraestrutura disponíveis nos programas; (vi) explorar novos tópicos de pesquisa; e, finalmente, (vii) estimular a formação pós-graduanda, com o estabelecimento de uma nova cultura de desenvolvimento de pesquisa em equipe. (PROCAD, 2006)

As pesquisas em ESFTBC analisadas nesta tese, quando defendidas na UFSC, são vinculadas ao CORDIAL através do PROCAD ou ao TRACOR. O projeto TRACOR

Tem por objetivo conduzir pesquisas sobre fenômenos tradutórios utilizando ferramentas com base em corpora eletrônicos disponíveis online. Além disso, o TRACOR atua na produção de conhecimento científico e tecnológico que venha a ser empregado na formação de tradutores (TRACOR, 2013)

Sediado atualmente na UFSC, o projeto TRACOR conta com uma equipe de professores, mestrandos e doutorandos que atuam no Programa de Pós-Graduação em Estudos da Tradução (PPGET) e no Programa de Pós-Graduação em Inglês (PPGI). O projeto abriga o COPA-TRAD (Corpus Paralelo de Tradução e quatro subcorpora: (i) o COPA-

LIJ (Corpus Paralelo de Literatura Infanto-Juvenil); (ii) o COPA-MUM (Corpus Paralelo de Multimodalidade), (iii) o COPA-MET (Corpus Paralelo de Meta-Discurso em Tradução) e (iv) o COPA-RAC (Corpus Paralelo de Resumos Acadêmicos).

É portanto no contexto destes três projetos de pesquisa, com o intuito de contribuir para o arcabouço metodológico de corpus, voltado à análise de corpora paralelos bilíngues de pequena dimensão, que se enquadra a presente tese.

1.3. OBJETIVOS E PERGUNTAS DE PESQUISA

1.3.1. Objetivo geral

Buscando contribuir ao projeto CORDIAL e ao TRACOR, esta investigação visa a criar e oferecer um possível método de compilação de corpus pautado nas tendências observadas e nas necessidades metodológicas emergentes em 20 dissertações de mestrado em ESFTBC produzidas no Brasil, de 2003 a 2010. Busca, assim, oferecer para pesquisadores em corpus um ponto de partida para o planejamento metodológico de suas pesquisa, mostrando uma forma de criar, de modo econômico, flexível e personalizado, um programa que responda as suas necessidades, ao apresentar algumas possibilidades de trabalho dentro de um Sistema Operativo com Processador de Texto e com Aplicativo de Planilhas.

1.3.2. Objetivos específicos

- a) Resgatar e listar todas as dissertações e teses em ESFTBC publicadas no Brasil até 2010 (incluindo as publicadas neste ano), encontradas a partir de uma pesquisa bibliográfica;
- b) Descrever e analisar as características metodológicas de todas as dissertações em ESFTBC no Brasil até 2010, buscando traçar o perfil metodológico dessas dissertações;
- c) Identificar as necessidades metodológicas de compilação de corpus mais emergentes dessas dissertação;
- d) Criar, explicar e oferecer um método que responda à maioria das necessidades encontradas, explorando a programação de *Macros* em Suítes de Aplicativos para Escritório (expecificamente Processadores de Texto e Aplicativos de Planilhas).

1.3.3. Perguntas de pesquisa

São quatro as perguntas de pesquisa (PP) que guiam esta investigação:

- i. **PP1:** Quantas e quais são as dissertações e teses publicadas em ESFTBC no Brasil até 2010?
- ii. **PP2:** Qual é o perfil metodológico das dissertações em ESFTBC no que se refere ao perfil de pesquisa e de compilação de corpus?
- iii. **PP3:** Como se deu o desenvolvimento diacrônico das pesquisas em ESFTBC no Brasil até 2010?
- iv. **PP4:** Como pode ser desenvolvida em um Processador de Texto e Aplicativo de Planilhas uma alternativa eficiente (econômica e rápida), acessível e dinâmica de compilação de corpus que responda às necessidades apontadas pela maioria das dissertações em ESFTBC?

2 REVISÃO DE LITERATURA

(...) in each of the three branches of translation studies, there are two further dimensions that I have not mentioned, dimensions having to do with the study, not of translating and translations, but of translation studies itself. (...) there is a dimension that might be called the methodological or meta-theoretical, concerning itself with problems of what methods and models can best be used in research in the various branches of the discipline (how translation theories, for instance, can be formed for greatest validity, or what analytic methods can best be used to achieve the most objective and meaningful descriptive results), but also devoting its attention to such basic issues as what the discipline itself comprises.)
(HOLMES, 2000)

O presente capítulo de referencial teórico mostra *com que campos teóricos* esta tese dialoga. Os objetivos desta tese (ver Capítulo 1, de Introdução) se fundamentam em premissas metodológicas de estudos de corpus¹¹, como aquelas da Linguística de Corpus (LC) e dos Estudos da Tradução Baseados em Corpus (ETBC), levando-nos a um diálogo com pesquisadores e estudiosos de ambas as áreas. Entretanto, por serem enfocados nos Estudos da Tradução, os diálogos estabelecidos com a LC são meramente metodológicos. A maioria da bibliografia que trata de Metodologia de Corpus menciona a LC e os ETBC (cf. Baker, 1993/1995 e Olohan, 2004).

Este capítulo se divide em quatro seções: (2.1) em que se definem os campos teórico-metodológicos da tese, discorrendo sobre os métodos de corpus em áreas da linguística e da tradução; (2.2) que se discutem particularidades pertinentes ao projetar uma pesquisa em ETBC, tais como o desenho, a compilação e a anotação do corpus, o processamento

¹¹ Entendem-se como Estudos de Corpus todos os estudos que incorporam a Metodologia de Corpus, tais como os estudos na área da LC e dos ETBC.

e a exposição dos dados. Nessa seção também são descritos os programas utilizados para análise lexical: os programas prontos e seu funcionamento e os programas adaptáveis em suas potencialidades.

2.1. CAMPOS TEÓRICO-METODOLÓGICOS

Nesta seção exploram-se os campos teóricos-metodológicos com os quais esta tese dialoga. Esta pesquisa é desenvolvida dentro do contexto dos Estudos da Tradução¹², mais especificamente dos Estudos da Tradução Baseados em Corpus (ETBC)¹³. O objeto de estudo, entretanto, encontra-se na interface dos ETBC com os Estudos Sistêmico-Funcionais da Tradução (ESFT)¹⁴. Já a análise e a proposta desta tese localizam-se nos estudos de corpus¹⁵ voltados para os Estudos da Tradução, que traz em si discussões sobre a Linguística de Corpus¹⁶ e os ETBC, explorados nas seções 2.1.1 e 2.1.2. Na seção 2.1.3, são resenhados três artigos (MASON, 2008); (Scott, 2008) e (GARRETSON, 2008) que dialogam com a fase de análise das metodologias aplicadas as pesquisas em ESFTBC e de proposta de um método para estudos de corpus paralelos bilíngues, no que se refere à programação de ferramentas de análise de corpus.

2.1.1. A interface entre os ETBC e os ESFT

Assim como os ETBC representam a relação da disciplina dos ET com arcabouços metodológicos da LC, os ESFT remete à relação dos ET com o arcabouço teórico da LSF. Com o projeto CORDIALl inicia-

¹² Confira (FROTA, MARTINS e CRISTINA, 1994); Holmes, 2000; (KRUGER, 2002); (Pagano & Vasconcellos, 2003); (WILLIAMS e CHESTERMAN, 2002); (GUERINI e TORRES, 2006); Venuti, 2000, entre outros.

¹³ Confira Baker (1993/1995), Olohan (2004), Bowker (2001), Alves e Vasconcellos (no prelo)

¹⁴ Confira Vasconcellos (2009); Assis (2012); Pagano e Vasconcellos, (2005)

¹⁵ Confira (MCENERY, XIAO e TONO, 2006) sobre o caráter metodológico e teórico dos Estudos de Línguas baseados em Corpus (*Corpus-Based Language Studies*).

¹⁶ Confira (KENNEDY, 1998); (GHADESSY, HENRY e ROSEBERRY, 2001); (BIBER e REPPEN, 1998); (MEYER, 2002); (ANDERMAN e ROGERS, 2008)

se a interface entre ambos os estudos, desenvolvendo pesquisas que investigam problemas relativos aos ET através da metodologia proposta pela LC e a análise linguística sugerida pela LSF.

O caráter interdisciplinar dos Estudos da Tradução permite-a interagir com outras disciplinas, apontando novas áreas de interação¹⁷, como ocorre com o surgimento dos ETBC. Nos anos 1990, as correntes metodológicas da Linguística de Corpus e teóricas dos Estudos Descritivos da Tradução (EDT) deram origem ao que Baker (1993 e 1995) denomina Estudos da Tradução Baseados em Corpus (ETBC), cuja metodologia voltada ao estudo de corpus conduz a descobertas relativas à natureza dos textos traduzidos e de suas características universais (Kruger, 2002, p.79). Em 1993, Mona Baker prevê que a compilação de vários tipos de corpora compostos por textos-originais e por textos-alvos, junto com o desenvolvimento da Metodologia de Corpus, possibilitaria que estudiosos da tradução descobrissem “a natureza dos textos traduzidos como um evento comunicativo mediado” (Baker, 1993, p.243) a partir de investigações de termos, denominados por ela, “universais”¹⁸ da tradução. Para os ETBC a tradução é tida como um texto mediador de eventos comunicativos. A metodologia adotada, baseada no estudo de textos autênticos e representativos em relação tradutória, compilados eletronicamente, é a característica-chave deste estudo. Tymoczko (1998, p.653-657) observa que esta metodologia pode revelar mais que o caráter universal das traduções, permite analisar características micro e macro textuais e combiná-las na interpretação dos elementos tradutológicos estudados. Em outras palavras, os ETBC contemplam tanto o processo quanto o produto tradutório (LAVIOSA-BRAITHWAITE, 1996), considerando os detalhes textuais da tradução assim como os elementos culturais internos e externos ao texto. Essa visão compromete integrar os

¹⁷ cf. (WILLIAMS e CHESTERMAN, 2002)

¹⁸ Segundo (MAURANEN, 2008), o conceito de “universalidade” é por muitos questionado, sendo outros termos e conceitos como “lei” e “tendências” adotados como alternativas menos radicais: “Since the introduction of the idea of universals, the subject has been much debated in Translation Studies. Many scholars have hastened to express their doubts about the feasibility of such features, or dispute the concept, while most agree that it is nevertheless a very interesting idea. To many, the notion of ‘universals’ has clearly been too radical, and alternative terms and concepts like ‘laws’ and ‘tendencies’ have appeared more acceptable. The disputes in part reflect different understandings of the term ‘universal’, but even so there seems to be more conceptual than terminological unity in the field” (Mauranen, 2008, p.34)

estudos linguísticos aos estudos culturais da tradução, envolvendo ideologia e texto/tradução.

A incorporação da LSF em pesquisas empíricas e descritiva dos Estudos da Tradução vem sendo sinalizada desde a década de 1960 por M.K. Halliday (Pagano e Vasconcellos, 2005). Segundo Pagano e Vasconcellos (2005), tal sinalização foi realizada pelo autor em pelo menos quatro momentos: (i) em 1962, quando define a tradução em relação à ordem e, operando no estrato léxico-gramatical, oferece um modelo de tradução assistida por computador; (ii) em 1964, quando afirma que o processo da tradução é uma seleção de categorias e elementos da língua de partida ‘equivalentes’ à categoria e elementos na língua de chegada, sendo que, em níveis linguísticos e contextuais, ambos os textos mantêm uma relação mútua; (iii) em 1985/1994, momentos em que aponta aplicações da Linguística Sistêmico-Funcional à tradução em treinamento de tradutores e intérpretes e em desenho de softwares para traduzir; e (iv) em 2001, quando define parâmetros de uma “boa tradução” e associa mais explicitamente a Linguística Sistêmico-Funcional à tradução, problematizando a “noção de equivalência enquanto um parâmetro vago e abstrato” (Pagano & Vasconcellos, 2005). Essa possibilidade de interface entre os ET com a LSF se desenvolveu em uma tradição de pesquisa no âmbito internacional e nacional, vindo a denominar-se de *Estudos Sistêmico-Funcionais da Tradução* (ESFT) (cf. Vasconcellos, 2009). No contexto internacional, Vasconcellos e Pagano (2005, p.12-13) mencionam trabalhos realizados por: Munday (1998 e 2002), que propõem “o uso de uma abordagem linguística como teoria de análise textual aos Estudos da Tradução”; (MAIA, 1998), que “analisa o uso dos pronomes de primeira pessoa em um romance em português e sua tradução para o inglês e em outro romance em inglês e sua tradução para o português”; (GHADDESSY e GAO, 2001), que, ao usar mais de um corpus restrito, examinam “a organização temática em comentários políticos em inglês, publicados em um livro didático chinês e sua tradução para o chinês”. No contexto brasileiro, os ESFT têm sido explorados desde 1997, com a publicação da tese de Vasconcellos e com a publicação da dissertação de (CAMPESATTO, 2002).

No Brasil, observa-se que no final da década de 1990 começa a ocorrer a interface entre ESFT e ETBC: nota-se a publicação de pesquisas em ETBC que utilizam o arcabouço teórico da LSF (cf. Alves, 2012 e Alves e Vasconcellos, no prelo) e pesquisas em ESFT que se apoiam no arcabouço metodológico da LC (cf. Assis, 2012). A partir de uma pesquisa bibliométrica, Alves (2012), Alves e Vasconcellos (no prelo) e

Assis (2012)¹⁹ resgatam e mapeiam algumas destas pesquisas no Brasil. Assis (2012), que rastreia trabalhos acadêmicos em ET publicados no Brasil vinculados à Linguística Sistêmico-Funcionais (utilizando ou não o arcabouço metodológico da LC), aponta para um total de 25 pesquisas publicadas de 1997 a 2010 no Brasil: “onze dissertações e cinco teses [defendidas] na UFMG e sete dissertações e duas teses na UFSC” (idem, p.64). Alves (2012) e Alves e Vasconcellos (no prelo) resgatam pesquisas em Estudos da Tradução Baseados em Corpus (com ou sem o arcabouço teórico da LSF) em nível de mestrado e de doutorado, publicados no Brasil entre 2006 e 2010. Das 278 pesquisas em ET encontradas, 47 são vinculadas à Linguística de Corpus, representando 16% das pesquisas em tradução deste período²⁰. Destas 47 pesquisas, 16 foram produzidas na UFSC, 10 na UFMG, 5 na UNESP (SJRP), 5 na UFRGS, 4 na USP, 4 na PUC-SP, 2 na UNESP e 1 na UNB. Na presente tese, identifica-se que a partir de 2003, com a publicação da dissertação de Osilene Cruz, as pesquisas, em ESFT passam a analisar textos mais extensos (deixando de ser contos para ser romances na íntegra), justificando a utilização de ferramentas propostas pela Linguística de Corpus. Assim, o arcabouço teórico-metodológico da LC passa a ser incorporado nas pesquisas de abordagens discursivas dos estudos da tradução em interface com a LSF: “Os pesquisadores trabalhando em tal interface acabaram por se tornar os impulsionadores da conjunção LSF e ET, a que vieram por incorporar as metodologias de estudos baseados em corpus” (Fleuri, Pagano, & Vasconcellos, 2009).

Em suma, da combinação de três disciplinas, Linguística de Corpus (LC), Linguística Sistêmico-Funcional (LSF) e Estudos da Tradução (ET) surgiram três áreas de estudos: Sistema e Corpus (fruto da interface entre LSF e LC), Estudos Sistêmico-Funcionais da Tradução (interface entre LSF e ET) e Estudos da Tradução em corpus (interface entre LC e ET). A combinação simultânea destas três áreas é o que nesta tese chamamos de Estudos Sistêmico-Funcionais da Tradução Baseados em Corpora/Corpus (ESFTBC).

¹⁹ Confira anexos 1, 2 e 3.

²⁰ “15,99% das teses e dissertações desenvolvidas no País adotaram metodologias, referenciais teóricos e princípios da Linguística de Corpus – um percentual elevado, considerando que as outras duas opções não agrupadas que o seguem são Traduções Comentadas (com 15,65%) e Traduções Intersemióticas (com 6,12%)” (Alves e Vasconcellos, no prelo).

2.1.2. A Metodologia de Corpus nos ETBC

Desde o surgimento do computador, a forma de se fazer pesquisa em linguística e, conseqüentemente nos ET, tem se transformado. O uso do computador na linguística, segundo Mason (2008, p.142), “mudou de simplesmente ser uma glorificada máquina de digitar artigos a ser uma essencial ferramenta de pesquisa”²¹. Questões práticas sobre ferramentas utilizadas na Metodologia de Corpus (MC) são discutidas em diversas áreas de estudos de linguagem, tais como: Linguística de Corpus (i.e. Kennedy, 1998, p. 57-82) e Estudos da Tradução Baseados em Corpora (i.e. Olohan, 2004, p.1-24). Entendem-se como “estudo de corpus” todos os estudos que incorporam a Metodologia de Corpus (FILLMORE, 1992).

Nessa metodologia o termo *computador* está presente na carga semântica da palavra *corpus*— texto naturalmente ocorridos e legíveis por ferramentas computacionais. O corpus é portanto o objeto de estudo e computador é o meio de pesquisa em campos de estudo que utilizam metodologias de corpus. Essa metodologia é aplicada em áreas que trabalham com descrição e análise linguística em corpora, elaboração, descrição e análise de funcionamento de ferramentas tecnológicas computacionais para trabalho com textos, estudos de aplicações de métodos que envolvam *software* de processamento linguísticos e/ou textuais, etc. O termo *corpus-based*, ou apenas a inclusão da palavra corpus/corpora à uma pesquisa, faz emergir o caráter metodológico do estudo, que em si implica a utilização de ferramentas computacionais para a compilação do corpus, processamento e análise dos dados.

Os termos “*corpus*” e “*computador*” estão então implicitamente conectados ao caráter metodológico de estudos de corpus, como o caso da Linguística de Corpus e dos Estudos da Tradução Baseados em Corpus. Observa-se na bibliografia revisada que a Metodologia de Corpus é aplicada a áreas que investigam textos ocorridos naturalmente através de ferramentas computacionais. Bowker (2001, p.346) aponta quatro principais características da abordagem baseada em corpus:

²¹ “The use of the computer in linguistics has changed from simply being a glorified typewriter for writing up articles to an essential research tool”.

(Mason, Oliver, 2008, p.141-156).

(Acesso online em fevereiro/2013:

http://dialnet.unirioja.es/servlet/listaarticulos?tipo_busqueda=EJEMPLAR&revisa_busqueda=1873&clave_busqueda=201121)

Primeiramente, é baseada na análise de uma coleção comparativamente grande²² e cuidadosamente selecionada de **textos ocorridos naturalmente**, armazenados em uma forma legível por máquinas (i.e. um corpus). Em segundo lugar, pelo fato de analisar **usos reais de padrões de linguagem** no corpus, é empírico e por isso objetivo. Em terceiro lugar, a abordagem baseada em corpus se aproveita de **ferramentas computacionais** e de métodos de manipulação de corpus para organizar dados de formas que torna possível identificar itens e padrões que seriam difíceis de se observar em outros tipos de fontes. Outra vantagem de computadores é que eles oferecem análises consistentes e confiáveis (i.e., eles não mudam de idéia ou se distraem). Por último, a abordagem baseada em corpus combina **técnicas quantitativas e qualitativas**; um computador é capaz de quantificar mecanicamente características linguísticas, mas o aprendiz de tradução [ou o pesquisador de traduções] é responsável por explorar e interpretar estes e outros dados para aprender [descrever e analisar] o uso de padrões de linguagem. (grifo meu)

O *computador*, então, em metodologia de corpus é a parte fundamental dos métodos aplicados nestes estudos. Às disciplinas que utilizam esta metodologia, “o computador introduziu incrível velocidade, total confiança, replicabilidade precisa, estatística confiável e a possibilidade de lidar com uma grande quantidade de dados” (Kennedy, 1998, p.5). Ou pelo menos parte disso se esperaria que um computador realizasse em pesquisas linguísticas que utilizam tal metodologia²³.

²² A característica relacionada ao tamanho da coleção de texto para que possa ser definida ou não como corpus é questionável. Sinclair (2001) afirma que a diferença entre um corpus grande e pequeno deve ser metodológica, já que não se pode falar apenas em tamanho, seja ele absoluto ou relativo: “corpora pequenos são desenhados para intervenção humana antecipada [Early Human Intervention – EHI] enquanto corpora grandes são desenhados para intervenções humanas tardias [Delayed Human Intervention—DHI]” (SINCLAIR, 2001).

²³ A análise das dissertações em ESFTBC aponta que alguns processamentos de corpus continuam lentos, ainda que sejam utilizados programas computacionais de última geração, fazendo com que muitos

Nos Estudos da Tradução Baseados em Corpus, por exemplo, Olohan (2004) argumenta que estes estudos carregam em si um forte caráter metodológico, mais que o estabelecimento de um paradigma:

Destaca-se que é mais produtivo ver o uso de corpora como uma metodologia de pesquisa, com seus pontos fortes e suas limitações, do que vê-lo como um paradigma ocupando um pólo ou outro. (Olohan, 2004, p.3)

Se computador é o meio pelo qual uma pesquisa de corpus é realizada, ou seja, seu uso é intrínseco à metodologia (Kennedy, 2008, p.5)²⁴, e se os ETBC são de cunho essencialmente metodológicos (Olohan, 2004, p.3), o computador e as ferramentas de manuseio de corpus são, então, peças-chaves nos estudos de corpus, podendo passar de instrumento a objeto de estudo. A evolução dessa tecnologia nos estudos de corpus abriu e continua abrindo espaço para que pesquisas tecnológicas sejam associadas a este estudo, no que se refere à relação das análises linguísticas, de resultados e de métodos e ferramentas aplicados, ou de desenvolvimento de programas de processamento de corpus e de dados.

Nota-se que de 1960 a 1990, década em que Kennedy (1998) publica seu livro, *An introduction to Corpus Linguistics*, as tecnologias do computador se aperfeiçoaram enormemente, em termos de armazenamento, velocidade, aplicativos, ferramentas e programas. De 1990 a 2010 as tecnologias se desenvolveram ainda mais. Em duas décadas não apenas evoluíram muito as tecnologias dos PCs e seus sistemas operativos, como também surgiram os dispositivos em formato de tabletes, como *Ipad* e *Iphones* (Apple Inc.), e seus aplicativos. Além disso, foram criados novos programas de processamento de corpus e de dados que respondem às necessidades das novas tendências de pesquisa e de aplicações às pesquisas. O programa de análise lexical *WordSmith Tools* cf. (Scott, 2001) e (Scott, 2008), por exemplo, foi desenvolvido na década de 1990 e continua até hoje sendo aperfeiçoado para atender a ampla demanda em pesquisas que envolvem metodologia de corpus²⁵.

pesquisadores busquem realizar a tarefa manualmente. Ver mais a respeito no Capítulos 4 e 6 desta tese.

²⁴ “Corpus linguistics is thus now inextricably linked to the computer, which has introduced incredible speed, total accountability, accurate reliability, statistical reliability and the ability to handle huge amounts of data”.

²⁵ O site de Mike Scott aponta algumas pesquisas desenvolvidas com e sobre o *WordSmith Tools*: (que usa *WST*)

Nos anos 2000, Laurence Anthony²⁶ desenvolveu um programa de concordância *freeware* (gratuito e disponível para baixar online), com funções semelhantes do *WordSmith Tools*, chamado *AntConc*²⁷, e ainda lançou recentemente, em 2012 e 2013, um programa de concordância paralela (*AntPConc*), de *profiling word* (*AntWordProfiler*), de análise estrutural de text (*AntMover*), entre outros. Ainda na entrada dos anos 2000, (BARLOW, 2003) lançou um programa de concordância paralela chamado *ParaConc*, também amplamente utilizado em pesquisas que realizam alinhamento de corpus e concordância em corpus alinhados.

Esta evolução tecnológica mostra uma abertura do campo de pesquisa focado no estudo de ferramentas computacionais voltadas aos estudos linguísticos. As pesquisas que envolvem Metodologia de Corpus não realizam estudos apenas em análise textuais e linguísticas, elas também se preocupam com a avaliação e desenvolvimento de métodos e de ferramentas computacionais, conforme explica Kennedy (1998, p.9) “o trabalho na linguística de corpus é atualmente associados a diversas atividades”, divididas em quatro grupos de pesquisadores:

O primeiro grupo de pesquisadores envolve os elaboradores de corpus ou compiladores de corpus. Tais acadêmicos se dedicam a desenhar e compilar corpora, a coleção de textos e sua preparação e armazenamento para análises posteriores. / **O segundo grupo de pesquisadores se dedicam a desenvolver ferramentas para analisar corpora.** Importantes contribuições ao desenvolvimento de software, especialmente para a análise sintática de corpora, têm sido associadas particularmente, mas não exclusivamente, a pesquisadores em Linguística Computacional. Esses pesquisadores têm se dedicado ao uso de corpora para desenvolver, entre outras coisas, algoritmos para processamento de linguagem natural e o modelamento de teorias linguísticas. / O terceiro grupo de pesquisadores envolve linguistas

http://www.lexically.net/wordsmith/corpus_linguistics_links/papers_using_wordsmith.htm

(que cita o WST):

http://www.lexically.net/publications/citing_wordsmith.htm

²⁶ <http://www.antlab.sci.waseda.ac.jp/software.html> (Fonte acessada em julho/2013)

²⁷ Confira http://research.ncl.ac.uk/dectete/toon/assets/docs/AntConc_Guide.pdf para obter um guia explicativo de utilização e descrição do AntConc.

descritivos cuja principal preocupação é fazer uso de corpora computadorizados para descrever confiavelmente o lexicon e a gramática da língua, (...). É o aspecto probabilístico de corpus baseado em estudos descritivos da língua que especialmente distingue-os de convencionais campos de estudos descritivos em linguística ou lexicografia. (...) / A quarta área de atividade, que está entre os resultados mais inovativos da revolução de corpus, é a exploração de descrição linguística baseada em corpus no uso de uma variedade de aplicações, tais como aprendizagem e ensino de línguas e processamento de linguagem natural por máquinas, incluindo o reconhecimento da fala e a tradução. (Kennedy, 1998, p.9, grifo meu)

Esse enunciado explica que atualmente a maioria dos pesquisadores concentram-se na descrição linguística baseada em corpus, embora haja grupos que se preocupam em estudar desenho de corpus, e outros que se concentram em estudar métodos para análise e processamento de texto. Olohan (2004) também observa que o caráter metodológico de pesquisas em tradução baseada em corpus permite, por exemplo, imaginar pesquisas sobre a “aplicação da metodologia a diferentes tipos de tradução”. Considerando a explicação de Kennedy sobre os grupos de pesquisadores, esta tese se localiza no segundo grupo de estudo, uma vez que busca oferecer uma alternativa de manipulação em ferramentas computacionais para propor uma metodologia para futuras pesquisas vinculadas a este projeto CORDIAL, realiza, assim, uma meta-reflexão sobre as metodologias baseadas em corpus de pesquisas em Estudos Sistemico-Funcionais da Tradução associadas a tal projeto.

Kennedy (1998) observa que desde a década de 1990 aponta-se para a necessidade para estudos enfocados em metodologias de análise de corpus. Esta necessidade foi apontada na mesa-redonda do encontro sobre linguísticas baseadas em corpus, em março de 1993 na Georgetown University. Segundo Kennedy (1998, p.11) foram identificados tópicos que apresentavam

uma particular necessidade de investigação e de disseminação neste momento em que as linguísticas estavam retornando à uma abordagem mais textual da língua: o desenho e desenvolvimento de corpora de texto-falado; ferramentas de busca e processamento de corpora

on-line; avaliação crítica de corpora on-line e de ferramentas de processamento de corpus; **questões metodológicas em análises baseadas em corpus**; aplicações e resultados em linguísticas e disciplinas relacionadas, incluindo ensino de línguas, linguística computacional, linguística histórica, análise do discurso e análise estilística. (grifo meu)

Nessa época já se notava a necessidade de se trabalhar com questões metodológicas relacionadas a estudos em análises baseadas em corpus. Esse encontro em Georgetown University em 1993 também ressaltou a necessidade de avaliações de ferramentas de processamento de corpus. Já se via desde os anos 1990 essa lacuna nos estudos baseados em corpus no campo das linguísticas. Ainda hoje observa-se uma escassez de estudos na área dos estudos de corpus que se preocupem em analisar e propôr métodos e ferramentas (cf. Mason, 2008). Olohan (2004, p.16) aponta que a orientação não-prescritiva das estruturas teóricas para métodos de corpus em estudos de tradução pressupõem o desenvolvimento de pesquisas majoritariamente enfocadas em análises linguísticas de traduções. Tal lacuna é também observada e constatada na presente tese. Os resultados da análise da aplicação de Metodologia de Corpus em ESFTBC obtidos nesta tese apontam para uma escassez de estudos metodológicos voltados para compilação de corpus (ver seção 1.1 desta tese). Isso se deu, possivelmente, por falta de um estudo que avaliasse a aplicação dos métodos da investigações e reorientasse o processo.

Com base nas discussões acima é possível localizar a presente pesquisa no âmbito dos ETBC a partir do viés metodológico baseado em corpus. Assim como a LC, os ETBC assumem estudos sobre a metodologia de Corpus (Kennedy, 1998 e Olohan, 2004) e sobre as ferramentas computacionais utilizadas em pesquisas de corpus.

2.1.3. Estudos de método relacionados à Metodologia de Corpus

Em 2008 a revista *International Journal of English Studies* publica um volume especialmente dedicado a questões relativas ao uso e desenvolvimento de *softwares* de análise linguísticas, intitulado *Monograph: Software-aided Analysis of Language*. Nessa revista há três artigos que abordam exclusivamente a programação e o desenvolvimento de *software* por linguistas para realizar tarefas voltadas a suas pesquisas. Esta tese dialoga com o artigo de Oliver Mason (2008), *Developing*

Software for Corpus Research, quanto à proposta de programa em programas adaptáveis, desenvolvendo um método economicamente acessível de processamento de corpus, sem necessidade de ser um programador. Quanto ao artigo de Mike Scott (2008), *Developing WordSmith*, o diálogo se estabelece entre linguistas de corpus com interesse e iniciativa de pensar na Metodologia de Corpus a partir das ferramentas de análise. Como linguista ele começou a programar, a princípio de modo rudimentar, o programa que hoje é um dos mais usados nos estudos de corpus—o *WordSmith Tools*. Finalmente, no artigo intitulado *Desiderata for Linguist Software Design*, Gregory Garretson (2008) apresenta uma série de diretrizes para pesquisadores a procura de *software* de análise linguística e para programadores que desenham tais *software*. Esse artigo se comunica com a presente pesquisa quando estabelece um diálogo entre pesquisadores e programadores e a colaboração que um oferece ao trabalho do outro—relacionando principalmente o papel do linguista de corpus na tarefa de programação. Também apresenta uma reflexão sobre os elementos a se considerar, durante a etapa de planejamento da pesquisa, na adoção de um *software* ou na criação de um *software* específico para a pesquisa.

2.1.3.1. Desenvolvendo *Software* para pesquisa em corpus (Mason, 2008)

Mason (2008) aponta que os estudos baseados em corpus (i.e. a Linguística de Corpus) “dependem muito da programação de computador e infelizmente poucos pesquisadores são capazes de desenhar seu próprio *software*, tendo então que depender de programas existentes para fazer seus processamentos” (idem, p.155). Os programas existentes apresentam em geral algumas limitações (idem, p.141)—muitas vezes não respondem às necessidades das pesquisas, levando-as a se adaptarem aos programas existentes, a se apoiarem em um segundo *software* ou a conduzirem grande parte do trabalho de compilação e análise manualmente.

Nesse sentido, conhecer as capacidades dos *software* disponíveis e saber escolher qual programa e ferramentas usar é essencial no planejamento da pesquisa em corpus. As escolhas, ressalta Mason, “dependem da natureza da análise e da disponibilidade do programa” (p.142). Saber que dentre essas escolhas há também a possibilidade de se criar programas que respondam às necessidades da sua pesquisa é uma importante informação que pode ajudar muito o pesquisador a planejar seu estudo (ver seção 2.2 desta tese).

Para fins de atender às necessidades de cada pesquisador, é possível afirmar que nem toda tarefa de programar precisa ser complexa

e envolver um projeto com profissionais de áreas de programação. De fato, menciona Mason, a maioria dos *softwares* de processamento de corpus disponíveis no mercado foram criados por ‘entusiastas amadores’ (i.e. *WordSmith Tools*) (idem, p.144). Mason argumenta que “é facilmente possível adquirir algumas habilidades necessárias para se criar pequenos programas em linguagem de *script*, acelerando muito o processamento de texto e de dados” (idem, p.155). Ele exemplifica essa possibilidade, explicando e demonstrando três possíveis linhas de comando escritas no ‘terminal’ de um Sistema Operativo, como o Windows, Macquintosh ou Linux, para: criar lista de frequência, contar número de palavras terminadas em *-ing*, por exemplo, e encontrar palavras em comum entre dois textos (idem, p. 150-153). Para realizar essas tarefas, como argumenta Mason (idem, ibidem), pode-se usar uma linguagem simples, acessível a iniciantes em processamento, lidando com um sistema operativo familiar ao pesquisador. O que acontece geralmente é que grande parte dos pesquisadores não têm consciência das opções disponíveis e não sabem por onde começar a buscar informação para desenvolver uma simples programação. O artigo de Mason (assim como a presente tese) oferece para pesquisadores em corpus um ponto de partida para o planejamento metodológico de suas pesquisa, mostrando uma forma de criar, de modo econômico, flexível e personalizado, um programa que responda as suas necessidades, ao apresentar algumas possibilidades de trabalho dentro de um sistema operativo, o *Windows* no caso desta tese.

2.1.3.2. Desenvolvendo *WordSmith Tools* (Scott, 2008)

Mike Scott (2008), no artigo intitulado *Developing WordSmith*, descreve, como linguista aplicado ao ensino de línguas e a corpus, sua trajetória na programação do seu *software* de análise lexical de textos, o *WordSmith Tools*. Ele explica, em uma linguagem acessível a linguistas de corpus, o contexto em que se deu o desenvolvimento do programa e os princípios em quais se baseia seu desenho. Comenta que o que o impulsionou a desenvolver tal programa foi o desejo de praticar um *hobby* auto-didata (elaborar ferramentas de computador) as quais poderia relacionar suas teorias, trabalho e interesses (idem, p.96).

Ao explicar como construiu o *WordSmith Tools*, Scott menciona que opta em manter o programa aberto (independente de uma língua específica) para ser usado em pesquisas em diversos idiomas. Portanto, ele escreveu *scripts* que lessem por exemplo *UN* (United Nation) como uma palavra e *Mrs.* como uma sentença, deixando a escolha de

especificação para o usuário (idem, p.98-99)²⁸. Outro caso que menciona foi o de ter optado em mostrar no painel de controle muitas configurações (*settings*) misteriosas, mesmo que isso tornasse o programa mais complexo. Isso fez com que os usuários tivessem a opção de escolher ver ou não os detalhes, já que muitas pesquisas requerem ajustes detalhados (idem, p. 99-100). Tais escolhas exemplificam como um software comercial pode ser mais aberto para aceitar uma vasta gama de variedades de pesquisas. Se, por um lado, isso viabiliza, por exemplo, o uso tanto por pesquisadores lituanios como por pesquisadores hispânicos, por outro lado o nível de intervenção manual aumenta, devido à demanda por especificações para o programa se encaixar ao perfil da pesquisa. Seu relato nos oferece uma ideia das questões a se pensar durante a elaboração de um programa.

Ao relatar sua experiência de linguista-programador, Scott mostra que o linguista de corpus é capaz de aprender a escrever *scripts* de análise textuais (idem, p.104). Entretanto ressalta que poucos linguistas de corpus de fato programam ou analisam métodos em relação à funcionalidade das ferramentas computacionais. A razão disso é explicada por analogia: o linguista de corpus geralmente se coloca para o computador, como o motorista para o carro—a maioria deles estão mais interessados na paisagem que no instrumento em si. O linguista se preocupa em usar a ferramenta computacional para executar a análise textual, e não observa tanto a análise para ver como melhorar a ferramenta, assim como os motoristas: “enquanto o carro está funcionando permanecem na estrada, se querem subir uma colina, tendem a ir caminhando” (idem, p.104). Ou seja, mudamos a estratégia para suprir deficiências da máquina e não mudamos a máquina para suprir as necessidades da pesquisa. Por isso, como observaremos no Capítulo de Análise, há tantos pesquisadores que ao encontrarem dificuldade no alinhar o corpus com o programa, optam por realizá-lo manualmente.

O comentário de Scott corrobora a menção de Olohan (2004) e Kennedy (1998) de que apesar de a Metodologia de Corpus englobar também a descrição das ferramentas e seu funcionamento, a maioria dos linguistas de corpus parecem estar mais interessados em descrever as particularidades textuais e linguísticas do corpus do que as ferramenta—

²⁸ Isso explica porque é praticamente impossível realizar, no WST, o alinhamento completamente automático de um texto em inglês e português, ou seja, sem ter que preparar os textos ou arrumar o alinhamento após processado no WST.

mesmo que o funcionamento e manuseio da ferramenta influencie diretamente o resultado das descobertas linguísticas.

2.1.3.3. Desideratas para desenho de *software* linguístico (Garretson, 2008)

Ao contrário do que argumenta Mason (2008), Garretson (2008)²⁹ sustenta que programar é uma atividade conjunta entre programador e linguistas de corpus. Isso se dá porque Garretson vê a tarefa de programar como algo realizado em plataformas familiarizadas por cientistas da computação, não como Mason (2008) que considera o uso de ferramentas familiarizadas por usuários comuns, como o sistema operativo *Windows* (MS-DOS), *Mac* ou *Linux*. Entretanto, apesar de sustentar seu artigo na premissa de que o desenvolvimento de um *software* deve ser um trabalho conjunto entre linguista e programador, Garretson não descarta, inclusive aponta como uma situação ideal, o fato de o próprio linguista ser programador de suas próprias ferramentas:

Acredito que linguistas deveriam ser encorajados a aprender técnicas de programação (cf. Biber et al. 1998, p.254-256) e posso testemunhar que é possível que *software* linguísticos sejam desenvolvidos por indivíduos que não tenham um *background* em ciências da computação (p.75).

Garretson (2008) inicia o artigo resgatando alguns trabalhos de linguistas de corpus na descrição de *software* de análise de corpus, aplicados em estudos específicos. Toma como *linguistas de corpus* “todos que usam corpus eletrônico ou base de dados para pesquisa ou para ensino” (idem, p.69), isto incluiria, desta forma, pesquisadores em ETBC. Garretson afirma que a bibliografia nesta área é ainda escassa, “relativamente poucos autores discutiram requerimentos específicos dos tipos de *software* considerados neste artigo” (idem, p.69). O primeiro trabalho mencionado é o escrito por (ALEXA e ZUELL, 2002), que “discutem a dicotomia entre pacotes de *software* quantitativos e qualitativos” (idem). Em seguida, apresenta o trabalho de (WIECHMANN e FUHS, 2006), que revisa dez programas de concordância disponíveis em 2006. Já (SMITH, HOFFMANN e RAYSON, 2008) enfocam na discussão da falta de “boas ferramentas para listas de anotação de linhas de concordância. Dentro das considerações de arquitetura, (DAVES, 2005) “discute outro aspecto do

²⁹ Outro livro citado nesta tese, que trata de assunto semelhante ao que trata Garretson (2008) é Barnbrook (1996).

desenho de *software* linguístico, chamado de uso de base de dados relacional” (idem, p.70). Leech (2005), em um guia de prática de anotação de corpus, oferece várias observações relevantes ao desenho de corpus, em relação a anotação (idem, p.71).

Após essa breve revisão bibliográfica, Garretson expõe cinco questões que um pesquisador em estudos de corpus deve considerar antes de iniciar a pesquisa, em relação a (i) os requerimentos de *software*; (ii) o formato dos dados; (iii) o uso de *software* existentes; (iv) alguns problemas de *softwares* prontos; e (v) a criação de um *software* específico para a pesquisa. Quanto ao último item, deve-se considerar como adquirir a linguagem de programação ou a ajuda de um programador, gratuitamente através de acadêmicos ou estudantes voluntários que saibam programação, através de colaborações de pessoal da área de ciências da computação ou de técnicos de informática da universidade, ou através de acordos e projetos conjuntos de pesquisas com áreas da linguagem natural ou linguística computacional. Ou então, considerar a contratação de um programador, ou ainda investir que alguém envolvido no projeto comece a aprender a programar; a primeira opção pode ser muito cara e a segunda demorada (idem, p.74-75).

A partir daí ele descreve uma desiderata para desenhar *software*. Ele chama de “desiderata” os objetivos e princípios a se considerar quando se trabalha em um projeto de *software* linguístico. Evitando-se detalhes e discussões técnicas, estas desideratas estão exibidas em grupos temáticos: i) princípios de desenho geral; ii) teoria linguística e dados linguísticos; iii) a interface de usuário; iv) documentação; v) teste, depuração e tratamento de erro; vi) capacidades de busca e vii) performance. Aqui serão destacadas apenas as que dialogam de alguma forma com esta tese.

Para os princípios de **desenho geral de um *software***, destaca-se aqui sua menção a:

- planejar o tempo em relação ao desenvolvimento de um *software*;
- aproveitar a que já existe;
- Tornar o programa flexível;
- Buscar acomodar o programa a uma grande gama de usuários;
- Tornar o programa mais claro que engenhoso.

Quanto à **teoria linguística e dados linguísticos**, destacam-se as sugestões de:

- Separar os *markups* das anotações. Chama-se de *markup* “as informações extras adicionadas durante a criação do texto, que normalmente seriam perdidas” (idem, p.78) e de anotação “o registro de uma análise teórica dos dados” (idem). Aconselha-se inclusive que os *markup* sejam anotados em documentos diferentes das anotações.
- Permitir que os usuários forneçam suas próprias categorias, não limitando a análise desnecessariamente.

Quanto à **interface de usuário**, destacam-se as sugestões de:

- Deixar o desenho da interface o menos técnica possível, evitando deixá-lo confuso;
- Oferecer um conjunto de opções, evitando que o usuário tenha que saber o que digitar;
- Usar um usuário gráfico, mais do que uma interface de linha de comando. “Nada é mais intimidante a um não programador que uma interface no estilo de terminal: uma tela negra com um cursor piscando. (...). Uma interface gráfica (i.e. usando o windows e um *mouse*), não é apenas mais fácil de usar, como também é mais familiar à maioria dos usuários” (idem, p.80).
- Apresentar os resultados em estágios.
- Tornar fácil de instalar e fazer *updates* no programa.

Quanto à **documentação da programação**, destacam-se as sugestões de:

- Documentar assim que se começa a programar;
- Preferir por oferecer informação de mais, ao invés de informação de menos;
- Documentar de modo que o leitor compreenda;
- Documentar o código para que outros programadores tenham acesso;

Quanto ao **teste, depuração e tratamento de erro**. Garretson enfatiza que todo programa tem que ser testado e a maioria dos *bugs* serem removidos antes de oferecer o programa aos usuários. Destacam-se as sugestões de:

- Testar o programa a medida que o desenvolve;
- Utilizar “beta-usuários”, ou seja, oferece o programa a um grupo de usuários para testá-lo;
- Relatar erros apropriadamente;

Quanto à **capacidade de busca de padrões em dados textuais**, destacam-se as sugestões de:

- Equilibrar complexidade (muitas opções de busca) e velocidade na busca. “Tal equilíbrio deve ser dado de acordo com a função do programa” (idem, p.86).
- Usar *unicode* em expressões gerais, ou seja, no caso de representação de letras maiúsculas, optar por “\p{Lu}”, que serve para todas as línguas, do que o estilo antigo “[A-Z]”, que serve apenas para o inglês.

Apesar de Garretson discutir o desenvolvimento de *software* sofisticado, que se insere em uma linguagem de programação profissional, a linguagem de programação tem suas semelhanças, sejam elas para programas simples ou complexos. O planejamento previsto antes de se executar uma pesquisa de análise linguística e algumas desideratas de elaboração de *software* são válidos nos propósitos de pesquisa desta tese. Alguns tópicos aqui citados serão retomados nos próximos capítulos.

2.1.3.4. Reflexões sobre Mason (2008), Scott (2008) e Garretson (2008)

A interface entre os ET, a LSF e a LC, denominada nesta tese de ESFTBC, torna-se tradição de pesquisa em 2003 a partir do projeto CORDIAL. Nessa tradição, a metodologia de corpus, que concebe o corpus como o objeto de estudo e o computador como o meio de pesquisa, veio sendo aplicada em análise e descrição de textos em relação tradutória (cf. Baker, 1995 e Olohan, 2004). Os conceitos teórico-metodológicos dos ETBC, provenientes da LC, inserem-se nos ESFTBC. Segundo Kennedy (1998), o trabalho na LC é associado a diversas atividades, entre elas a de compilar e elaborar corpus; descrever elementos linguísticos/textuais através de corpora computadorizados; realizar descrição linguística em corpus para aplicar à aprendizagem e ensino de línguas e ao processamento de linguagem natural por máquinas; e desenvolver ferramentas para analisar corpora. Essa última atividade, em que se enquadra a presente pesquisa, é descrita por Mason (2008), Scott (2008) e Garretson (2008).

Em muitos casos, por falta de opção no mercado ou desconhecimento dos pesquisadores em relação aos programas disponíveis (Mason, 2008), ou por desconhecimento das opções de

desenvolver ferramentas que respondam exatamente à necessidade da pesquisa (Garretson, 2008), pesquisadores desenham suas pesquisas para a ferramenta de análise textual, e não vice-versa³⁰ (cf. Rodrigues, 2005 e Fleuri, 2006).

A linguística de corpus, assim como qualquer outra área em estudos de corpus, por ser um campo essencialmente metodológico (Olohan, 2004; Kennedy, 1998), permite que o foco do estudo se direcione ao método e suas ferramentas em si (como realizamos nesta tese), além da usual aplicação da metodologia na descrição e análise textual (como ocorre nas vinte dissertações analisadas nesta tese). Estudos sobre métodos e ferramentas de análise linguística são ainda escassos na área da LC e dos ETBC, conforme apontou o encontro de linguistas de corpus em Georgetown University em 1993 (Kennedy, 1998). Tal escassez faz com que a programação se torne um estudo mais matemático (Mason, 2008) e voltado ou associado à linguística computacional (Kennedy, 1998).

Entretanto, Mason (2008), Scott (2008) e Garretson (2008) mostram que examinar o funcionamento de ferramentas e de suas aplicações aos métodos de corpus não se tem que ser tarefa exclusiva do cientista computacional. De fato, não é necessário que o pesquisador seja um especialista em linguagem de programação e em desenvolvimento de ferramentas para análise textual para realizar estudos nessa área: qualquer linguista de corpus pode perfeitamente desenvolver esta tarefa, como prova Scott (2008). Segundo Garretson (2008), pensar a ferramenta e a possibilidade de programar deve ser uma das considerações a se considerar na hora de se planejar uma pesquisa em corpus.

Afinal, não há realmente como evitar a entrada e ou a consolidação dessa atividade de pesquisa nos estudos de corpus, como prevê Meyer (2002, p.141)—no futuro todos os linguistas de corpus serão um pouco programadores e seus *scripts* serão passados adiante de modo que um dia possam apoiar suas pesquisas em programas escritos e adaptados pelos próprios pesquisadores³¹. Vale destacar, entretanto, que

³⁰ Foi o que também ocorreu em minha dissertação de mestrado cujo corpus teve de ser reduzido e as unidades de análise limitadas por conta do longo tempo que se tomava em compilar e anotar o corpus.

³¹ “Muito provavelmente a próxima geração de linguista de corpus vai ter um *background* de programação muito melhor. Assim, estes linguistas de corpus serão capazes de aplicar o conhecimento da linguagem *Perl* ou *Visual Basic* para escrever *scripts* específicos para a análise de textos, e a medida que estes *scripts* vão se proliferando, poderão ser passados para frente e talvez tornar

apesar de defender esta ideia e de dialogar, até um certo ponto, com a linguagem de programação, esta tese não lida com a linguagem complexa de programação, mas sim oferece alternativas de programação básica em *VBA* e *Macros* para desenvolver *Templates* em um Processador de Texto e um Aplicativo de Planilhas, flexíveis e passíveis de serem reprogramados e adaptados. Esses *Templates* se apresentam como uma forma diferente e alternativa de compilação de corpus, que permite o pesquisador manipular organizadamente seu corpus (pois todos os documentos se encontram concentrados em um único programa) durante sua pesquisa. Todos os dados e corpus guardados em um único documento permite que seja repassados a futuros pesquisadores de modo que possam remanipular os dados e corpus já compilado de acordo com seus objetivos de pesquisa. Estes *Templates* visam a agregar rapidez e organização a não apenas pesquisas em ESFTBC, como também a pesquisas em ETBC.

2.2. METODOLOGIA DE CORPUS EM ETBC: PROJETO DE PESQUISA E PROGRAMAS PARA PROCESSAMENTO DE CORPUS

Com a localização teórica da presente pesquisa explicitada na seção anterior, passa-se à discussão sobre questões que envolvem o planejamento e a execução de pesquisas em metodologia de corpus. Nesta seção, são expostas questões debatidas por (BARNBROOK, 1996) sobre pontos metodológicos a se considerar na fase do projeto de uma pesquisa baseada em corpus. A Metodologia de Corpus requer a definição do desenho do corpus, a organização do processo de sua compilação, o planejamento do método de anotação e processamento do corpus e muita atenção, na redação da pesquisa, ao modo de exposição e à viabilização dos dados. Autores como Baker (1995), Olohan (2004), Fernandes (2006), Anderman & Rogers (2008), Borja, (2008), Varantola (2002), Kennedy (1998), Feitosa (2005), Leech (2005) entre outros trazem questões sobre desenho, compilação, anotação e processamento de corpus em estudos linguísticos voltados à tradução.

Segundo Barnbrook (1996), um projeto de pesquisa em corpus deve planejar o programa a ser adotado. Os programas citados nesta tese recebem três denominações: ‘programas prontos’ (*readly available*), ‘programas encomendados’ (*custom-built program*) e, como se denomina

obsoleta a necessidade de desenhar programas específicos de análise de textos” (Meyer apud Scott, 2008, p.103).

nesta tese, ‘programas adaptáveis’. Masons (2008), Garretson (2008), Barnbrook (1996) e Bowker (2004) usam os seguintes termos para designar *software* voltados a pesquisas de análise linguística de acordo com sua disponibilização:

- (i) “Programas prontos” (*readily available tools/program, ready-made software; existing software*) são *software* desenvolvidos especificamente para lidar com corpus, conforme Scott (2008). Os exemplos mencionados nesta tese, e que são os mais referidos em pesquisas em corpus, são o *WordSmith Tools* (Scott, 2011) e a opção gratuita o *AntConc* (Anthony, 2011);
- (ii) “Programas encomendados” (*custom-built program/tools*) são *software* a serem desenvolvidos para cumprir o propósito de uma determinada pesquisa. Em geral requerem um orçamento para seu desenvolvimento e a associação do pesquisador a um programador;
- (iii) Há ainda o que Bowker (2004) menciona de programas existentes não específicos para tradução (*existing non-translation specific tools*), que aqui chamaremos de ‘programas adaptáveis’. Os programas adaptáveis, ou que participam de uma esfera de “consumidor final”, como explica Sardinha (2004)³², são, no caso desta tese, os programas de processadores de textos (i.e. *MS Word*) e de aplicativos de planilha (i.e. *MS Excel*). Além de oferecerem ferramentas que automatizam a edição e oferecem informações quantitativas sobre determinados elementos, como: contadores de palavras, “corretores ortográficos, resumidores (por exemplo, a função ‘Auto Resumo’ do *Microsoft Word*), sintetizadores de voz, tradutores e digitadores (Via Voice, por exemplo)” (idem, *ibidem*), oferecem, segundo Rasinger (2008)³³, a possibilidade de se analisar dados estatísticos explanatórios e descritivos, como é o caso do *Microsoft Excel*, que possuem funções de programação básica, chamada de *Macros*.

³² Fonte:

http://www.revel.inf.br/files/entrevistas/revel_3_entrevista_tony_berber_sardinha.pdf

³³ Fonte: <http://linguistlist.org/pubs/reviews/get-review.cfm?SubID=2230895>

De acordo com Barnbrook (1996, p.165), “as qualidades mais importantes, as quais todos os que utilizam o computador para explorar e analisar línguas necessitam, são imaginação, flexibilidade e, claro, paciência”. Nada melhor que as “ferramentas já existentes e específicas a não-tradutores” (Bowker, 2004, p.23), como o caso das ferramentas aqui utilizadas do *MS Word* e *MS Excel*, para explorar a imaginação do pesquisador de corpus. Para Bowker (2004), o programa encomendado é uma alternativa no caso da inexistência de um programa pronto e o programa adaptável é uma alternativa para se continuar a trabalhar enquanto o programa encomendado não está pronto. Esta tese mostra que os programas adaptáveis servem não somente como uma alternativa de espera, mas também como o primeiro programa cogitado. Para um pesquisador em ETBC chegar à conclusão de qual programa utilizar em sua pesquisa precisa antes projetar, considerando alguns elementos-chaves. Nesta seção são explanadas questões sobre o projeto de pesquisa e o funcionamento dos programas prontos e adaptáveis.

2.2.1. O projeto de pesquisa em ETBC

Barnbrook (1996, p.5-20) argumenta que o desenho, a compilação e o processamento de um corpus se respaldam no projeto e objetivos. Para ilustrar essa relação ele usa como exemplo sua própria pesquisa e as etapas de projeto pelas quais passou. Barnbrook aponta que para desenvolver uma pesquisa em corpus, o pesquisador tem que pensar em uma série de questões pragmáticas relacionadas ao método e ao material escolhido.

Antes de cogitar a utilização de corpus eletrônico ou a condução manual de análise linguística em um projeto de pesquisa, “há que se considerar em detalhes todos seus convenientes e inconvenientes” desta metodologia (idem, p.11). As vantagens de adotar a metodologia de corpus³⁴, segundo Barnbrook (idem) são: “(a) a velocidade, a precisão e consistência de processamento; (b) a capacidade de realizar vários processamentos dos resultados; (c) a facilidade com a qual os dados podem ser manipulados, selecionados, ordenados e formatados; (d) a ausência de qualquer tendenciosidade humana” (idem, p.11-12). As desvantagens são: “(a) restrição de corpora eletrônicos e trabalho extra em sua preparação para que possam ser processados; (b) modificações no

³⁴ Barnbrook (1996) se refere à metodologia de corpus como “o uso de computador em pesquisas linguísticas” e seus exemplos são todos voltados ao método de corpus.

corpus em função do processo de análise requisitado, devido à natureza das operações dos programas; (c) trabalho extra envolvido no desenvolvimento e teste de programas; (d) a falta de um conhecimento dos programas e ferramentas computacionais” (idem, p.12). O pesquisador é quem julga o que mais lhe convém considerando os objetivos e medindo o tempo de pesquisa em relação aos recursos disponíveis. Em pesquisas de corpora de pequena dimensão, por exemplo, uma das vantagens é a replicabilidade e a possibilidade de executar testes com os dados da várias e diferentes formas e uma desvantagem é que em geral o tempo gasto na compilação do corpus e na aprendizagem do programa a processá-lo pode ser praticamente o mesmo que o tempo gasto em uma análise manual deste corpus.

Barnbrook (1996) aconselha que se realize uma análise em uma pequena amostra do corpus que se pretende analisar, testando o corpus escolhido, o modo de compilação que deverá ser aplicado e o programa para seu processamento. Este teste visa a mostrar: (i) se os dados obtidos podem responder às perguntas de pesquisa, verificando a relevância da escolha do corpus; (ii) se os programas utilizados geram os dados necessários, verificando se as ferramentas disponíveis cumprem com os objetivos da pesquisa e se é necessário criar outros programas ou adequar os objetivos aos programas; e (iii) se o corpus necessita muita alteração para poder ser processado. Esses elementos verificam a viabilidade da pesquisa em termos de recursos, tempo e custo aplicados durante as etapas de compilação, processo e análise.

Determinar o desenho do corpus ajuda a definir o método de análise a se adotar, em relação ao programa ou a forma de processamento e de análise (mais manual ou mais automática?). Barnbrook (1996, p.19) aconselha considerar todas as alternativas antes de se comprometer com um método.

2.2.1.1. Desenho do corpus

Ao se pensar um projeto de pesquisa em corpus, o pesquisador deve se assegurar que o corpus selecionado responda aos objetivos da pesquisa (Barnbrook, 1996, p.24). O desenho de um corpus, que está relacionado ao modo de apresentação deste conforme tais objetivos, refere-se, segundo Baker (1995, p.229) ao tamanho, a representatividade, o local (i.e. Austrália ou EUA), o tempo (i.e. Antigo ou contemporâneo), o gênero (i.e. Jornal, ficção, acadêmico etc.), o tipo, entre outros fatores que são vistos a seguir. A diferença de corpus entre a LC e os ETBC é que o corpus utilizado na LC é monolíngue e não inclui traduções, mesmo

que sejam para a língua em questão (idem, p.230); já nos ETBC, traduções devem fazer parte do corpus estudado (idem, ibidem) ou no propósito da pesquisa.

Algumas discussões típicas sobre o desenho de corpus em Linguística de Corpus são também recorrentes nos ETBC. Na presente subseção são discutidas algumas questões sobre o desenho do corpus que: (i) se assemelham entre a LC e os ETBC, como as considerações em relação ao tamanho e o ‘equilíbrio’ do corpus e (ii) que cabem exclusivamente aos ETBC, como as considerações em relações a um corpus ser paralelo ou comparável (e multilíngues). Barnbrook (1996) e Kennedy (1998) concentram-se no desenho do corpus para pesquisas em linguística, enquanto Borja (2008), Varantola (2002) e Fernandes (2006) —com base em Baker (1995), (ZANETTIN, 2000) e Olohan (2004), descrevem o desenho de corpus em pesquisas nos ETBC.

Na visão de um linguista, a questão sobre o tamanho³⁵ do corpus está relacionada à questão de representatividade, que se conecta também ao objetivo/propósito da pesquisa (Barnbrook, 1996, p.24). Um corpus é considerado suficientemente grande se é representativo à proposta de sua pesquisa. Segundo Barnbrook (idem), uma amostra do corpus é capaz de revelar, não só a seleção dos métodos, como também o perfil necessário, isto é, o desenho do corpus. Kennedy (1998, p.60) concorda com a relação entre o tamanho do corpus e a representatividade e entre representatividade e objetivo da pesquisa. Afirma que questões relacionadas à representatividade “são essencialmente relacionadas com a qualidade de um corpus” (idem, p.66), e que o tamanho tem muito a ver, também, com a quantidade de textos:

Essas questões não se relacionam apenas com o total de número de palavras (tokens) e palavras diferentes (types) em um corpus, mas com quantas categorias o corpus deveria conter, quantas amostras o corpus deveria conter em cada categoria, e quantas palavras deveriam haver em cada amostra (idem, ibidem).

De acordo com a definição pautada na quantidade de textos, são considerados corpora de grande dimensão o *British National Corpus*, o

³⁵ A extensão do corpus conta no planejamento da pesquisa e do tipo de programa adotar: um corpus muito pequeno permite cogitar uma análise manual, enquanto um corpus muito grande não oferece outra escolha a não ser adotar o modo automático de análise.

The Canadian Harsards, o *The German-English Parallel Corpus of Literary Texts*, o *LacioWeb*³⁶, entre outros.

Kennedy (1998) acrescenta o elemento ‘equilíbrio’ (*balance*) ao desenho do corpus, definindo-o como “o peso entre secções diferentes em um corpus geral” (idem, p.62). A questão sobre o ‘equilíbrio’ é levantada também em corpora desenhados para representar não apenas a língua como um todo, mas um específico domínio, gênero, tópico ou assunto. Além de tamanho, representatividade e equilíbrio, Kennedy adiciona dois tipos de corpus relacionados ao desenho: o estático e o dinâmico: “Um corpus pode ser uma coleção estática de textos selecionados a partir de um certo princípio, intencionado para ser típico de toda língua ou um aspecto da língua em um determinado tempo” (p.60), enquanto um “corpus dinâmico ou monitor é mais análogo a uma figura em movimento que a um *snapshot*, é assim chamado por oferecer meios de monitorar padrões em mudança de uso no tempo” (p.61).

Nos ETBC, Baker (1995) inicia a discussão sobre o desenho de corpus em pesquisas voltadas essencialmente aos Estudos da Tradução.

³⁶ Segundo Sardinha (SARDINHA, 2000), que enfoca no campo da Linguística de Corpus, “Na língua portuguesa, há vários corpora eletrônicos de destaque, tais como o Corpus de Araraquara, o de São Carlos (NILC), o CRPC (Corpus de referência do português contemporâneo), o Banco de Português, o PORTEXT, o Tycho-Brahe (português histórico) e o Corpus Natura, para citar apenas alguns. A pesquisa com corpora eletrônicos no Brasil já vem de longa data. Biderman (1978, pp.265-266) cita o corpus do 'Frequency Dictionary of Portuguese Words' como um dos primeiros corpora eletrônicos de português. Esse corpus continha 500 mil palavras de português europeu, referentes a publicações de 1920 a 1940. O dicionário de frequências feito a partir dele foi concluído em 1972, mas permanece inédito (Duncan Jr, 1972). Biderman (1978, pp. 65-67) ainda menciona vários outros corpora pioneiros no Brasil, usados para pesquisas no campo da Estatística Léxica, destacando-se os compilados por Jean Roche (Universidade de Toulouse, França, na década de 1960), J. Hutchins (Academia Naval de Anápolis, EUA, anos 1970), Cléa Rameh (Universidade Stanford, EUA, 1972), além daquele compilado por ela mesma (Maria Teresa Biderman, USP, 1969) e de uma série de corpora de textos literários de autores brasileiros construídos e analisados por uma equipe do ITA (São José dos Campos). Castilho et al. (1995) oferecem um panorama dos projetos de criação e informatização de corpora em várias regiões do Brasil. O levantamento indicou que havia um interesse na criação de corpora por parte de vários grupos de pesquisa, embora o índice de informatização estivesse apenas pouco acima de 50%. Castilho et al. (1995) concluem pormenorizando o que seria o Banco de Dados da Língua Portuguesa, um corpus de língua escrita e falada, o qual não foi concretizado”.

Parece-lhe claro que a LC baseia-se somente em corpora monolíngues e que cabe aos ETBC refinar esses critérios de acordo com as necessidades de pesquisa desta área (idem, p.230):

“Textos traduzidos sempre tiveram uma relação incomoda com a linguística de corpus. Eles têm sido excluídos de corpora monolíngues, em que são geralmente tratados como não-representativos da língua estudada, independente da direção da tradução” (Baker, 1996, p.175).

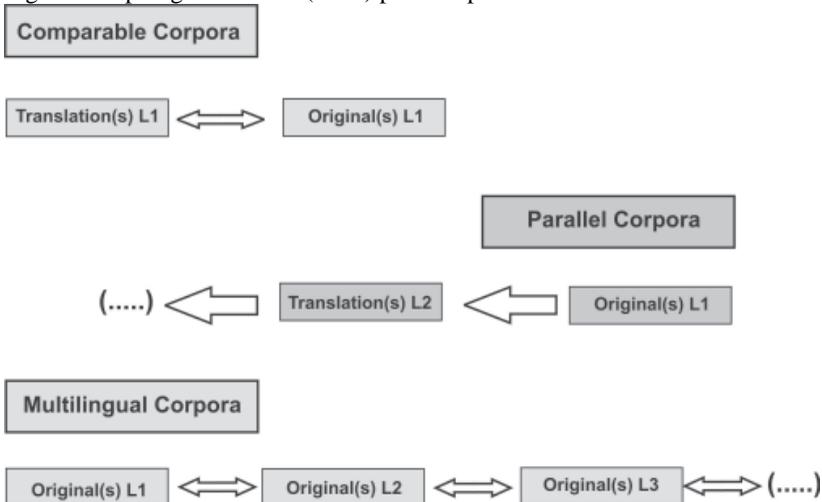
A partir deste ponto, Baker (1995, p.230) propõe três tipos de corpus para se começar a estabelecer os critérios e a discussão quanto ao desenho de corpus nos ETBC, são eles: corpora paralelos; corpora multilíngues e corpora comparáveis. O corpus paralelo é composto por “textos na língua-fonte A e suas versões traduzidas para a língua B. (...) Estes corpora permitem-nos, por exemplo, estabelecer objetivamente como os tradutores lidam na prática com dificuldades de tradução” (Baker, 1995, p.231). O corpus multilíngue

Se refere a um conjunto de dois ou mais corpora monolíngues em línguas diferentes, construídos nas mesmas ou diferentes instituições em base a critérios de desenho semelhantes. (...) Este corpus permite-nos estudar itens e características linguísticas nos seus próprios ambientes, em vez do modo como são usados em textos traduzidos. (...) Estes corpora podem prover acesso aos padrões naturais de uma determinada língua alvo (idem, p.232).

Já o corpus comparável, para Baker, é uma mistura entre o corpus paralelo e multilíngue: “Corpora comparáveis consistem de duas coleções separadas de textos na mesma língua: um corpus é composto por textos originais em uma determinada língua [A] e o outro por traduções nesta língua [A] de uma dada língua [B] ou línguas-fonte [B, C, D]” (idem, p.234).

Os tipos de corpus descritos por Baker (1995) são ilustrados por Fernandes (2006) da seguinte forma:

Figura 1: Tipologia de Baker (1995) para Corpora em ET



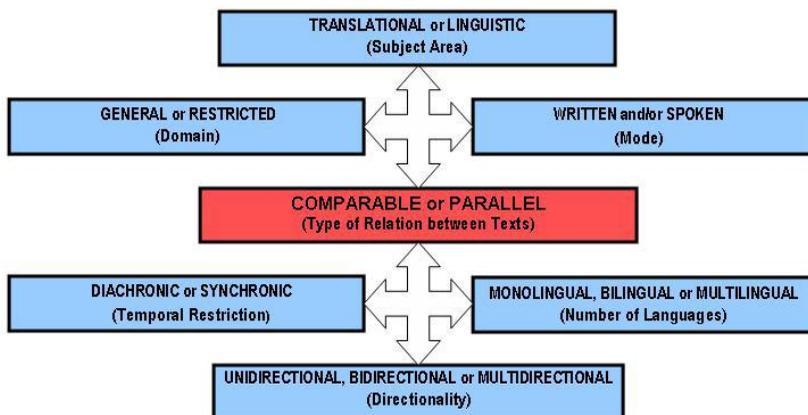
Fonte: Fernandes, 2006, p.90

Os corpora que Baker (1995) descreve representam apenas o início de uma longa discussão a este respeito, conforme ela mesma previu e como afirma Fernandes (2006). Fernandes (idem, p.89) propõe uma forma mais flexível de classificação de corpus em ETBC, fundamentando-se na proposta de Baker (1995) e em outros autores como Olohan (2004), (KENNY, 2001) e Zanettin (2000). Ele inicia sua proposta argumentando que os corpora multilíngues não têm sido utilizados conforme descritos por Baker (1995) e por isso sugere que sejam utilizados apenas os termos corpus paralelo e corpus comparável:

Tal classificação [corpus multilíngue] não parece ter entrado no campo, uma vez que o termo corpora comparáveis multilíngues tem sido geralmente usado para substituir corpora multilíngue (cf. Teubert, 1996 and Kenny, 2001). Além disso, vale notar que no primeiro livro introdutório dos ETBC, intitulado *Corpora in Translation Studies* escrito por Maeve Olohan (2004), a autora enfoca somente em corpora paralelos e comparáveis, o que pode indicar uma mudança de perspectiva no modo de classificação dos tipos de corpora. (Fernandes, 2006, p.91)

Seguindo a abordagem que os ET têm dado aos estudos de corpus, Fernandes (idem, ibidem) propõe uma classificação de corpora em ET mais flexível, ilustrada da seguinte forma:

Figura 2: Critérios de Classificação para Corpora in ET



Fonte: Fernandes, 2006, p. 91

Como se pode observar, a classificação gira em torno dos tipos de corpora “paralelos” e “comparáveis”. Esses tipos são diferenciados por seus contrastes: no corpus comparável “os textos são agrupados com base na semelhança *textual* (i.e. tópico, tipo de texto, funções comunicativas, etc.)” (idem, p.91), enquanto no corpus paralelo “os textos são agrupados com base na semelhança *tradutológica* (i.e. um texto é a tradução do outro)” (idem, ibidem). Conforme a Figura 2, cada um desses eixos podem se interagir com outras características relacionadas ao desenho de corpus, refinando-o. Seis tipos de possíveis relações são apontados: (i) a relação de área, (ii) de domínio, (iii) de modo, (iv) de restrição temporal, (v) de número de línguas envolvidas no corpus e (vi) de direcionalidade (idem, ibidem).

Na mesma linha de Fernandes, McEnery & Xiao (2008, p.19) excluem o termo “corpus multilíngue” e se baseiam nos dois eixos de corpus paralelo e comparável para classificar o desenho de um corpus em ETBC. Seguindo esta mesma linha de raciocínio, Borja (2008) afirma que a combinação de vários elementos binários permite gerar uma infinidade de formas de se determinar o desenho de um corpus: “Para criar um corpus teremos que decidir, entre outras, as seguintes dicotomias:

- Impresso/eletrônico

- Online/off-line
- Oral/escrito
- Todo texto/fragmentos de texto
- Estático/ dinâmico
- Sincrônico/diacrônico
- Grande/pequeno
- Geral/especializado
- Fontes literárias/fontes não-literárias
- Representativo (de uma língua especializada, variedade de língua, gênero etc.) / não-representativo (amostra de corpus)
- Organizado em torno de um gênero/ não organizado em torno de um gênero
- Monolíngue/multilíngue
- Paralelo multilíngue/ comparável multilíngue
- Multilíngue alinhado em paralelo/ multilíngue não-alinhado em paralelo
- Sem anotação/ com anotação
- Para uso profissional/ para uso em pesquisa
- Para pesquisa em geral/ para pesquisa específica
- Para uso profissional em tradução geral/ para uso profissional em tradução especializada” (p.246).

Borja (2008, p. 245-250) busca descrever um desenho de corpus ‘ideal’ para o tradutor. A autora resgata discussões de linguistas como (ATKINS, CLEAR e OSTLER, N, 1992) e (MCENERY e WILSON, A., 1996) que enfatizam que os objetivos da pesquisa em corpus influenciam o resultado de desenho deste. Observa-se que Borja (2008) considera os mesmos elementos que destaca Fernandes (2006), mas adiciona algumas possíveis combinações considerando também o ‘campo profissional da tradução’ em contraste com o ‘campo da pesquisa’. Menciona elementos tipicamente resgatados no desenho de corpora em linguística que servem igualmente aos ETBC, como a possibilidade de ser oral ou escrito, estático ou dinâmico e grande ou pequeno. Borja também menciona o estado primário do corpus, antes do processamento, como a possibilidade de estar impresso (e aqui eu adicionaria, “ou falado”) ou eletrônico, e online ou offline (como o corpus que se encontra em um CD).

A esta lista eu acrescentaria o elemento relacionado ao ‘modo de compilação’, por interferir no método: ‘corpus compilado’ seriam os corpora montados, já compilados como os grandes corpora como o *British National Corpus*, *Australian National Corpus* etc.; e ‘corpus não-

compilado’, ou seja, aqueles cujos textos encontram-se impressos ou gravados e são compilados pelo próprio pesquisador.

Krista Varantola (2002) relaciona o modo de compilação à funcionalidade do corpus segundo o propósito da pesquisa, podendo ser considerado ‘descartável’ ou ‘permanente’. A autora usa como exemplo de ‘corpus descartável’ aquele montado com um certo propósito—como o de pesquisa durante o ato de traduzir (em uma aula de tradução, por exemplo), e depois de usado é descartado ou reciclado:

Os corpora descartáveis (...) podem ser vistos como um tipo extremo na escala de corpora comparável. Tipicamente, o texto-fonte a ser traduzido é meramente um texto; um material na língua-fonte que forma o material de fundo para o texto que está para ser produzido na língua-alvo, ou um texto já na língua-alvo que precisa ser editado ou atualizado. O volume do corpus consiste de textos eletrônicos comparáveis na língua-alvo compilados para ser traduzido, editado ou para tarefas de produção de texto. A maioria da fonte disponível para esse tipo de material é, naturalmente, a Web, mas nada impede que o compilador utilize fontes mais permanentes como enciclopédias e publicações em CD-ROM. Em geral, o corpus resultante é descartável e não precisa ser tecnicamente elaborado. Sustentando ainda a metáfora, se surge a necessidade, este material descartável pode ser reciclado e refinado para formar parte de uma coleção mais permanente. (Varantola, 2002, p.174-175).

Olohan (2004), Fernandes (2006) e Borja (2008) apresentam variações de tipos de corpus, mas todos concordam que essas variações são relacionadas em ETBC a um corpus paralelo ou comparável. Para Fernandes cada um desses tipos de corpus pode se relacionar com outras seis categorias de desenho de corpus. Borja (2008), assim como Varantola (2002), aponta para outras categorias dicotômicas, além dessas seis. Olohan (idem), assim como Barnbrook (1996) e Kennedy (1998), debate também sobre a característica relativa ao tamanho do corpus. Questões sobre o tamanho do corpus são debatidas tanto nos ETBC quanto na LC. Segundo Barnbrook, Kennedy e Olohan, o tamanho do corpus pode ser definido conforme sua representatividade (dizendo a respeito da qualidade do corpus) e o número de textos, palavras e toques (dizendo respeito à quantidade).

A proposta metodológica desta tese abrange corpora eletrônicos bilíngues paralelos de pequena dimensão não-compilados voltados para uso em pesquisas. Esse desenho diz respeito ao formato do corpus e sua compatibilidade ao processamento no programa aplicado ao método.

2.2.1.2. Compilação do corpus

Para Olohan (2004) o corpus é compilado de acordo com seu desenho. A compilação de um corpus é provavelmente a etapa menos automatizada e que mais tempo toma em uma pesquisa em corpus. Na década de 1990 (cf. Barnbrook, 1996 e Kennedy, 1998) o corpus manual em contraste com o eletrônico era mencionado, apesar de já não mais cogitado; nos anos 2000 em diante (cf. Olohan, 2004 e Fernandes, 2006) considera-se que todos os corpora estão ou devem estar em formato eletrônico e são trabalhados somente através do computador. Por isso, esta é uma etapa obrigatória em toda pesquisa cujo corpus já não se encontre compilado. A compilação consiste em transformar o corpus em um formato eletrônico legível pelos programas de computador que o processarão. Esta seção se dedica a mostrar de modo geral as ações implicadas neste processo, segundo Olohan (2004, p.48-56) e de acordo com a análise das dissertações em ESFTBC nesta tese.

De acordo com Olohan (idem, ibidem), são seis ações a se tomar durante a etapa de compilação de um corpus ao formato eletrônico:

(i) **Fundação de um projeto de compilação de corpus:** esta etapa foi discutida por Barnbrook (1996) e Garretson (2008) e resgatada aqui por Olohan (2004). Por pressupor o uso de aparatos eletrônicos, a serem projetados, esta etapa envolve custos e conhecimentos básicos de tais tecnologias (ou auxílio técnico), (cf. Garretson, 2008). O tempo gasto também deve ser considerado, pois, dependendo do formato em que se apresentam os textos esta etapa pode se mostrar muito lenta, afetando o desenvolvimento da pesquisa. O projeto visa a constatar a viabilidade da pesquisa em termos de custo, apoio técnico e tempo aplicado:

nota-se que mesmo a construção de um corpus pequeno por pesquisadores individuais implica algum custo, dependendo das instalações já disponíveis (...); pode haver um custo adicional ao ter que comprar livros para serem escaneados ou permissão para serem escaneados e para comprar escâneres e programas de análise de corpus (p.49);

(ii) **Transformação dos textos em formato eletrônico:** “mesmo com o auxílio de um *software*, transformar textos impressos em formato eletrônico pode ser um processo bem tedioso” (Olohan, 2004, p.49). Kenny (2001 apud Olohan, 2004, p.50) relata que gastou uma hora para escanear 50 mil palavras e outra hora para corrigir 6 mil palavras. Este é portanto um processo muito lento, que demanda paciência e dedicação à tarefa manual por parte do pesquisador (ou assistente de pesquisa). Isso porque essa etapa requer muita correção “manual” do pesquisador, reformatação, revisão e edição dos textos. A quantidade de trabalho aplicado nesta etapa depende de dois elementos: do programa que processará o corpus e do formato inicial dos textos que irão compor o corpus (isso inclui o idioma em que se encontram os textos e a compatibilidade de leitura nesta língua pelo programa—o OCR de muitos escâneres interpretam melhor textos em inglês que em outras línguas). Assim que transformar os textos em formatos eletrônicos, há que prestar atenção no formato de documento em que estes textos serão salvados. Para isso há que verificar o programa que processará o corpus e o tipo de formatação textual este demanda. Por exemplo, ferramentas que requerem textos em formato de *Plain Text*, pedem que os textos estejam livres de qualquer formatação de fonte ou de linhas e parágrafos, além da eliminação de elementos extratextuais como gráficos, figuras etc. (idem, p.51). Então, antes de iniciar essa etapa é importante verificar o programa a ser usado e o formato textual requerido.

(iii) **Obtenção da permissão de direitos autorais:** em geral, se se deseja guardar uma cópia completa ou cópia de ‘parte substancial’ de um texto, é necessário obter permissão dos direitos autorais (idem, p.50). Cada país e companhia tem suas regras em relação à concessão dos direitos autorais a cópias do material. É importante que isso seja verificado antes de iniciar o processo de compilação e o autor (e tradutor) seja consultado. Geralmente, afirma Olohan (idem,, p.51), muitos autores e companhias não veem problemas em permitir cópias para somente fins de pesquisa, uma vez que seja especificado que esse material será utilizado apenas para tal fim e que outros pesquisadores terão acesso a partes deste texto, que não será divulgado publicamente na íntegra.

(iv) **Anotação do corpus:** se um texto deve ou não ser anotado “depende dos objetivos da pesquisa, do *software* de corpus utilizado e do quanto esse corpus será compartilhado com outros pesquisadores” (idem, p.51). A etiqueta (*tag*) tem que obedecer certos quesitos de leitura do programa que processará o corpus. Por exemplo, anotar texto com formatação de fonte, destacando o objeto anotado em *italico* ou **negrito** não é uma boa ideia se o programa lê textos em formato

RTF, que elimina qualquer tipo de formatação. Muitas etiquetas são nomes, siglas, ou códigos entre parênteses angulares < > ou traço baixo (*underscore symbols*). As etiquetas podem ser acrescentadas ao corpus manualmente pelo pesquisador ou automaticamente através de *tagging systems*, *tagsets* ou *lemmatizers*, que são programas de etiquetamento baseados em uma abordagem probabilística (idem, p.52). Segundo Kennedy (1998, p.218) esses programas apresentam uma margem de erro de 4%, o que é considerado bom, mas que não dispensa o trabalho manual de verificação de etiquetas.

(v) **Alinhamento do corpus paralelo:** esta etapa é necessária em caso de corpora paralelos alinhados³⁷. Segundo Olohan (2004, p.55) “é possível alinhar dois corpora automaticamente, mas o grau de sucesso desta operação depende das línguas envolvidas e do mecanismo de alinhamento utilizado”. Assim como o processo de escaneamento, o alinhamento automático também necessita ser verificado, já que o computador não é capaz de ler automaticamente se uma frase foi traduzida em duas ou vice-versa, ou outras forma de reestruturação textual durante o ato de tradução. Além disso, observou-se que certos programas, como o *WordSmith Tools* (Scott, 2011) lê letra maiúscula e ponto final como marcação de sentença (cf. Scott, 2008), isso faz com que se considerem abreviações como tais, ou seja, *Prof.*, por exemplo, uma sentença completa. Isso tem que ser arrumado antes de o alinhamento automático ser efetuado.

Depois de o corpus ser alinhado “um programa de concordância paralela pode ser usado para produzir instâncias de ocorrências de uma palavra ou estrutura no texto-fonte e na sua tradução equivalente, ou vice-versa” (Olohan, 2004, p.55). Dois exemplos de programas deste tipo, mencionados por Olohan são o *Multiconcord* e o *Paraconc*, que alinham automaticamente e realizam a concordância em corpora alinhados. Segundo (VINTAR, 2008) (p. 165) “se se consegue a licença de cópia do *Trados WinAlign* versão 6.5 ou mais alta, o alinhamento se torna uma tarefa fácil”, há alguns programas no mercado de memória de tradução, como o *ATRIL's DejaVu*, e de concordância em paralelo, como o *Paraconc*, que podem ser comprados por um preço razoável. Ainda assim, aponta o autor, o alinhamento de sentenças tende a ser um processo não completamente automático, pois as ferramentas processam os pares de

³⁷ Apesar de muitos autores insinuarem que o corpus paralelo é em geral exposto lado-a-lado, como um alinhamento (cf. Baker, 1995 e Fernandes, 2006), Borja (2008, p.245) menciona sem detalhar, o contraste no desenho de corpus entre ‘paralelo alinhado’ e ‘paralelo não-alinhado’.

sentença, mas a correção deste alinhamento é realizada manualmente pelo pesquisador.

(vi) **Coleta e codificação de outros dados:** além das anotações, que são marcas gramaticais dos textos que compõem o corpus, o pesquisador pode necessitar de marcações (*Markups*) que são marcas adicionais ao documento, que indicam a proveniência do texto, data de publicação, tipo-textual etc.

“Marcação (*Markup*) é o termo usado para se referir a estes tipos de informação adicional codificada em um documento que acompanha um texto. As marcações agem como um conjunto de instruções para que programas de computador exibam. (...) Se um corpus tem que ser compartilhado com outros pesquisadores ou disponibilizado na internet, é importante usar padrões e formatos reconhecidos para codificar os dados” (Olohan, 2004, p.56).

Em relação à tarefa de transformar os textos impressos ou falados em formato eletrônico, o que foi observado durante a minha pesquisa de mestrado Fleuri (2006) e durante a leitura e análise das vinte dissertações descritas nesta tese³⁸, e segundo Olohan (2004, p.49-50), é que o tempo de compilação aumenta conforme a apresentação da versão inicial dos textos. Textos impressos (ou falados) demandam mais tempo e trabalho de correção dos erros de escaneamento, que textos em PDF, e estes demandam mais que textos em formato *Word Document* (.doc) ou *Plain Text* (.txt).

Apresenta-se abaixo a ordem das etapas de compilação do corpus, considerando a versão dos textos, o tempo considerado para a compilação e os procedimentos em sequência. Esta ordem foi apresentada em 2011 no *I Seminário de Pesquisas em Tradução e Corpora*,

³⁸ O que se observou nas dissertações analisadas foi que 60% dos pesquisadores citaram alguma forma de trabalho e tempo de pesquisa gasto na compilação do corpus, especialmente no que se refere à tarefa de transformar o texto impresso em formato eletrônico: “Esta primeira etapa consistiu em digitalizar os romances através de “escaner” e salvá-los em formato “txt”. Muita atenção foi demandada na conferência da digitalização porque, durante o processo, ocorrem muitas alterações no corpus, dentre elas, omissão de palavras, orações e até mesmo parágrafos; apresentação dos parágrafos em formato de colunas; palavras com letras trocadas. A revisão do processo de digitalização foi fundamental e, para maior confiabilidade, uma releitura dos romances, em sua totalidade, foi realizada” (Cruz, 2003).

(Universidade Federal de Santa Catarina-Brasil) (Fleuri, 2011a) e no *International Conference on Translation and Cross-Cultural Communication*, na (Universidade de Queensland/Austrália) (Fleuri, 2011b).

➤ **Texto impresso³⁹: tempo longo**

1. Escanear,
2. Transformar em Doc ou Txt,
3. Corrigir erros de escaneamento (passar por um corretor ortográfico),
4. Preparar para ser alinhado (eliminar pontos de abreviação, quebrar texto em sentenças etc.),
5. Preparar para ser etiquetado (isolar os elementos a serem classificados e rotulados),
6. Etiquetar,
7. Revisar,
8. Coletar os dados (quantificar, analisar, comparar etc.).

➤ **Texto em PDF: tempo médio**

1. Transformar em Doc ou Txt,
2. Preparar para ser alinhado (eliminar pontos de abreviação, quebrar texto em sentenças etc.),
3. Preparar para ser etiquetado
4. Etiquetar,
5. Revisar,
6. Coletar os dados.

➤ **Texto em Doc ou Txt: tempo mínimo**

1. Corrigir (caso já não se apresente com os erros de leitura do escâner corrigidos),
2. Preparar para ser alinhado (eliminar pontos de abreviação, quebrar texto em sentenças etc.),
3. Preparar para ser etiquetado
4. Etiquetar,
5. Revisar,
6. Coletar os dados.

³⁹ No caso de compilação de textos falados o processo de escaneamento é substituído pela transcrição, e o tempo de compilação pode aumentar dependendo da qualidade da gravação e do escaner.

➤ **Texto em Doc ou Txt corrigido, alinhado e etiquetado: pronto.**

1. (Re)etiquetar, caso necessário
2. Coletar os dados

Quanto à etapa de preparação do texto para ser alinhado, há alguns pontos a se considerar. Para se alinhar dois textos no *Viewer and Aligner* do *WordSmith Tools*, selecionam-se dois textos em relação tradutória e os alinha ao nível da sentença ou do parágrafo, ou seja, no caso do alinhamento em sentenças o programa toma a sentença um do texto um (T1) e alinha acima da sentença um do texto dois (T2). Esse programa reconhece como ‘sentença’ todo fragmento textual delimitado por uma letra maiúscula e um ponto (. ? !). É portanto muito comum que estes programas interpretem abreviações (Sr. Paulo) como duas sentenças (Sr. e Paulo...), pois *Sr.* inicia com uma letra maiúscula e termina com um ponto, neste caso, o ponto final. Esta má interpretação gera erros de alinhamento em casos de haver abreviações nos textos alinhados. Portanto, se os pontos das abreviações não são apagados no texto gravados em *Plain Text*, antes de serem processados no *Viewer and Aligner*, certamente erros de alinhamento ocorrerão. Há outros elementos, além da presença de abreviações, que causam desalinhamento (cf. Morinaka e Alves, 2004 – em anexo), tais como: as reticências, as escolhas do tradutor de traduzir uma sentença em duas e algumas características de gênero e estilo textual.

Uma vez que o corpus esteja em formato eletrônico, corrigido, anotado e alinhado, com as permissões de direitos autorais consentidas, a etapa de compilação está concluída e que o corpus está pronto para ser processado, seus dados coletados e analisados.

2.2.1.3. Anotação de corpus

Apesar de a anotação fazer parte da etapa de compilação, encontrou-se pertinente abrir uma seção somente para esta etapa, visto que este é um ponto observado durante a análise das vinte dissertações e da proposta metodológica desta tese. Segundo Leech (2005), “anotação de corpus é a prática de adicionar informação interpretativa a um corpus” — um corpus anotado “representa uma fonte compartilhável para pesquisas em ciências humanas e sociais” (idem). Uma pesquisa que requer a anotação de elementos textuais para se obter a informação

desejada durante o processamento, precisa considerar um código de anotação. Este código pode ser criado pelo próprio pesquisador, satisfazendo as necessidades da sua própria pesquisa, ou ser adotado a partir de um código criado e padronizado, permitindo que outros pesquisadores processem este corpus:

“Algumas pessoas (...) preferem não anotar seu corpus: para eles, o corpus não- anotado é o corpus ‘puro’ o qual desejam investigar—o corpus sem adulteração com informação que é suspeita, possivelmente refletindo preferências ou até mesmo erros de um anotador. Para outros, anotar é um meio de torna o corpus mais útil—um enriquecimento do corpus em sua forma crua. A partir desta perspectiva, provavelmente a visão da maioria, adicionar uma anotação em um corpus é ‘adicionar valor’, que pode ser usado em pesquisas de indivíduos ou de equipes que realizam anotação, mas que também pode ser passado a outros que achem isso útil aos seus propósitos” (Leech, 2005).

Esta seção se dedica a descrever como a anotação tem sido tratada na tradição da LC, segundo Leech (2005), e no âmbito dos ESFTBC, de acordo com Feitosa (2005)—quando se descreve *Código de Rotulação Sistêmico-Funcional* (CROSF).

Os tipos de anotações e de programas desenvolvidos para anotar são questões muito exploradas na LC e ainda pouco discutidas nos ETBC. Olohan (2004), que pode ser reconhecida como pesquisadora de ETBC, menciona etiquetas comumente utilizadas por linguistas de corpus, como a usada pelo *British National Corpus* (BNC), que se baseia em *Part-of-speech tagging* (POS-tagging). Leech (2005)⁴⁰ levanta questões práticas a respeito de anotação de corpus em estudos linguísticos mencionando alguns tipos de anotação mais comuns e alguns programas de anotação automática. Além do *POS-tagging*, Leech (2005) lista mais seis tipos anotação, correspondendo a diferentes níveis de análise linguística de um corpus, tais como: (i) anotação fonética, “que adiciona informação sobre como uma palavra em um corpus oral foi pronunciada”; (ii) anotação semântica, “que adiciona informação sobre categorias semânticas de palavras”; (iii) anotação pragmática, “que adiciona informação sobre os tipos de ato do discurso (ou do diálogo) que ocorre em um diálogo falado”; (iv) anotação do discurso, “que adiciona informação sobre

⁴⁰ Fonte consultada em julho/2013 <http://www.ahds.ac.uk/guides/linguistic-corpora/chapter2.htm>

conectores anafóricos em um texto”; (v) anotação estilística, “que adiciona informação sobre apresentação discursiva e de pensamento” e (vi) anotação lexical, “que adiciona a identidade do lema de cada forma de palavra em um texto”.

Leech verifica algumas tentativas de padronização os sistemas de anotação em (a) *POS-tagging*, como o *Brown Family of Corpora*, usado pelo BNC; (b) anotação sintática, como o esquema de análise (*parsing scheme*) utilizado em SUSANNE Corpus (Sampson, 1995), ou o *Penn Treebank* (Marcu, Santori & Marcinkiewicz, 1993); (c) anotação prosódica, como o ToBI (*Tones and Break Indices*), “que vem com sua própria plataforma de processador de discurso”, o INTSINT (HIRST, 1991) e o TSM — *Tonetic Stress Marks* (KNOWLES, WILLIAMS e TAYLOR, 1996) e (d) anotação pragmática, como o DAMSL manual (*Dialog Act Markup in Several Layers*) (Allen & Core, 1997), o esquema TRAINS, VERBMOBIL, o *Edinburgh Map Task Corpus* e o SPAAC (Leech & Weisser, 2003).

Um corpus pode ser anotado manualmente ou automaticamente. Segundo Leech (2005)

Etiquetadores automáticos podem atingir 98% de anotações corretas. Entretanto, mesmo que essa porcentagem seja considerada boa, o etiquetamento automático é geralmente seguido por uma etapa de pós-edição, em que requer a interferência humana na correção dos erros do etiquetador automático, ou para resolver problemas de ambiguidade.

A anotação manual também pode ser propensa a erros, além de ser um processo muito lento. No caso de se optar por adotar um etiquetador automático Leech (idem) aponta que há alguns programas desenvolvidos para projetos grandes, como é o caso do *CLAWS tagger* e o *Template Tagger* que são usados pelo *Brown Family of Corpora* e o BNC⁴¹. Leech (idem) conclui que não importando o método utilizado na anotação,

Para obter um ótimo corpus etiquetado, é necessário passar por um trabalho manual, geralmente em larga escala. O corpus etiquetado automaticamente tem que ser pós-editado por uma equipe de seres humanos, que podem passar horas nisso. Se o resultado desse trabalho tornar o corpus

⁴¹ “Such programs or packages can be licensed for your own annotation work. (For CLAWS, see the UCREL website <http://www.comp.lancs.ac.uk/ucrel/>.)” (Leech, 2005)

mais útil, deve-se construir uma versão etiquetada do corpus, que pode então ser disponibilizada para quem queira usar a etiqueta como um trampolim para sua própria pesquisa (Leech, 2005).

Com a intenção de padronizar o sistema de anotação das pesquisas em ESFTBC do projeto CORDIAL, é desenvolvido, em 2005, por Marcos Feitosa, o *Código de Rotulação Sistêmico-Funcional* (CROSF), respaldado nas três Metafunções da Gramática Sistêmico-Funcional. A dissertação de Marcos Feitosa intitulada *Uma proposta de anotação de corpora paralelos com base na Linguística Sistêmico-Funcional*⁴² visa a propor uma anotação de corpora mais eficiente e menos propensa a erros de digitação, os quais interferem nos dados das pesquisas em ESFTBC datadas antes de sua criação. Desta forma, a proposta de Feitosa é oferecer “um modelo de anotação instrumentalizado através de um código numérico para a Rotulação de corpora com base na Gramática Sistêmico-Funcional de Halliday” (Feitosa, 2005). Esse código, denominado “CROSF” (Código de Rotulação Sistêmico-Funcional), foi criado por meio de sucessivos testes em um corpus paralelo bilíngue de pequena dimensão. Destes testes foram desenvolvidos quatorze protótipos, e posteriormente à data de publicação de sua pesquisa, o CROSF15 foi elaborado, representando o último deles. Feitosa (2005) testa o CROSF14 em um estudo de caso em relação à Metafunção Textual do corpus analisado. Tal estudo demonstrou que o código numérico “facilitou as buscas no corpus eletrônico, permitindo melhor visibilidade do texto anotado e buscas de diferentes combinações de escolhas temáticas” (idem).

O código é composto de 7 dígitos numéricos compreendidos entre parênteses angulares.

O código funciona de acordo com a disposição ab cdefg. Isso implica que as posições a e b estão relacionadas entre si e as posições cdefg estão interligadas de modo que g é compreendido por f, que está inserido em e, o qual, por sua vez, é dominado pela posição c, sendo que d também é compreendido por c, conforme será explicado a seguir. A posição a (primeiro dígito) indica se o elemento da oração é Tema (simples ou múltiplo),

⁴² Dissertação disponibilizada online, no site:
http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_acti on=&co_obra=83927

ou se é Rema ou ainda N-Rema. A posição b (segunda posição) indica qual a posição do elemento classificado em relação a Temas simples e múltiplos. Se for elíptico, o algarismo 0 é utilizado; se ele for o primeiro ou o único, utiliza-se o algarismo 1; caso ele seja o segundo, é utilizado o algarismo 2; e assim sucessivamente. Já a posição c refere-se a uma das três Metafunções descritas por Halliday: a Metafunção Ideacional corresponde ao algarismo 1; a Metafunção Interpessoal, ao 2; e a Metafunção Textual, ao 3. A posição d concerne ao fato de o elemento ser Marcado ou Não-Marcado, na Metafunção Ideacional, ou Modulação/Modalização na Metafunção Interpessoal, não tendo subcategoria alguma para esta posição na Metafunção Textual, a qual é, por isto, obrigatoriamente preenchida com o dígito 0 (zero). Por fim, as posições efg correspondem a uma análise mais específica do elemento na oração. Deste modo, caso se trate da Metafunção Ideacional, por exemplo, e o algarismo que ocupa a posição e seja 1, tem-se que o elemento Ideacional é um Participante (sem interpolação) da oração. Caso se deseje, todavia, uma análise ainda mais apurada, as posições fg cobrem este tipo de análise. Por exemplo, (...), caso se deseje saber qual é a função do Participante de uma oração envolvendo um Processo Material, pode-se utilizar estas duas posições para indicar se o Participante é, por exemplo, “Ator” ou “Meta” (Feitosa, 2005, p.61).

Fleuri (2006) adapta a tabela de Feitosa (2005) voltada para a descrição da Metafunção Ideacional. Esta tabela ilustra a explicação de Feitosa apenas citada:

Quadro 1: CROSF-14 Metafunção Ideacional

< a	b	C	d	e	f/ >
Tema/ Rema	Posição	Metafunção	Marcado ou não	Relativo à Metafunção (Ideacional)	Relativo Metafunção (Ideacional) Participante
1: tema simples 2: tema múltiplo 3: Rema Simples 4: Rema múltiplo 5: minor clause * 6: N- Rema	0: elíptico 1: primeiro 2: segundo 3: terceiro 4: quarto 5: quinto 6: sexto 7: sétimo 8: oitavo 9: nono	1- Ideacional 2- Interpessoal 3- Textual	1: não- marcado 2: marcado	1: participante sem interpolação 2: participante com interpolação 3: processo 4: processoparticipante 5: circunstância sem interpolação 6: circunstância com interpolação 7: atributivo Preposto 8: oração 9: estrutura Tematizada	11: ator 12: meta 13: receptor 14: cliente 21: experienciador 22: fenômeno 31: portador 32: atributo 33: identificado 34: identificador 35: característica 36: valor 41: dizente 42: receptor 43: verbiagem 44: alvo 45: locução 51: comportante 61: existente 70: ambiente 80: extensão Processo 10: material 20: mental 30: relacional 40: verbal 50: comportamental 60: existencial Circunstância 10: localização * 20: extensão * 30: modo * 40: causa * 50: contingência * 60: acompanhamento * 70: papel * 80: assunto * 90: ângulo* Oração 10: sem interpolação 20: com interpolação Estrutura Tematizada 10: predicado * 20: equativo * 30: preposto * 40: comentário *

Fonte: Feitosa, 2005

Muitas pesquisas baseadas em corpus têm se revelado, segundo Feitosa (2005), laboriosas não apenas em função das tarefas de digitalização dos textos e de compilação do corpus, como também em função do processo manual de anotação dos corpora. É o caso relatado por Feitosa em 2005 dos pesquisadores do NET (Núcleo de Estudos da Tradução) da FALE (Faculdade de Letras) da UFMG (Universidade Federal de Minas Gerais),

Foi implementada uma proposta de análise da organização temática em corpora paralelos. As primeiras análises realizadas pela Professora Dra. Adriana Pagano e pelos próprios alunos mostraram que, além de se tratar de um trabalho demorado, o processo de anotação dos corpora vinha

apresentando falhas devido a erros de digitação na hora da inserção de Rótulos. A anotação vinha sendo efetuada através de palavras, com o nome de cada categoria inserido por extenso em cada Rótulo. Esses Rótulos acabavam sendo longos e mais propensos a erros de digitação, o que interferia nos resultados fornecidos pelo programa de análise de corpora. Além disso, a ausência de um padrão de anotação entre os membros do NET dificultava um diálogo mais profícuo entre os pesquisadores. Como cada pesquisador estabelecia seus Rótulos, o acesso ao corpus anotado de um outro pesquisador tornava-se mais difícil, pois implicava a necessidade de se aprender como foi efetuada a Rotulação de seu corpus (Feitosa, 2005, p.14).

Muitas pesquisas vinculadas ao projeto CORDIALL adotaram o código para etiquetar o corpus de suas pesquisas, como veremos no capítulo 4 desta tese. Observa-se que todas as pesquisas que utilizaram o CROSF14 analisavam o perfil ideacional (Metafunção Ideacional) de entidades lexicais (personagens ou determinados nódulos) em corpus paralelo bilíngue de pequena dimensão, construído com a finalidade de servir aos propósitos de cada pesquisa. Muitas das pesquisas em ESFTBC que não utilizaram o CROSF lidaram com outras instâncias Sistêmico-Funcionais (i.e. Alves, 2007; Araújo, 2007; Figueredo, 2007; Filgueiras, 2007; Nunes, 2009; De Souza, 2010; Rodrigues, 2010; Assis, 2009; e Jesus, 2008) cuja descrição das funções o código não cobria. Já outras investigações acabaram por adotar códigos criados pelos próprios pesquisadores, ainda que lidassem com a análise de alguma Metafunção como é o caso de Rodrigues Junior (2006); Zanella (2006); e Baldissera (2010).

O CROSF, apesar de não responder a todas as necessidades de etiquetamento de muitas pesquisas em ESFTBC, revelou-se uma opção de padronização de anotação em Metafunções Sistêmico-Funcionais, permitindo que corpora criados para pesquisas em ESFTBC sejam armazenados e reutilizados, cumprindo com a utilidade, mencionada por Leech (2005), que a anotação em corpus apresenta: a de poder “ser disponibilizada para quem queira usar a etiqueta como um trampolim para sua própria pesquisa” (idem, *ibidem*).

2.2.1.4. Processamento dos dados

Processar um corpus significa passá-lo por um programa computacional e obter os dados requeridos na pesquisa, por exemplo, obter número de palavras, palavras-chaves, uma lista de palavras, concordância de termos, a colocação, etc. O processamento é a etapa mais automatizada da pesquisa em corpus. Através dele se obtêm os valores quantitativos dos dados ou torna-se possível isolar os termos linguísticos a serem analisados. Em geral o processamento se dá através de ferramentas dos programas de análise lexical. No planejamento do projeto da pesquisa em corpus, deve-se considerar a informação que se necessita obter dos dados para se pensar as ferramentas computacionais que permitem extrair tais dados e os programas que possuem tais ferramentas. Esta etapa passa a ser mais detalhada quando se descrevem os programas prontos e adaptáveis. Para fins de ilustração, cito aqui algumas pesquisas cuja discussão centra no processamento de corpora em estudos da tradução encontradas em Anderman & Rogers (2008), tais como: (VÁRADI, 2008), (SIMOV, SIMOV, *et al.*, 2004), (AHMAD, 2008), (PIOTROWSKI, 2008) e (RAMBOUSEK e CHAMONIKOLASOVÁ, 2008).

2.2.1.5. Exposição dos dados quantitativos

A forma de exposição dos dados e da análise na redação da pesquisa não é muito explorada na bibliografia de ETBC. Entretanto, observa-se que é uma questão importante, pois esta é a única forma de o leitor e outros pesquisadores possuem para acessar a informação processada a que se referem os resultados. Conforme o exemplo mencionado no Capítulo 1 desta tese, a forma de exposição dos dados das pesquisas de Zuniga (2006), Fleuri (2005) e Geysel (2009) dificultou a coleta e a reutilização de tais dados no que seria inicialmente a presente tese. A disponibilização dos detalhes de seus dados é importantes para a constatação da validação da pesquisa e para permitir prosseguimento desta em futuras investigações.

Observa-se que pesquisas em ETBC (SERPOLLET, 2008) e (SANTOS, 2008) tendem a expor os dados quantitativos uma tabelas: em geral a primeira coluna é dedicada ao valor numérico, com o valor de cada item analisado exposto em cada linha e na última, o valor total; na seguinte coluna é exposto o valor percentual, na mesma distribuição. Os dados referentes ao *corpus A* é exposto lado-a-lado ao *corpus B*, principalmente em casos de corpora paralelos. Portanto, os valores numérico e percentual são explicitados, assim como os valores parcial e total. Em geral, um gráfico segue a tabela, facilitando a visualização para

o leitor. O que se observa é que a exposição de uma tabela nessa estrutura é mais importante que o gráfico, sendo este um artifício ilustrativo opcional, assim como o valor numérico tende a ser mais explicativo que o valor percentual.

2.2.2. Programas prontos

Nesta seção descrevo dois programas: o *WordSmith Tools* (Scott, 2011) e o *AntConc* (Anthony, 2011). Selecionamos o *WordSmith Tools* porque é o mais mencionado nas dissertações analisadas nesta tese e o *AntConc* por ser um programa gratuito que apresenta ferramentas com funções semelhantes ao do *WordSmith Tools*. A explicação é resumida e centrada nas ferramentas que foram mais requisitadas nas dissertações analisadas.

2.2.2.1. *WordSmith Tools* (Scott, 2010)

Nesta seção é fornecida uma explicação do *WordSmith Tools* (doravante *WST*) quanto a seu funcionamento e sua aplicabilidade. O site oficial do Mike Scott⁴³ oferece um detalhamento do programa e a possibilidade de baixar um demonstrativo da versão 5 do *WST*.

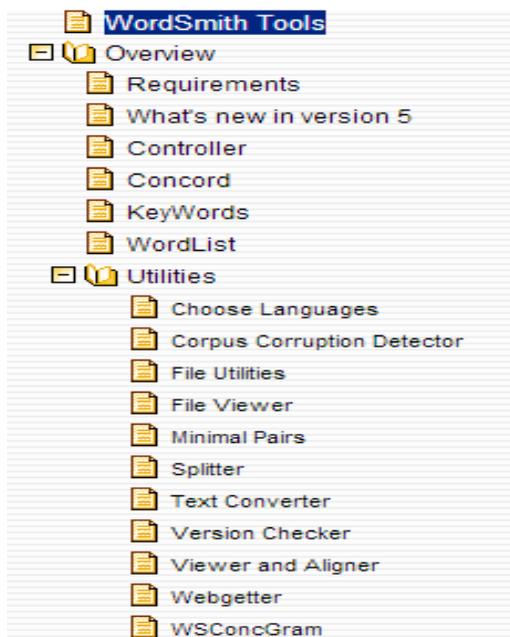
O *WordSmith Tools* é um programa de análise lexical, ou seja, “é um conjunto integrado de ferramentas que observam como as palavras se comportam em textos”⁴⁴. Ele existe no mercado desde 1996 e em 2013 já se encontra em sua sexta versão. É um *software* pago, com o custo mais baixo de GBP 50 mais taxas, cerca de R\$ 174⁴⁵, para a obtenção da licença a um único usuário. O site oficial ilustra em detalhes o funcionamento do programa e permite que se baixe gratuitamente um demonstrativo antes de que a compra seja efetuada. As ferramentas e as utilidades do *WordSmith Tools* versão 5, que iniciou em Junho de 2007, três anos depois do lançamento da versão 4, são as seguintes:

Figura 3: Ferramentas e utilidades do *WordSmith Tools* (Scott, 2011)

⁴³ Explicação detalhada deste programa se encontra no site oficial do WST: <http://www.lexically.net/downloads/version5/HTML/> e <http://www.lexically.net/wordsmith/index.html>

⁴⁴ Fonte: <http://www.lexically.net/downloads/version6/HTML/?wshell.htm> (acessado: 27 de junho de 2012).

⁴⁵ Valor e conversão verificados em julho/2013.



Fonte: Scott, 2011.

O programa contém quatro ferramentas⁴⁶e onze utilidades. As ferramentas mais utilizadas nas 20 dissertações analisadas são (i) *Concordância* (*Concord*); (ii) Palavras-chave (*Keywords*) e (iii) Lista de Palavras (*WordList*), além da utilidade do Visualizador e Alinhador (*Viewer and Aligner*). Para processar os corpora, o programa demanda três quesitos: um computador razoavelmente atualizado; a instalação do *Windows 2000* ou outra versão mais recente e uma coleção de textos sem formatação (em formato ou convertido ao formato *Plain Text*).

A ferramenta de concordância, o *Concord*, possibilita observar qualquer palavra ou frase em seu contexto – para que se possa ver o que a acompanha” (SCOTT, 2010b). A ferramenta *WordList* “permite visualizar uma lista de palavras ou grupos de palavras em um texto, organizadas em ordem alfabética ou por frequência” (idem, ibidem). Com o *Keywords* “é possível encontrar as palavras-chaves de um texto” (idem, ibidem). A utilidade de alinhamento, o *Viewer and Aligner*, produz uma

⁴⁶ Scott se refere ao que chamamos de programa, software, de ferramentas, programa (program) e de utilidades, ferramentas (tools).

versão alinhada de dois ou mais textos em línguas diferentes, com frases ou parágrafos alternados de cada um deles (ver seção 3.3.2.2 desta tese).

Atualmente identificam-se muitas pesquisas em ETBC que fazem uso do *WordSmith Tools*⁴⁷. O *WST* é um programa bastante completo, pois apresenta ferramentas e utilidades que permitem o pesquisador a investigar corpora monolíngues ou multilíngues, realizando concordâncias, identificando colocações de palavras no texto e palavras-chave, construindo uma lista de palavras ou alinhando dois ou mais textos em paralelo: “As ferramentas têm sido usadas pelo *Oxford University Press* em seu trabalho lexicográfico de preparação de dicionários, por professores e estudantes de línguas e por pesquisadores que investigam padrões linguísticos em várias línguas, em diversos países”. O site oficial da versão 5 do *WST* relaciona centenas de pesquisas entre dissertações e teses, artigos, livros, capítulos e recursos online que se pautam na utilização do *WST*, muitas delas utilizadas para estudar corpora e tradução, entre elas algumas pesquisas brasileiras (BUENO, 2005).

2.2.2.2. *AntConc* e *AntPConc* (Anthony, 2010)

O *AntConc* (ANTHONY, 2012) é um programa gratuito que apresenta funções semelhantes ao *WST*. Portanto é definido como um programa de multi-plataforma para conduzir pesquisas linguísticas e aulas de língua guiadas por dados. Pode ser usado em diversos sistemas operativos como o *Windows*, *Macintosh OSX* e *Linux*. Sua tecnologia também demanda textos em formato *Plain Text* para poder processá-los. É um programa relativamente novo: a primeira versão 1.0 foi lançada em 2002 e atualmente, em 2013, o programa se encontra em sua 3ª versão 3.2.4 (37 programas foram lançados até agora). Por ser desenvolvido no Japão, o *AntConc* desenvolve bem a função de trabalho com a interface entre inglês e japonês, chinês e coreano, além de outras línguas com mesmo sistema alfabético que o inglês.

Assim como o *WordSmith Tools*, o *AntConc* apresenta ferramentas de concordância, mapeamento da concordância (*concordance plot*), lista de palavras, gerador de palavras-chaves e

⁴⁷ Como será visto em detalhes no capítulo de análise, 85%, ou seja, 17 dos 21 programas utilizados em pesquisas em ESFTBC no Brasil, até 2010, utilizaram o *WST* (Scott, 1999 a 2004) em suas pesquisas.

colocações⁴⁸. As ferramentas do *AntConc* não se diferem muito das ferramentas homônimas do *WST*. Quanto à interface em ambos programas, no *AntConc* observa-se que elas são apresentadas em uma interface simples, objetiva e fácil de aprender, enquanto o *WST* apresenta mais opções para detalhar a apresentação dos resultados e os dados investigados.

Diferentemente do *WordSmith Tools*, ele não apresenta utilidades de alinhamento, não se apresentando muito útil em pesquisas em corpora multilíngues que pretendem analisar corpora paralelos alinhados. Para alinhar e realizar concordâncias em alinhamento, está sendo lançado o *AntPConc* (ANTHONY, 2012). O *AntPConc* é um programa de concordância paralela também gratuito para *Windows*, disponível para ser baixado online. Ainda encontra-se em fase experimental (acaba de ser lançado no site oficial, em junho/2013) e não há informação suficiente a seu respeito para aqui detalhá-la. Isso indica que a carência de alinhamento do *AntConc* será suprida muito em breve, com a proposta de Laurence Anthony.

O *AntConc* não é muito mencionado em pesquisas em ETBC, apesar de ser uma das únicas opções gratuitas que desempenha funções semelhantes ao *WST*. Anthony (2012) menciona que seu uso tem sido mais direcionado a atividades pedagógicas que requisitam recursos de consulta em corpus, mesmo que funcione muito bem em outras pesquisas em corpus. O *AntConc* permite que se analise uma grande quantidade de textos e suas ferramentas se assemelham a do *WST*. É uma alternativa bastante acessível e econômica para quem não necessita uma grande gama de detalhamentos de seus dados e que não tem recursos financeiros para investir na compra de um *software* pago, ou que prefere testar suas hipóteses e seu corpus gratuitamente antes de realizar tal investimento.

2.2.3. Programas adaptáveis

Os programas que aqui chamamos de adaptáveis, como o caso dos Suítes de Aplicativos para Escritório (i.e. Sistemas Operativos *Windows*, *Mac* ou *Linux*), não são muito discutidos nas bibliografias encontradas de ETBC ou de LC. Eles são mencionados, bem vagamente, como uma opção, mas suas ferramentas não são exploradas a fundo neste sentido. Como foi visto, o trabalho que aponta as vantagens de se

⁴⁸ Ver detalhes de utilização no site oficial de Laurence Anthony:

http://www.antlab.sci.waseda.ac.jp/antconc_index.html/

http://www.antlab.sci.waseda.ac.jp/software/antconc330/AntConc_readme.pdf

trabalhar com esse tipo de programa é o publicado por Mason (2008). Esta seção concentra-se na descrição destes programas, principalmente no que se refere às ferramentas que são exploradas durante a elaboração dos *Templates* apresentados no capítulo 5 desta tese. O objetivo desta seção não é descrever exaustivamente todas as possibilidades de exploração de corpus através de programas adaptáveis, mas explorar algumas possíveis alternativas de se trabalhar com tais programas e inspirar novas ideias.

Há algumas vantagens e desvantagens de se utilizar programas adaptáveis em pesquisa de corpus. A grande vantagem é que esses programas, por permitirem ser programados pelo próprio pesquisador em uma plataforma de sua familiaridade geralmente são muito econômico ou gratuito (i.e. *Linux*). Elaborar uma função nesses programas exige um trabalho extra do pesquisador em planejar as funções e a execução. Também demanda conhecimentos básicos a avançados de programação em *Macros*, *VBA* e fórmulas. Embora o pesquisador possa precisar buscar um auxílio de um especialista em informática, muita informação pode ser conseguida na internet: em manuais online, em fóruns do *MS Office* e tutoriais no *Youtube*.

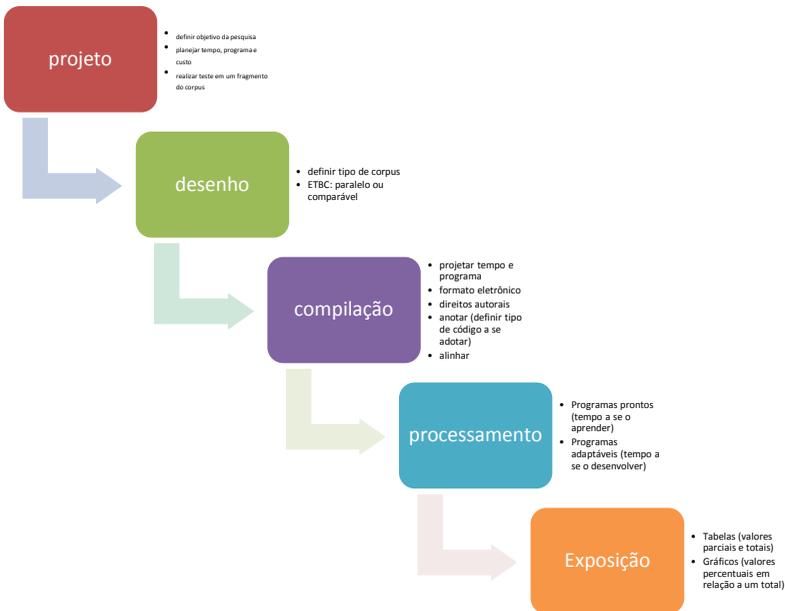
No caso da presente pesquisa são explorados os *Macros* no Processador de Texto *MS Word*, para preparar os textos para serem alinhados, e no Aplicativo de Planilha *MS Excel*, para a formatação dos textos alinhados, arrumar o alinhamento semi-automaticamente e para oferecer o cálculo automático e em gráficos das etiquetas inseridas no programa. Observou-se que nas pesquisas em ESFTBC analisadas nesta tese, o uso de Aplicativos de Planilhas se limita às suas funções básicas, familiares no dia-a-dia de usuários-pesquisadores, tal como a organização de textos em tabelas e/ou de gráficos (cf. Zanella, 2006), a contabilização de dados coletados (cf. Fleuri, 2006), para organização e revisão de rótulos (cf. Feitosa, 2005), o alinhamento manual corpus bilíngue (cf. Alves, 2007). Dentre as 24 dissertações e teses em ESFTBC, seis mencionam explicitamente utilizar Aplicativos de Planilha como o *MS Excel* e o *Open Office Calc.*), são elas as pesquisas de: Nunes (2010), Feitosa (2005), Alves (2007), Zanella (2006), Jesus (2010), Fleuri (2006).

As ideias de programação de *Macros* para a manipulação de textos podem ser infinitas, dependendo da criatividade do pesquisador. Muitas dessas funções são explicitadas no Capítulo 3, de metodologia. Na presente tese utilizam-se o Processador de Texto e o Aplicativo de Planilha da *MS Office*.

2.2.4. Reflexões sobre o planejamento e execução de pesquisas em corpus.

Uma pesquisa em corpus requer um projeto e, se possível, um estudo piloto para avaliar se o orçamento, os recursos e o tempo disponíveis são suficientes para se adotar uma Metodologia de Corpus (Barnbrook, 1996). Uma vez identificado que os objetivos da pesquisa demandam de fato o uso de tal metodologia, define-se o desenho do corpus (cf. Kennedy, 1998; Olohan, 2004; Fernandes, 2006, Borja, 2008 e Varantola, 2002), o modo de compilação deste (Olohan, 2004), a necessidade de anotação e, em caso afirmativo, o código a ser adotado ou criado (cf. Leech, 2005 e Feitosa, 2005), o modo de processamento do corpus (Anderman & Rogers, 2008), a definição do tipo de programa (cf. Bowker, 2004; Garretson, 2008, Mason, 2008 e Sardinha, 2004) e, por fim, o modo de expor os dados quantitativos. Este processo que vai do planejamento à execução de pesquisas em corpus pode ser ilustrado da seguinte forma:

Quadro 2: Processos de planejamento e execução de uma pesquisa em corpus—do projeto de pesquisa à exposição dos dados.



O projeto é necessário, pois conduz à definição do desenho do corpus que deve responder aos objetivos da pesquisa. (Barnbrook. 1996) Este desenho influencia diretamente o modo de compilação deste corpus. A compilação deve considerar o modo de processamento dos dados que devem ser expostos de modo a fornecer o máximo de informação ao leitor, não apenas o valor percentual ou parcial do que se encontrou durante o processamento.

A escolha do tipo de programa interfere na necessidade de se aprender detalhes das ferramentas de programas já prontos, ou de se aprender detalhes de programação básica para adaptar um programa às necessidades da pesquisa. Observa-se que tanto os programas prontos (i.e. *WordSmith Tools* e o *AntConc*), quanto os programas adaptáveis (i.e. Processadores de Textos e os Aplicativos de Planilhas) apresentam vantagens e desvantagens. O que se observa em pesquisas realizadas no campo da LC e dos ETBC é que os Aplicativos de Planilhas são utilizados para desempenhar suas funções básicas de organização de textos e de dados em tabelas e gráficos. O trabalho realizado nestes programas são essencialmente manuais e limitados, tendo seus recursos de gravação de macros não explorados. O Aplicativo de Planilha utilizado nesta tese, o *MS Excel*, demonstra-se muito útil para trabalhar em corpora descartáveis multilíngues de pequena dimensão, pois pode armazenar uma quantidade controlada de texto (apesar de grande é limitada e por isso “de pequena dimensão”) e ser programado para cumprir propósitos específicos de uma determinada pesquisa (por isso “descartável”), tendo se demonstrado útil durante o alinhamento de corpora bilíngues e multilíngues. Já pesquisas que utilizam o *WST* relatam a praticidade de realizar concordâncias automáticas e de obterem dados estatísticos do corpus em segundos, entretanto também relatam dificuldades em alinhar textos bilíngues de modo completamente automático. Ambos os programas, tanto os prontos quanto os adaptáveis, para serem explorados em sua totalidade demandam um tempo de aprendizagem. O *WST* e o *AntConc* demandam um estudo e uma prática de suas funções, enquanto os Aplicativos de Planilhas e os Processadores de textos demandam estudo de programação de *Macros* e de fórmulas. Possivelmente o caráter autoexplicativo dos programas de concordância lexical e o fato de estarem inseridos na tradição do campo da Linguística de Corpus faz com que muitos linguistas e tradutores escolham lidar com métodos de manipulação de corpus no *WST* e *AntConc* e a escolherem o Aplicativo de Planilha para organizar, de modo manual, textos (ou fragmentos dos corpora) e dados em gráficos e tabelas para cumprir determinados propósitos de sua pesquisa.

Os *Templates* desta tese, criados nos programas adaptáveis, não se propõem a substituir os programas prontos, pois estes são muito mais requintados e apresentam ferramentas difíceis ou impossíveis de serem reproduzidas naqueles programas. Esses *Templates* se propõem a oferecer uma nova perspectiva de trabalho com corpora descartáveis e de pequena dimensão, de modo a evitar a conversão de formatos de textos e a transição de um programa a outro, facilitando a compilação e o manuseio do corpus durante a pesquisa. Além disso, os programas adaptáveis oferecem uma alternativa econômica de trabalho em um ambiente computacional familiar a todos os usuários de *Windows*, *Macintosh* ou *Linux*.

O presente capítulo dialogou principalmente com teóricos da LC (Kennedy, 1998; Barnbrook, 1996; Stubb, 1996; Mason, 2008; Scott, 2008; Garretson, 2008; Leech, 2005; entre outros) e dos ETBC (Olohan, 2004; Fernandes, 2006; entre outros) para mostrar que ocupa um espaço essencialmente em estudos que utilizam a Metodologia de Corpus para análise e descrição de textos em relação tradutória. A reflexão sobre ferramentas que processam corpora paralelos bilíngues estreita as relações desta pesquisa aos ETBC. Para estabelecer tal diálogo, foram definidos os campos teórico-metodológicos da tese, discorrendo sobre: a interface entre ETBC e ESFT gerando o que aqui nomeamos de ESFTBC; os estudos de corpus em áreas da linguística e da tradução; particularidades relacionadas ao projeto e execução de pesquisas que utilizam a metodologia de corpus (o desenho, a compilação e a anotação do corpus, o processamento e a exposição dos dados); programas de análises lexicais (*WordSmith Tools* e *AntConc*) e de programas de processamento de texto e de planilhas eletrônicas.

3 METODOLOGIA

The use of machine in linguistic analysis is now established.
(Firth, 1957, p.31)

O presente capítulo descreve os procedimentos adotados nesta tese para a coleta de pesquisas em ESFTBC, o modo como se deu sua descrição e análise, os procedimentos de elaboração dos *Template Editor* e *Template Corpus Paralelo*, incluindo as fórmulas e macros criados e os procedimentos para o estudo piloto. Este capítulo está organizado da seguinte forma: (3.1) Procedimentos de coleta de pesquisas em ESFTBC; (3.2) Procedimentos de descrição das dissertações; (3.3) Procedimentos de criação dos *Templates*; (3.3.1) *Template Editor*; (3.3.2) *Template Corpus Paralelo* e (3.4) Procedimentos no estudo piloto.

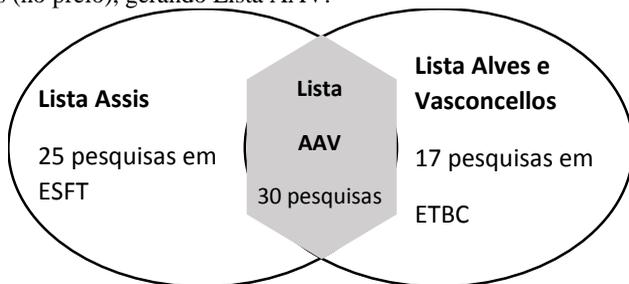
3.1. PROCEDIMENTOS PARA A COLETA DE PESQUISAS EM ESFTBC

São três os procedimentos de coleta de pesquisas em ESFTBC, adotados nesta tese: (i) cruzar os dados da *Lista de Assis* (Assis, 2012) com os dados da *Lista de Alves e Vasconcellos* (Alves e Vasconcellos, no prelo) para gerar uma terceira lista denominada *Lista Assis-Alves-Vasconcellos* (Lista AVV), contendo teses e dissertações em ESFT e ESFTBC; (ii) coletar nos currículos dos orientadores dessas pesquisas, em bibliotecas universitárias e em portais acadêmicos mais dissertações e teses nos ESFT e ESFTBC, gerando a *Lista Assis-Alves-Vasconcellos-Fleuri* (Lista AAVF); e (iii) separar desta lista as pesquisas em ESFT das pesquisas em ESFTBC, obtendo a *Lista ESFTBC*. Cada procedimento engloba suas etapas, que são listadas e descritas nesta seção.

Em relação ao primeiro procedimento—cruzar os dados da *Lista de Assis* com os dados da *Lista de Alves e Vasconcellos*, junta-se em uma lista os trabalhos em ESFT e ESFTBC presentes em cada uma dessas duas listas. O procedimento de cruzar os dados observados de ambas as listas, além de envolver a leitura dos artigos citados e a observação das listas expostas em tais artigos, consiste em:

- a) Ler o título das 17 pesquisas em ESFTBC de Alves e Vasconcellos e das 25 pesquisas em ESFT de Assis para verificar o envolvimento destas pesquisas com as áreas de ET e LSF;
- b) Juntar a lista de 17 pesquisas com a de 25 e eliminar as repetidas. Isso gerou uma terceira lista contendo pesquisas em ESFT e ESFTBC. Essa junção é ilustrada na seguinte figura:

Figura 4: Ilustração do cruzamento da lista de Assis (2012) e da de Alves e Vasconcellos (no prelo), gerando Lista AAV.



A lista AAV apresentam um total de 30 pesquisas em ESFT e ESFTBC entre teses e dissertações publicadas no Brasil desde 1997 a 2010, produzidas na Universidade Federal de Minas Gerais (UFMG) e na Universidade Federal de Santa Catarina (UFSC). Ao combinar ambas as listas na *Lista AAV*, mantenho as ocorrências repetidas, acrescentando 5 pesquisas da lista de Alves e Vasconcellos e 12 pesquisas da lista de Assis. O resultado é uma lista de 30 pesquisas,

Nota-se que o artigo de Alves e Vasconcellos (no prelo) rastreia pesquisas envolvidas nos ETBC, com ou sem arcabouço da LSF, enquanto o artigo de Assis (2012) identifica os trabalhos acadêmicos envolvidos nos ESFT, com ou sem o arcabouço da LC. Ambos realizam suas pesquisas a partir de uma investigação bibliométrica realizada online: Alves e Vasconcellos (idem) buscam por pesquisas publicadas entre 2006 e 2010 no site do governo brasileiro “Domínio Público”, enquanto Assis (idem) parte de uma pesquisa realizada por Pires (2009) e segue com a busca em sites da Capes e do CNPq gerando uma lista de pesquisas publicadas entre 1997 e 2010. Alves e Vasconcellos (idem) identificam 47 pesquisas em ETBC realizadas no Brasil sendo que 17 delas envolvem simultaneamente os ET, LSF e LC, ou seja, encontram-se na área que aqui denominamos de ESFTBC. Assis identifica 25

pesquisas em ESFT publicadas de 1997 a 2010, sem distinguir quais também têm interface com a LC.

Como nem Assis (2012) nem Alves e Vasconcellos (no prelo) investigaram, como parte do método da pesquisa bibliométrica, o CV lattes dos orientadores das pesquisas listadas, parto para o segundo procedimento, o de buscar nos currículos dos orientadores dessas pesquisas, em bibliotecas universitárias e em portais acadêmicos mais dissertações e teses nos ESFT e ESFTBC. Essa busca visa garantir que o máximo de pesquisas publicadas até 2010 sejam incluídas na lista de investigações em ESFTBC. Este procedimento inclui as seguintes etapas:

- (i) Listar todas as orientações defendidas, presentes nos currículos Lattes das três orientadoras mais citadas na lista de Assis e de Alves e Vasconcellos⁴⁹, que são: Dra. Célia Magalhães (UFMG), Dra. Adriana Pagano (UFMG) e Dra. Maria Lúcia B. Vasconcellos (UFSC);
- (ii) Excluir os trabalhos acadêmicos que já haviam sido citados por Assis (2012) e por Alves e Vasconcellos (no prelo) e os que não citavam nenhuma forma de envolvimento com Estudos da Tradução e LSF. No Curriculum da Maria Lúcia Vasconcellos identificaram-se 12 pesquisas, entre teses e dissertações cujos resumos anunciam sua inserção nos Estudos da Tradução com interface com a LSF com ou sem o arcabouço metodológico da LC; dentre os orientandos de Adriana Pagano, são identificados 13 trabalhos entre teses e dissertações; o mesmo procedimento é seguido com o CV da pesquisadora Célia Magalhães, sendo identificadas 13 pesquisas (ver anexo);
- (iii) Buscar pelo resumo dessas pesquisas: (i) online, digitando o título e/ou o nome do completo dos pesquisadores no *Google*; (ii) nos portais da Capes (Plataforma Lattes)⁵⁰ e/ou CNPq;

⁴⁹ Dr. Leila Bárbara (USP), Dr. Lincoln Fernandes (UFSC), Dra. Viviane Herbele (UFSC) e Dr. Ronaldo Lima (UFSC) foram também mencionados como coorientadores de muitos dos trabalhos citados na *Lista AAV*, e portanto considerou-se investigar seus currícula também. Uma busca mais rápida levou-me a entender que as afiliações de Dra. Bárbara, Dra. Heberle são principalmente com a LSF e não necessariamente com os ET, e as de Dr. Fernandes e Dr. Lima são com os ET, apresentando um foco mais voltado à LC e não à LSF. Portanto não foi encontrado, dentro do arco temporal de 1997 a 2010, no curriculum destes pesquisadores nenhum trabalho, além dos citados na lista, que envolvessem ET e LSF ou ET, LSF e LC orientado diretamente por eles, mas sim em ETBC.

⁵⁰ Fonte: <http://buscatextual.cnpq.br/buscatextual/busca.do?metodo=apresentar>

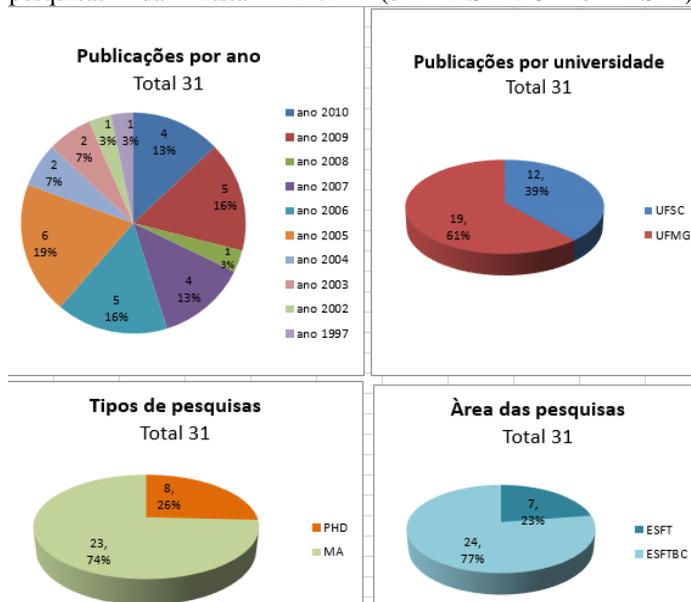
- (iii) Bibliotecas universitárias; e se em nenhuma dessas opções o resumo é encontrado, busca-se o contato do pesquisador e entra-se em contato por e-mail;
- (iv) Buscar, nos resumos, palavras-chaves como: “corpus”, “linguística de corpus”, “corpora”, “linguística sistêmico-funcional” ou termos relacionados (tais como “Processos”, “Metafunção”, “Sistema de”, “Estrutura Temática” etc.), “tradução”, “corpus paralelo bilíngue, trilingue ou multilíngue”, ou qualquer outra palavra que indique envolvimento dos ET com a LC e/ou LSF.

Ao final dessas quatro etapas do segundo procedimento, apenas uma pesquisa é acrescentada à *Lista AAV*: a dissertação de (CAMPESATTO, 2002), que registra envolvimento com ESFT. Esse acréscimo gera uma quarta lista, que chamo de *Lista AAVF*. Essa busca bibliográfica em currículos dos pesquisadores e sites resultou na identificação de muitos trabalhos na área da Análise do Discurso e Linguística Sistêmico-Funcional, dos Estudos da Tradução e Linguística de Corpus, ou dos Estudos da Tradução em interface com a LSF e LC, mas com ênfase em áreas não enfocadas ao estudo descritivo da tradução—i.e. (LIMA, 2008) (BRAGA, 2007); (SILVA, 2007). Essas pesquisas não foram incluídas, pois o enfoque de minha tese está na análise/descrição de textos em relação tradutória. No curriculum de Maria Lúcia Vasconcellos são identificadas três pesquisas envolvidas com os ET e a LC, sem o uso do arcabouço teórico da LSF: (KOBELINSKI, 2005) e de (FERNANDES, 2004).

O terceiro e último procedimento consiste em separar da *Lista AAVF* as pesquisas em ESFT das pesquisas em ESFTBC, obtendo a *Lista ESFTBC*. Esse procedimento passa pelas etapas de:

- (i) Gerar a *Lista AAVF* em uma planilha no *MS Excel*, dedicando colunas para o ano, universidade vinculada, tipo de pesquisa e as área de estudo identificadas nas listas observadas e geradas até então, ou seja, ESFT, ESFTBC e ETBC, conforme exposto no quadro abaixo:

(ii) Gerar gráficos a partir desta tabela. Gráfico 1: Balanço das pesquisas da *Lista AAVF* (em ESFTBC e ESFT)



Como o Gráfico acima mostra, das 31 pesquisas, da *Lista AAVF*, 77% pertencem aos ESFTBC e 23% aos ESFT. Essas etapas geraram um quinta lista que denomino de *Lista ESFTBC* composta por 24 trabalhos, 20 dissertações e 4 teses, apresentando envolvimento simultâneo e exclusivo com os ET, a LSF e a LC. O quadro abaixo expõe o resumo de cada lista gerada nesta tese em relação ao número de trabalhos inclusos:

Quadro 4: as cinco listas desta pesquisa: duas consultadas, a Lista Assis e a Lista Alves e Vasconcellos, e três geradas, a Lista AAV, a AAVF e a ESFTBC.

Listas geradas	No. trabalhos acadêmicos	Estudos envolvidos
<i>Lista Assis</i>	25	ESFT (com ou sem LC)
<i>Lista Alves e Vasconcellos</i>	47	ETBC (com ou sem LSF)
<i>Lista AAV</i>	30	ESFTBC e ESFT
<i>Lista AAVF</i>	31	ESFTBC e ESFT
<i>Lista ESFTBC</i>	24	ESFTBC

A *Lista ESFTBC* resgata todas as pesquisas encontráveis online (até dez/2012) publicadas no Brasil até 2010, relacionadas aos campos

disciplinares Estudos da Tradução e Análise /Descrição Textual, em interface simultânea e exclusiva com Linguística Sistêmico-Funcional e com a Linguística de Corpus.

3.2. PROCEDIMENTOS PARA DESCRIÇÃO DAS DISSERTAÇÕES

São cinco os *procedimentos para descrição das dissertações*: (i) na *Lista ESFTBC*, separar as dissertações das teses; (ii) elaborar um sistema de resgate das informações sobre os procedimentos metodológicos das dissertações; (iii) ler as dissertações e elaborar uma tabela com as informações coletadas; (iv) na segunda leitura, elaborar a descrição detalhada no capítulo de análise e (v) gerar tabelas e gráficos em uma planilha do *Excel* para oferecer o perfil das dissertações em ESFTBC.

A tabela com as informações coletadas, registra: as informações gerais, como título, nome do autor, ano de publicação, universidade afiliada, orientador, tipo de pesquisa (MA ou PhD), e informações específicas como objetivo, objeto de análise, tipo de corpus, unidade de análise, Sistema ou Metafunção da LSF, detalhes do modo de compilação do corpus, programas e ferramentas utilizados, anotação, marcações e tipos de código. Essas informações são organizadas da seguinte forma no capítulo de análise desta tese:

Quadro 5: Informações coletadas e descritas em cada dissertação em ESFTBC

Informação	Significado	Onde se encontra nas dissertações
Informações gerais	Título, ano, universidade afiliada, orientador, tipo de pesquisa	Encontradas nas páginas de abertura da pesquisa.
Objetivo da pesquisa	O objetivo geral da pesquisa	Geralmente encontrada no resumo e no capítulo introdutório
Objeto de análise	O que se analisa ou descreve nos textos que compõem o corpus	Informação dispersa, muitas vezes resgatada/construída apenas por meio de leitura cuidadosa do capítulo de análise
O corpus	Os tipos de corpora utilizados na pesquisa	Geralmente encontrada no capítulo

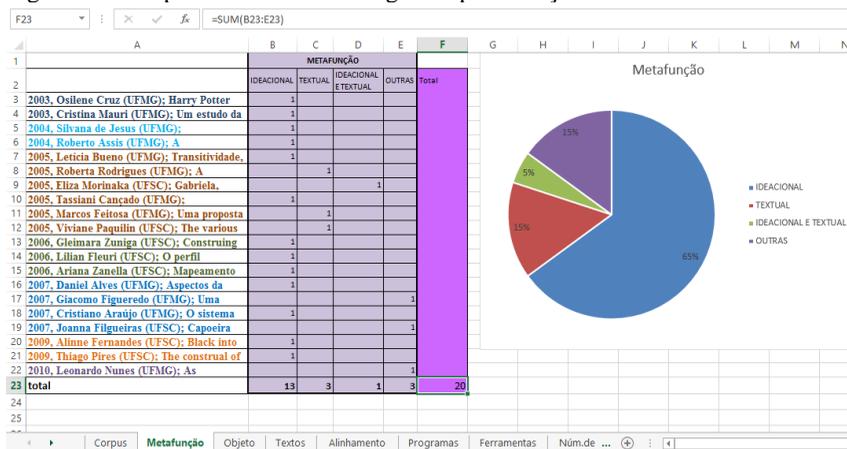
	e a versão em que apresentavam os textos que compunham o corpus	introdutório ou no desenho do corpus, no capítulo de metodologia;
Metafunção	A instância da Gramática Sistêmico-Funcional em que a investigação se fundou	Geralmente encontrada no resumo e no capítulo introdutório, ou sumário e no referencial teórico
Programa e ferramentas	Os softwares utilizados durante o processo de compilação e de processamento do corpus e dos dados.	Geralmente encontrada no capítulo de metodologia, onde se detalhavam os softwares e as suas ferramentas utilizadas durante o processamento do corpus e/ou a quantificação dos dados; na ausência de informação, usa-se o recurso do <i>Word</i> de <i>Find</i> , por meio de digitação de uma palavra—como “ <i>WordSmith Tools</i> ” e/ou o nome de ferramentas do software;
Anotação⁵¹ e Código de Rotulação	Indicação do modo como o pesquisador anotava, para a descrição e análise, o objeto de análise	Por vezes dispersa, ‘detectada’ a partir dos fragmentos de corpus expostos nos capítulos de análise, ou ainda

⁵¹ Nesta tese chamo de anotação ou rotulação/etiquetamento o código inserido dentro do corpus que classifica elementos linguísticos analisados e de marcação (ou *markup*) os códigos que classificam elementos textuais (ver capítulo de Referencial Teórico, seção 2.2.1.2, cf. Kennedy, 1998; Olohan, 2004). Muitos pesquisadores utilizaram os termos anotação ou rotulação intercambiavelmente, raramente utilizam o termo “marcação”; nesta tese faço a distinção entre o momento da descrição e análise para estabelecer o uso de cada termo;

		nos anexos das pesquisas;
Etapas Metodológicas	Passos seguidos pelo pesquisador para realizar a compilação do corpus, a fase de rotulação e a quantificação dos dados	Por vezes dispersa, encontrada nos capítulos de metodologia, de análise e de considerações finais;
Observações	Comentários desta pesquisadora a respeito de elementos relacionados à metodologia e às diferenças da referida pesquisa em relação aos anos anteriores	Essa informação não é encontrada nas dissertação.

A seguir, coleta-se as informações sobre cada um desses itens e as organiza em uma planilha de *Excel*, gerando gráficos (ver anexo e capítulo de Análise, seção 4.2), conforme ilustrado na Figura a seguir:

Figura 5: Exemplo da Planilha Excel gerada para criação de Gráficos.



Essas informações oferecem subsídios para a proposta de um método de compilação de corpus e processamento dos dados para esse perfil de pesquisa, pensando em meios de manipulação de corpus e dados

a partir da criação dos dois *Templates* do *MS Office*, conforme exposto na subseção a seguir.

3.3. PROCEDIMENTOS PARA A CRIAÇÃO DOS TEMPLATES

Visto que a maioria das dissertações em ESFTBC lidam com um corpus paralelo bilíngue (inglês/português) de pequena dimensão, os *Templates* aqui criados buscam responder às necessidades de compilação e de processamento deste tipo de corpus, elaborando alternativas automáticas de: edição de textos para serem alinhados, alinhamento e correção de desalinhamento, flexibilização do modo de rotulação, geração de tabelas e gráficos para visualização e quantificação dos dados. Os procedimentos de criação dos *Templates* são listados nesta seção.

No caso desta tese, esses *Templates* são desenvolvidos nos programas do *MS Word* e do *MS Excel*. Pelo fato de o *Excel* não ser muito apropriado para se lidar com edição de textos, apesar de ser um bom recurso para visualizar tabelas e gráficos, a combinação com o *MS Word* faz-se necessária. Utiliza-se o *MS Word* para editar os textos tornando possível seu alinhamento em sentenças mais automático: eliminam-se pontos de abreviações e quebra-se o texto em sentenças. O *MS Excel* é utilizado para alinhar o corpus paralelo de pequena dimensão, rotular o corpus e quantificar os dados em tabelas e gráficos.

Operar com o corpus no *MS Excel* e no *MS Word* apresenta a vantagem de se encontrar ferramentas que permitem realizar tarefas automaticamente, tais como a ferramenta: (i) de *Macros* em ambos os programas, (ii) de geração de Fórmulas e Gráficos, no *Excel*, e (iii) de busca automática possibilitada pelo *Find & Replace* através de linguagem dos *Wildcards*, no *Word*. Além disso, o *Excel* permite, por exemplo, que: as alterações dos dados sejam recalculadas automaticamente nos gráficos; a revisão dos rótulos e elementos rotulados seja facilitada devido à organização dos dados em tabelas; o pesquisador faça anotações referentes a decisões metodológicas, dentro das próprias planilhas, durante o processo de análise, sendo mais fácil para o pesquisador resgatar as decisões metodológicas tomadas.

Tomando as considerações de Barnbrook (1996) em relação ao projeto em pesquisas em corpus, no caso de escolher lidar com esses programas durante a compilação e processamento do corpus, o pesquisador deve, antes de planejar a elaboração dos *Templates*, observar as características de seu corpus e identificar as necessidades que este corpus apresenta no momento de sua compilação e seu processamento, de

acordo com seus objetivos de pesquisa (ver seção 2.2 desta tese). Se for identificada a necessidade de formatar textos, aconselha-se criar um *Template* no Processador de texto e se a necessidade for de lidar com textos alinhados ou tabulados, tabelas numéricas e gráficos e/ou cálculos e geração de dados quantitativos automáticos aconselha-se criar um *Template* no Aplicativo de Planilha.

Como mencionado, nesta pesquisa elaboro dois *Templates*: o *Template Editor* no Processador de Textos *MS Word*, para editar os textos que compõem o corpus a ser alinhado e o *Template Corpus Paralelo*, no Aplicativo de Planilha *MS Excel*, que organiza o alinhamento, rotulação dos elementos linguísticos/textuais e quantificação e visualização dos dados em tabelas e gráficos. Os passos a se seguir no planejamento e construção do *Template Editor*, são:

- (i) Digitalizar uma amostra do corpus e criar um corpus-teste;
- (ii) Observar suas necessidades de edição e definir os elementos textuais que demandam edição/correção. No caso desta tese são alguns erros de escaneamento mais recorrente, pontos de abreviações, quebra do texto em sentenças etc.;
- (iii) Estudar os recursos de edição automática que respondem à necessidade da edição. No caso desta tese o recurso estudado foi o de *Find & Replace*⁵²;
- (iv) Caso observa-se a necessidade de se criar uma linguagem de programação em *Macros* ou *WildCards*⁵³ (caso o *Find & Replace* seja utilizado), estudar tais linguagens;
- (v) Caso verifique-se a necessidade de utilizar os *Wildcards*, estudar possíveis combinações de *Wildcards* e de substituições;
- (vi) Testar tais combinações e substituições em uma amostra do corpus;
- (vii) Gravar essas combinações em *shortcuts* através da programação de *Macros*.

Os passos a se seguir no planejamento e construção do *Template Corpus Paralelo*, são:

- (i) Definir o tipo de corpus e as necessidades de processamento deste corpus. No caso desta tese, o corpus é paralelo bilíngue (inglês/português) de pequena dimensão e precisa ser

⁵² <http://office.microsoft.com/en-us/word-help/find-and-replace-text-and-other-data-in-your-word-2010-files-HA102350610.aspx>

⁵³ http://www.gmayor.com/replace_using_wildcards.htm

- alinhado, ter elementos textuais e linguísticos rotulados e quantificados;
- (ii) Definir os recursos de *Excel* que devem ser utilizados na elaboração do *Template*; estudar tais recursos, que podem ser: *Find & Replace*, tabela (como *Pivot Table* e de formatação de tabela), gráficos, filtros e de ordenação, ferramentas de dados, janelas, fórmulas e *macros*. No caso desta tese, estudam-se os recursos de tabela, de gráficos, de filtros e de ordenação, de fórmulas e de macros.
 - (iii) Se nesta fase de estudo são consultados fóruns⁵⁴, deve-se elaborar perguntas cujas respostas solucionem o problema encontrado. O elemento-chave, nesta etapa, não é descobrir exatamente a informação, mas obter das fontes disponíveis as informações que auxiliam na formulação de perguntas sobre o que se busca criar. Antes de formular perguntas vê-se necessário: (i) estudar o básico da linguagem computacional, (ii) levantar hipóteses, (iii) testá-las, (iv) observar os erros, (v) criar alternativas e então (vi) elaborar a pergunta que expressa o problema relacionado ao que se pretende realizar (ver as perguntas dos fóruns em anexo)⁵⁵;
 - (iv) Criar as tabelas com fórmulas e *macros*;
 - (v) Gerar gráficos a partir destas tabelas;
 - (vi) Testar em uma amostra do corpus e executar um teste;
 - (vii) Editar o desenho das planilhas e das pastas de trabalho (*Work Sheets*).

As informações sobre os recursos e ferramentas do *Word* e *Excel* podem ser encontrados online, em *links* relacionados ao *Microsoft* e manuais⁵⁶ impressos e online do *MS Word* e *Excel*, tutoriais no *Youtube* e fóruns sobre *MS Word* e *Excel*. Os fóruns representam um importante meio de ter perguntas específicas respondidas por especialistas das áreas mostra-se muito útil na fase de elaboração de fórmulas e programação de

⁵⁴ O fórum mais acessado foi <http://answers.microsoft.com/en-us> ao qual o usuário tem que se registrar antes de realizar suas perguntas (ver anexo para as perguntas e respostas realizadas no fórum durante o processo de criação dos Macros).

⁵⁵ O fórum foi acessado sob o nome de Victor Carreto, no anexo lê-se Lilian Fleuri onde se encontra a pergunta de “Victor Carreto”.

⁵⁶ Microsoft Office Specialist Excel 2003: Study Guide *Computers/Cerfitifation guides / Wiley* Autor Linda F. Johnson Edição ilustrada Editora John Wiley & Sons, Num. págs.504 páginas, Ano 2006.

Macros em VBA. Esses fóruns correspondem à consulta a um programador, mencionada por Garretson (2008), na seção 2.1.2.4.3 desta tese.

Esses são alguns possíveis procedimentos que sugiro seguir para compilar e processar corpus nos programas *MS Word* e *MS Excel*. A seguir expõe-se as etapas de elaboração do *Template Editor* e *Template Corpus Paralelo* propostos nesta tese.

3.3.1. *Template Editor*

Os procedimentos de criação do *Template Editor* desta tese são: (3.3.1.1) criar o corpus-teste; (3.3.1.2) estudar os recursos de edição automática, como o *Find & Replace* e a linguagem de *Wildcards*; (3.3.1.3) criar os códigos de *Wildcard*; (3.3.1.4) realizar testes e reestruturar códigos conforme os resultados; (3.3.1.5) fixar uma sequência de substituições e gravar macros em *shortcuts*. A seguir cada procedimento é detalhado com mais passos e com exemplos de o que se realizou nesta tese.

3.3.1.1. Montagem de um corpus-teste

Como apontado na seção 2.2.1.2. *Compilação do corpus*, é praticamente impossível, com a tecnologia atual realizar um alinhamento completamente automático, sem que se tenha que passar antes por uma preparação ‘manual’ do texto pelo pesquisador, que consiste em: eliminar os pontos de abreviações, realizar a quebra do texto em sentenças e reestruturar as unidades de alinhamento, após alinhamento. Considerando esses procedimentos, nesta tese montou-se um ‘corpus-teste’ que englobasse os problemas previstos em Alves e Morinaka (2004) e que isolasse alguns problemas detectados no corpus que seria utilizado no Estudo Piloto, ou seja, o corpus utilizado na minha dissertação, Fleuri (2006). O corpus-teste é composto pelos seguintes textos e elementos a seguir:

- 1) Texto original: **Translators Through History** (doravante **TTH**) (Deslile e WoodsWorth, 1995), salvo no *MS Word*, processado no *MS Word* 2010, fonte Verdana tamanho 10, parágrafo simples, 148 páginas, 104.811 palavras.
- 2) Texto traduzido: **Os Tradutores na História** (idem, tradução por Sérgio Bath) (doravante **OTH**), salvo no *MS Word*, processado no *MS*

Word 2010, fonte Verdana tamanho 10, parágrafo simples, 146 páginas, 103.482 palavras.

Figura 6: Elementos adicionados ao corpus para teste.

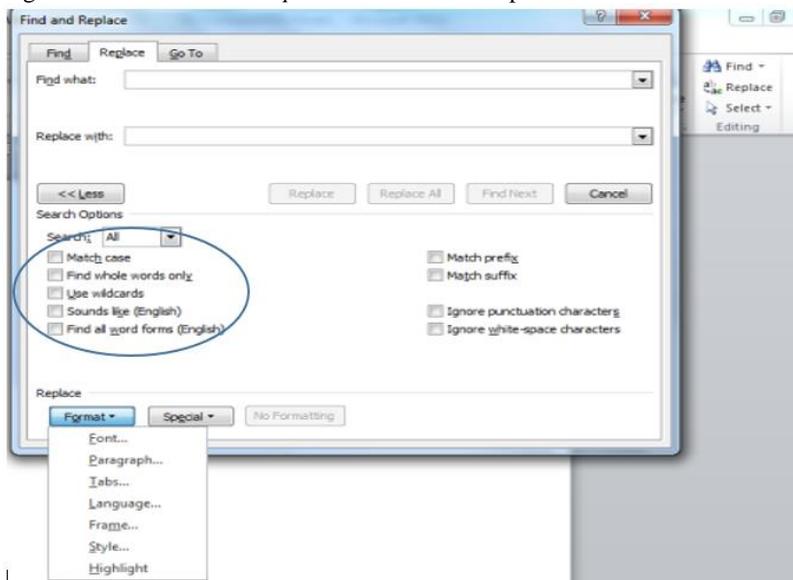
1. Pineapples; 2. Bears; a. Situations; b. Down Under; i. John; ii. Mary; I. Index; II. Mirror.
 ie Apples
 fig. 1
 fig. Mary
 fig. above
 U.S.A. were visited by
 U.S.A. Were visited by
 (Lilova 1979: 23). **In this article**
 [gasp]. **Well**
 into instruments of humanism, [...]. **Lord Smith said once**
 He runs <0010310>. **He jumps**
 He runs <mat>. **He jumps**
 The sequence code is {0uijkjkh}. **Therefore**
 Do you like... mmm... HELP... help... help... help...!
 “This is what it is”. **Said my brother.**
 “This is what it is.” **Said my brother.**
 This is what it is. “Said my brother”
 I went to the cinema . **It was yesterday.**
 I went to the cinema. **It was yesterday.**
 He was born in 1960. **He was...**
 The meeting is on the 8th. **I can’t wait!**
 Since the 1980’s. **In particular, translation**
 I was born in Brazil. **This is a**
 I was born in USA. **This is a**
 This is irresponsibility. **Therefore**
 FIT. **The translator**
 século XV. **The translator**
 Dr. John said he will be here in 10 mins. **He said**
 Prof. Smith is never late. **He just**
 Wait! **Don’t let me here.**
 He said “wait”! **And run.**
 WAIT! **Don’t let me here.**
 Peter? **Is that you?**
 He said “Peter”? **Is that you?**
 PETER? **Is that you?**
 Help...! **But nobody heard.**
 Peter...? **Is that you?**

Textos 1 e 2 representam o corpus adotado em Fleuri (2006) e a Figura acima representa o corpus criado com os elementos que devem ser codificados para que todos os tipos de sentenças sejam ‘quebradas’ automaticamente através do *Find & Replace*. Os elementos dessa Figura são incorporados ao texto em português e ao texto em inglês que compõem o corpus paralelo do Estudo Piloto—o corpus-teste.

3.3.1.2. Estudo dos recursos de edição automática—*Find & Replace* e a linguagem de *Wildcards*

Os comandos avançados mais acessados do *Find & Replace* são os botões “Formatar” e “Especial” e as opções destacadas na figura abaixo:

Figura 7: Tela do *Find & Replace* – Formatos e Especiais



Coloca-se o cursor no espaço ‘*Find*’ (Busca), marca-se a opção ‘*Use Wildcards*’ (Usar caracteres curingas) e digita-se uma combinação de códigos *Wildcards* nesse espaço. Busca-se online explicações dos códigos e realiza-se tentativas de combinações deste código para a desejada substituição. Os sites consultados nesta tese foram:

- <http://office.microsoft.com/en-us/word-help/find-and-replace-text-and-other-data-in-your-word-2010-files-HA102350610.aspx>

- http://www.gmajor.com/replace_using_wildcards.htm

3.3.1.3. Criação dos códigos de *Wildcard*

Após definido o corpus para o teste e estudados os recursos de edição automática, como o *Find & Replace* e a linguagem de *Wildcards*, definem-se as combinações de *Wildcards*. Como a criação de um só código que resolvesse todos os problemas é difícil de ser criado, desenvolve-se uma série de combinações de *Wildcards* de substituição, gravando cada um em um *Macro*. Esse conjunto de Macros gera o *Template Editor* (ver Quadro seguinte).

As combinações são inclusas no espaço ‘*Find*’ do *Find&Replace*, e são variáveis de *caracteres+pontos+espaço+caracteres*. Um ponto pode ser antecedido de letra minúscula ou maiúscula, sinais de parênteses, aspas (que pode vir antes ou depois do ponto), letras subscritas, outros pontos. Em geral, *ponto+espaço+ letra maiúscula* é uma combinação fixa. Em alguns casos pode ocorrer um sinal de parênteses, aspas ou, no caso do espanhol, um ponto (como *¿*) antes da letra maiúscula. Este *Template* não contempla estes casos, pois foi desenhado para editar textos em língua portuguesa e inglesa (não espanhola) e porque o ‘corpus-teste’ não apresentou casos de frases iniciadas por aspas ou parêntese (depois e um ponto).

Há outras particularidades, provindas ou não do escaneamento, que podem impedir ou dificultar a leitura exata do código, tais como espaços duplos, espaços entre palavra e ponto final e aspas depois do ponto (como a presença constante de *Tabs* no texto). Prevendo alguns desses casos, elaboro códigos de substituição para eliminar ou arrumar tais (possíveis) particularidades, expostos nas seção 3.3.1.5 desta tese.

3.3.1.4. Realização de testes e fixação da sequência de códigos em *Wildcards*

Definidas as combinações, parte-se aos testes de substituição. Digitam-se as combinações em *Wildcards* no espaço ‘*Find*’ e outras combinações no espaço ‘*Replace*’ conforme os exemplos abaixo.

No caso desta tese, o primeiro teste praticado no corpus-teste foi o de substituir *ponto+espaço+letra maiúscula* por *ponto+quebra de sentença manual+letra maiúscula*, visando testar a quebra de sentença, antes da eliminação dos pontos de abreviações:

Equação 1: Exemplo do primeiro teste do *Template Editor*

Find what	Replace with
. [A-Z]	./^[A-Z]

O *Template Editor* apresenta um total de onze substituições, resultado de inúmeros testes como os exemplificados na Equação 1, de Wildcards. Após os testes, fixa-se uma sequência e gravam-se comandos em *Macros*. A sequência das onze substituições e os comandos de *Macros* correspondentes são:

Quadro 6: combinações de *Wildcards* para substituições e respectivos *shortcuts* gravados em *Macros*.

	Ação	Shortcut gravado	Comando de substituição em Wildcards	
			Find what	Replace with
1.	Arrumar espaços duplos entre ponto e letra maiúscula	Alt+0	(\.\? !) ([A-Z ^c])	I 2
2.	Invertem as aspas localizadas depois do ponto final	Alt + 1	(\.\? !)(^c) ([A-Z ^c])	2 I 3
3.	Retira pontos de abreviações de duas letras, sendo a primeira maiúscula e a segunda minúscula (Mr. Sr.)	Alt + 2	([A-Z][bcdfghjklmnpqrstvwz]{1,2}). ([A-Z ^c])	I 2
4.	Elimina pontos de abreviações com quatro letras (como Rev. e Prof.)	Alt + 3	([PR][evrof]{1,3}). ([A-Z ^c])	I 2
5.	Elimina pontos de itens com letras minúsculas	Alt + 4	(<[a-z]{I}>). ([A-Z ^c])	I 2
6.	Quebra linhas em frases terminadas por parênteses (ou símbolos semelhantes) e pontos.	Alt + 5	(\)\ / > >'')(\.\? !) ([A-Z ^c])	I 2^ 3

7.	Quebra linhas após números (de três e quatro dígitos) mais pontos.	Alt + 6	<i>([0-9]{3,4})([. \? \!]) ([A-Z^c])</i>	I 2^ 3
8.	Quebra linhas em frases terminadas por palavras em letras maiúsculas e pontos	Alt + 8	<i>([!.][A-Z]{2})([. \? \!]) ([A-Z^c])</i>	I 2^ 3
9.	Quebra linhas em frases terminadas em aspas e ponto.	Alt+p	<i>(")([. \? \!]) ([A-Z])</i>	I 2^ 3
10.	Elimina os símbolos de Tab—podem causar problemas durante o alinhamento, elimina-se como um modo preventivo.	Alt+B	<i>^t</i>	<i>Não se digita nada</i>
11.	Para realizar todas os comandos de Alt+0 a Alt+B de uma só vez.	Alt+T	<i>Este Macro foi resultado da gravação contínua de todos os Macros anteriores, nesta sequência.</i>	
12.	Salvar o texto-fonte e repetir no texto-alvo			
	<i>Observação:</i> as aspas tem que ser copiadas do texto digitado no Word e coladas na sequência de Wildcards, para que o Find & Replace consiga ler e diferenciar uma aspa que abre o texto (“) da aspa que fecha o texto (”). O símbolo ^ significa “Quebra manual de linha” (Manual line break), que se encontra no ícone “special” da janela do “Find & Replace”.			

Durante a utilização do *Template Editor*, com o Quadro acima como guia, o pesquisador tem a opção de: realizar cada substituição por vez, com os *shortcuts* expostos; de realizar todas as substituições de uma só vez através do *shortcut* Alt+T ou de digitar cada sequência de *Wildcards*. Oferecer a tabela nesta seção permite que esta seja adaptada por futuros pesquisadores de acordo com as necessidades de seu corpus.

3.3.2. *Template Corpus Paralelo*

Os procedimentos de criação do *Template Corpus Paralelo* desta tese são: (3.3.2.1) definição do desenho do *Template*; (3.3.2.2) estudo dos programas prontos cujas ferramentas são replicadas; (3.3.2.3) criação de

macros e fórmulas; (3.3.2.4) teste de *macros*; (3.3.2.5) criação de fórmulas, tabelas e gráficos; e (3.3.2.6) edição do desenho das pastas do *Template*. A seguir cada procedimento é detalhado com mais passos e com exemplos de o que se realizou nesta tese.

3.3.2.1. Definição do desenho da planilha e de suas pastas de trabalho

A definição do desenho da planilha do *Template Corpus Paralelo* deve estar de acordo com as funções que se busca oferecer. No caso desta tese, crio uma pasta de trabalho (*worksheet*) para cada função (alinhamento, rotulação, geração de lista de elementos rotulados e gráficos que os quantifiquem e geração de tabelas). O *Template Corpus Paralelo* apresenta 5 pastas, denominadas de: (i) “Alinhamento; (ii) “Rotulação TF”; (iii) “Rotulação TA”; (iv) “Listas e Gráficos”, e (v) “Dados” (ver planilhas na seção 5.2 desta tese).

3.3.2.2. Estudo dos programas prontos para réplica de suas ferramentas

A programa pronto estudado foi *WordSmith Tools* (Scott, 2011) e sua ferramenta de alinhamento, o *Viewer and Aligner*. Observou-se particularmente o *layout* do alinhamento e as funções que permitem corrigir as unidades desalinhadas. Essas funções são:

- (i) *Join* (Juntar), que junta uma sentença com a de cima; permitindo solucionar desalinhamentos causados pela presença de abreviações e casos em que uma frase do texto-fonte tenha sido traduzido em duas frases;
- (ii) *Split* ou *Um-join* (Separar), que separa a frase no ponto em que se deseja tal separação; permitindo separar duas frases do texto-fonte que tenham sido traduzidas como uma frase;
- (iii) *Match by Capitals* (Busca por letras maiúsculas), que acelera a busca pelos casos de desalinhamentos, apontando os nomes próprios nas duas versões, original e traduzida. Este comando permite que o usuário identifique quando há um desalinhamento (por exemplo, na tradução encontra-se o nome próprio ‘Paris’ e no texto-original não, é possível que essas frases não sejam correspondentes) e o corrija a partir das funções *Join* ou *Un-Join*.

O *WordSmith Tools* não oferece uma função que permita inserir uma linha em branco no lugar da ausência de correspondência (cf. Scott, 2004-2005), no caso de frases presentes no texto-fonte e não traduzidos no texto-alvo.

Quanto ao *layout*, o *Viewer and Aligner*, quando alinha os textos por sentenças, posiciona-as uma abaixo da outra, formando a *unidade de alinhamento* (cf. Alves, 2000 e Assis, 2004). As linhas correspondentes ao original têm cores diferentes das do texto traduzido. Particularmente não acho este *layout* prático, visto que quando há um desalinhamento é difícil buscar as unidades de alinhamento. Uma visualização lado-a-lado, como realizam os pesquisadores que optam por alinhar seu corpus manualmente (i.e. Cruz, 2003), parece-me mais prática.

A partir do estudo das funções do *WST*, defino o *layout* da pasta de ‘Alinhamento’ no *Template Corpus Paralelo*. Os textos são alinhados lado-a-lado por sentença; cada coluna com cores distintas uma da outra; crio as opções “juntar”, “separar” e “buscar células” em *Macros* no *Excel*; nos casos de desalinhamentos causados por falta de tradução, o usuário do *Template* tem a opção de “inserir célula”, através de um comando homônimo já disponível no *Excel*.

3.3.2.3. Criação dos *Macros* e Fórmulas

No *Template Corpus Paralelo* desenvolvo cinco *Macros* de dois tipos: dois *Macros* gravados, que formatam as colunas após terem seus textos colados e três *Macros* em *VBA*⁵⁷, que arrumam o alinhamento: um busca as unidades desalinhadas, outro combina a célula de baixo com a de cima e outro separa uma célula em duas.

As ações gravadas em *Macros* na Pasta de alinhamento do *Template Corpus Paralelo* e seus respectivos *shortcuts* são expostos no seguinte quadro:

Quadro 7: Comandos e seus *shortcuts* – Pasta Alinhamento do *Template Corpus Paralelo*

	Ação	Shortcut gravado
1.	Formatar a coluna A: apaga as linhas em branco, deslocando as células para cima, wrap os textos	Ctrl+Q

⁵⁷ Há duas formas de criar *Macros*, gravando e programando em *VBA*. ‘Gravar *Macros*’ significa ativar o comando de gravar e começa a realizar a sequência de ações, desativando o comando quando essa sequência é finalizada. Esta é uma forma muito simples, podendo ser realizada sem necessidade de um estudo de programação de *Macros*. Programar *VBA* é mais complexo e necessita o estudo de programação de *Macros* em *Excel* ou *Word* para escrever comandos que realizem as ações desejadas.

	de todas as linhas, marca borda em torno de todas as células.	
2.	Formatar a coluna B: apaga as linhas em branco, deslocando as células para cima, wrap os textos de todas as linhas, marca borda em torno de todas as células.	Ctrl+W
3.	Buscar por unidades desalinhadas: leva o curso a linhas cujas sentenças estão possivelmente desalinhadas.	Ctrl+M
4.	Juntar célula abaixo (A2) com célula acima (A1), sem alterar a ordem da linha da coluna ao lado: copia o texto contido em uma determinada célula (i.e. A2) e o leva automaticamente ao final do texto da célula acima (i.e. A1).	Ctrl+J
5.	Separar uma célula em duas linhas, sem alterar a ordem da linha da coluna ao lado: levar o cursor ao ponto do texto desejado e dar o comando Alt+Enter; selecionar esta célula e acionar o comando Ctrl+S	Alt+Enter (comando do Excel) Ctrl+S (comando gravado no <i>Template</i>)
6.	<i>Salvar o texto-fonte e repetir no texto-alvo</i>	

O processo de gravar *Macros* para formatar as colunas, como os casos 1 e 2, ocorre da seguinte forma:

- (i) Acionar o botão de gravar *Macros (Record Macros)*
- (ii) Iniciar a sequência de ações, gravando-a *shortcuts*;
- (iii) Selecionar as linhas em branco;
- (iv) Deletar as linhas em branco deslocando células para cima;
- (v) *Wrap* textos;
- (vi) Adicionar borda;
- (vii) Aplicam esses procedimentos primeiro na coluna A e depois na coluna B, gravando com diferentes *shortcuts*.

O processo de criar *Macros* em *VBA*, como os casos 3, 4 e 5, conta com os seguintes procedimentos:

- (i) Criar hipóteses,
- (ii) Elaborar perguntas;

- (iii) Buscar respostas em manuais, tutoriais e fóruns especializados em programação *VBA* para Excel (ver anexo 8);
- (iv) Testar as respostas encontradas e as oferecidas no fórum, e reformular a pergunta caso necessário.

O *Macro* 3, que automatiza o processo de busca por sentenças desalinhadas, gravado com o *shortcut* Ctrl+M, demanda a criação prévia de uma fórmula que deve ser adicionada à programação do *Macro* em *VBA*. A criação da fórmula contou com a hipótese (constatada) de que em um corpus bilíngue português e inglês as sentenças que formam uma unidade desalinhada apresentam uma diferença de mais ou de menos dez palavras entre elas. A partir da constatação desta hipótese, elaborei uma fórmula de *Excel* para contar o número de palavras em uma célula. Em seguida elaborei outra fórmula para buscar pares de linhas A/B que apresentam uma diferença de 9 palavras absolutas (-9 ou +9)⁵⁸ entre as células A e B. As fórmulas são as seguintes:

Equação 2: Fórmulas para se buscar desalinhamento entre textos em Português e Inglês—Pasta Alinhamento no *Template Corpus Paralelo*

Função	Fórmula
Contar palavras de uma determinada célula da coluna A.	=LEN(A1)-LEN(SUBSTITUTE(A1,"",","))+1
Buscar pares de linhas A-B que apresentem uma diferença de 9 palavras absolutas (-/+9)	=IF(ABS((LEN(A1)-LEN(SUBSTITUTE(A1,"",","))+1)-(LEN(B1)-LEN(SUBSTITUTE(B1,"",","))+1))>=9,"MISMATCH",0)

Acrescento estas fórmulas em perguntas levadas a um fórum, com o intuito de criar um *Macro VBA* para buscar pares de células A/B com as condições desta fórmula (os *Macros* criados encontram-se nos anexos 9 e 10).

3.3.2.4. Teste e reelaboração dos *Macros*

Esta etapa engloba os processos de:

- (i) Realizar os testes várias vezes durante a elaboração dos *Macros*;

⁵⁸ Diminui-se uma palavra para garantir que todos os pares desalinhados fossem encontrados.

- (ii) Anotar os problemas encontrados,
- (iii) Reestudar os casos e
- (iv) Reformular as perguntas, levando-as novamente aos fóruns.

No caso do *Template Corpus Paralelo*, como será exemplificado no capítulo 6, de Estudo Piloto, o teste final mostrou que o *Macro* ‘Busca’ é capaz de captar todos os casos de desalinhamento, que podem ser corrigidos através do *Macro* ‘Juntar’ ou ‘Separar’, ou ainda, nos casos de não haver correspondente, uma linha em branco pode ser adicionada manualmente através de comandos oferecidos pelo próprio *Excel*, conforme exemplificado no Capítulo 4, de análise, desta tese.

3.3.2.5. Criação de fórmulas, tabelas e gráficos

No *Template Corpus Paralelo* há três tipos de fórmulas: as que organizam as etiquetas ao lado dos elementos etiquetados (criadas nas pastas ‘Rotulação TF’, ‘Rotulação TA’), as que quantificam dados e os expõem em tabelas e gráficos (criadas na pasta ‘Lista e Gráficos’) e as que quantificam nódulos no corpus (criadas na pasta ‘Dados’). O *layout* de cada pasta é ilustrado e explicado, para se entender onde e como as fórmulas são criadas.

O *layout* das pastas de ‘Rotulação do Texto-Fonte’ e ‘Rotulação do Texto-Alvo’, e a descrição do processo de elaboração de cada coluna, são expostos na figura a seguir:

Figura 8: Pasta de Trabalho de Rotulação do TF do *Template Corpus Paralelo*.

J	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2	PREFACE	ACTOR	GOAL	CLIENT	RECIPTE	SENSOR/IM	IDENTR	IDENTO	SAVER	VERB	RECEIVER	EXISTING	BEHAV				ELEMENTO	CATEGORIA
3	Turning translations into instruments of humanism, peace and progress such is our noble task.																	0
4	These are the words of Pierre-François Calle (1907-79), founding President of the International Federation of Translators: they reflect his personal philosophy, which he passed along to the Federation when it was established in 1953 (Llova 1979: 23).								1								REFLECT	IDENTFO
5	In article 6 of its bylaws, FIT proclaims its responsibility to "assist in the spreading of culture throughout the world".																	0
6	The tens of thousands of translators who belong to the seventy-three member organizations of our Federation								1								BELONG	IDENTFO
7	spare no effort to fulfil this mission.	1															SPARE	ACTOR
8	The work they perform on a day-to-day basis attests to the fact that translation permeates all facets of human activity and is an inexhaustible source of progress.																	0
9	People have translated since time immemorial.																	0
10	Long before FIT, translators served as vicararies in the vast chain through which knowledge was transmitted among groups of people separated by language barriers.								1								SERVE	IDENTFO
11	Ever since humans first devised writing systems, translators have been building bridges between nations, races, cultures and continents.	1															BUILD	ACTOR
12	Bridges between past and present, too. Translators have the ability								1								HAVE	IDENTR
13	to span time and space.	1															SPAN	ACTOR
14	They have enabled certain central texts - works of science, philosophy or literature to acquire universal stature.																	0
15	Translators breach the walls created by language differences, thereby	1															BREACH	ACTOR
16	opening up new horizons and broadening our vision of reality	1															OPEN UP	ACTOR
17	to encompass the entire world.	1															ENCOMPASS	ACTOR
18	Translators live off the differences between languages, all the while working	1		1													LIVE OFF	CLIENT
19	toward eliminating them" (Edmond Cary 1956: 181).	1															WORK	ACTOR
20	Yet translators have been widely scorned at times and																ELIMINATE	ACTOR
21	their work severely criticized.																SCORN	0
22	These educated men and women of letters have been																WORK	0
23	distrusted, even called turncoats and traitors.																	0

- (i) Na coluna A, da pasta de Rotulação do Texto-Fonte, cola-se o Texto-Fonte;
- (ii) As colunas B à P são dedicadas aos rótulos—cada coluna recebe o nome de um rótulo. O pesquisador manualmente deixa na coluna A somente os elementos a serem rotulados e em seguida marca “1” em cada célula correspondente ao elemento (na linha) e ao rótulo (na coluna) que corresponde ao elemento (veja que os elementos ‘anotados’ estão em negrito e o número 1 é marcado na coluna que o classifica);
- (iii) As colunas Q e R oferecem a visualização lado-a-lado dos elementos rotulados e o nome do rótulo, facilitando a revisão dos elementos rotulados;
- (iv) Na coluna Q o pesquisador anota manualmente o elemento que esta rotulando (isoladamente do contexto);
- (v) A coluna R recebe uma fórmula que realiza a cópia do nome do rótulo marcado na linha correspondente.

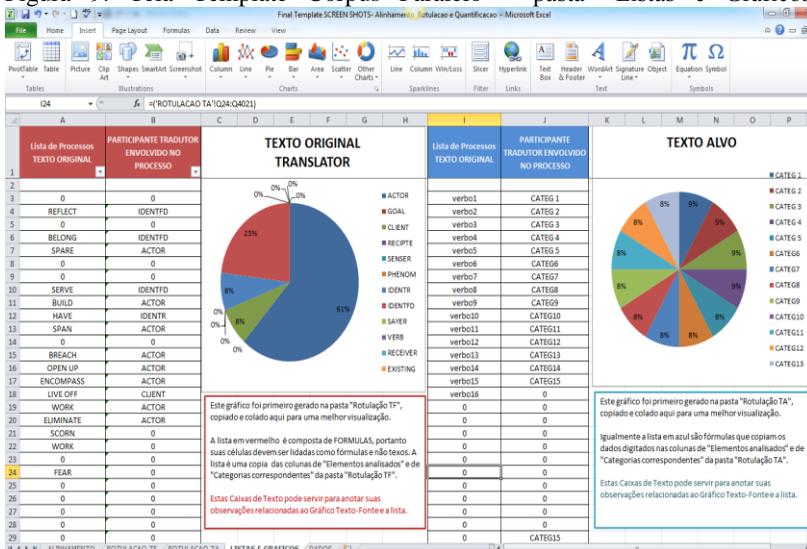
A fórmula acrescentada na coluna R é a seguinte:

Equação 3: Fórmula que transfere o nome da etiqueta ao lado do elemento etiquetado, nas pastas ‘Rotulação TF’ e ‘Rotulação TA’.

Fórmula	Significado
=IF(C3=1,\$C\$2,IF(D3=1,\$D\$2,IF(E3=1,\$E\$2,IF(F3=1,\$F\$2,IF(G3=1,\$G\$2,IF(H3=1,\$H\$2,IF(I3=1,\$I\$2,IF(J3=1,\$J\$2,IF(K3=1,\$K\$2,IF(L3=1,\$L\$2,IF(M3=1,\$M\$2,IF(N3=1,\$N\$2,IF(O3=1,\$O\$2,IF(P3=1,\$P\$2,IF(B3=1,\$B\$2,0)))))))))))))))))	Leva o nome da etiqueta, escrita das células B2 a P2, à coluna R, na linha correspondente em que foi anotada (com o número 1).

O layout da pasta ‘Lista e Gráficos’ apresenta uma lista e um gráfico para o Texto-Fonte (em vermelho) e para o Texto-Alvo (em azul), conforme ilustrado na Figura abaixo:

Figura 9: Tela Template Corpus Paralelo – pasta “Listas e Gráficos”



As listas, desenvolvidas no par de colunas A/B e I/J, são cópias das colunas Q e R da pasta de ‘Rotulação’ correspondente. Isso faz com que as mudanças geradas nessas pastas alterem automaticamente os dados na pasta ‘Lista e Gráficos’. Proponho criar gráficos a partir dessas listas, através do comando disponibilizado pelo *Excel*, de ‘inserir gráficos’.

O layout da pasta de ‘Dados’ é muito simples, conforme ilustrado na Figura abaixo:

Figura 10: Tela da quinta área de trabalho (pasta) do *Template Corpus Paralelo*

	A	B	C	D	E
2	ELEMENTOS TF	QUANTIDADES		ELEMENTOS TA	QUANTIDADES
3	Numero de nodulos			Numero de nodulos	0
4	Numero de Processos realizados pelo Participante rotulado			Numero de Processos realizados pelo Participante rotulado	16
5					

Elaborei duas tabelas nesta pasta, uma referente ao Texto-Fonte (TF), no par de colunas A/B, e outra ao Texto-Alvo (TA), no par de colunas D/E, com duas linhas, informando o número de nódulos presentes em cada texto e o número de Processos realizados pelo Participante rotulado. Criei esta pasta para responder a demanda de contagem dos nódulos “Tradutor” e “Translator”, no caso específico do Estudo Piloto. Deixei esta pasta como parte do *Template*, para que outros usuários tomem esta forma de contagem de nódulos como exemplo, caso optem em adaptar a pasta aos objetivos de suas pesquisas.

Nesta pasta, criei um total de quatro fórmulas, que são:

- (i) Duas que quantificam o número de nódulos, no caso do Estudo Piloto, os nódulos eram “translator” no TF e “tradutor” no TA.

Equação 4: Fórmulas que quantificam os nódulos “translator” e “tradutor”, na pasta ‘Dados’

Fórmula	Significado
=COUNTIF('ROTULACAO TF'!A3:A4000,"*?ranslator?*")	'ROTULACAO TF' corresponde à área de trabalho, ou seja, segunda pasta; A3:A4000 corresponde à coluna, * e ? são Wildcards que significam: * é qualquer ou nenhum caractere e ? é um qualquer caractere. Então * corresponde à todo o restante do texto da célula e ? a diferentes letras ou formatos de letras (ex. T maiúsculo ou t

	minúsculo, ou a presença ou ausência de um "s" indicador de plural) que podem ocupar essa posição.
=COUNTIF('ROTULACAO TA'!A3:A4000,"*?radutor?*").	'ROTULACAO TA' corresponde à área de trabalho, ou seja, segunda pasta; A3:A4000 corresponde à coluna, * e '?' são <i>Wildcards</i> que significam: * é qualquer ou nenhum caractere e '?' é um qualquer caractere. Então * corresponde à todo o restante do texto da célula e '?' a diferentes letras ou formatos de letras (ex. T maiúsculo ou t minúsculo, ou a presença ou ausência de um "s" indicador de plural) que podem ocupar essa posição.

- (ii) Duas que quantificam automaticamente o número de elementos que estão sendo analisados, ou seja, de elementos rotulados. A coluna Q da pasta "Rotulação TF" e "Rotulação TA" corresponde ao elemento analisado, um para cada linha considerada. Para quantificar automaticamente o "Número de elementos analisados" no TF e no TA são inseridas a seguinte fórmulas nas células B4 e E4:

Equação 5: Fórmulas que quantificam os elementos analisados, na pasta 'Dados'

Fórmula	Significado
=COUNTIF('ROTULACAO TF'!Q4:Q4001,"?*")	Número de elementos analisados no TF
=COUNTIF('ROTULACAO TA'!Q4:Q4001,"?*")	Número de elementos analisados no TA

3.3.2.6. Edição do desenho das planilhas

Ao final da criação do *Template* e seus *Macros*, Fórmulas, Tabelas e Gráficos, reviso o desenho de cada pasta, arrumando as cores, as bordas, a fonte etc.

Esta é a etapa final da elaboração dos *Templates*. É a etapa em que:

- (i) As tabelas são coloridas e as bordas são (re)desenhadas;
- (ii) As fórmulas e os *Macros* revisados;
- (iii) Os Gráficos editados.

Esses são os seis procedimentos que adoto na criação do *Template Corpus Paralelo*: (i) definir o desenho da planilha e suas pastas de trabalho; (ii) estudar os programas prontos cujas ferramentas desejam ser replicadas; (iii) criar os macros e fórmulas necessários que realizem tais réplicas; (iv) testar macros e reelaborá-los se necessário; (v) criar fórmulas, tabelas e gráficos que exponham os resultados quantitativos automaticamente; e (vi) editar o desenho das planilhas. Enquanto o capítulo de análise desta tese busca apresentar e descrever o funcionamento dos *Templates*, esta seção detalhou a criação das fórmulas e *Macros*, explicitando os recursos utilizados durante o processo de sua criação. Os detalhes de Macros e Fórmulas encontram-se em anexo.

3.4. PROCEDIMENTOS DE EXECUÇÃO DO ESTUDO PILOTO

Para a realização do Estudo Piloto adoto o mesmo corpus da dissertação de Fleuri (2006). Este corpus já havia tido seus erros ortográficos de escaneamento corrigidos, mas lhe faltava a preparação pré-alinhamento. É um corpus paralelo bilíngue (inglês/português) de pequena dimensão, composto pelos seguintes textos:

- 1) Texto original: **Translators Through History** (doravante **TTH**) (Delisle e WoodsWorth, 1995), salvo no *MS Word*, processado no *MS Word* 2010, fonte Verdana tamanho 10, parágrafo simples, 148 páginas, 104.811 palavras.
- 2) Texto traduzido: **Os Tradutores na História** (idem, tradução por Sergio Bath) (doravante **OTH**), salvo no *MS Word*, processado no *MS Word* 2010, fonte Verdana tamanho 10, parágrafo simples, 146 páginas, 103.482 palavras.

Durante a execução do Estudo Piloto adoto os seguintes procedimentos:

- (i) Rer ler ao capítulo de metodologia da dissertação de Fleuri (2006) e regatar os detalhes dos procedimentos seguidos durante:
 - a. Preparação para o alinhamento;

- b. Execução do Alinhamento;
 - c. Anotação do corpus;
 - d. Revisão das etiquetas;
 - e. Quantificação dos dados.
- (ii) Executar os mesmos procedimentos utilizando o corpus de Fleuri (2006) nos *Template Editor* e *Template Corpus Paralelo*, que são os seguintes:
- a. Preparação do corpus para alinhamento, realizado pelo *Template Editor*;
 - b. Execução do alinhamento, realizado na pasta ‘Alinhamento’ do *Template Corpus Paralelo*;
 - c. Classificação dos objetos analisados, realizada nas pastas ‘Rotulação TF’ e ‘Rotulação TA’ do *Template Corpus Paralelo*;
 - d. Revisão das etiquetas e correção das classificações, realizada na pasta ‘Listas e Gráficos’ do *Template Corpus Paralelo*;
 - e. Obtenção da quantificação dos dados;
- (iii) Comparar cada procedimento, o da dissertação de Fleuri (2006) e do Estudo Piloto, observando o tempo gasto, o número etapas executadas em cada processo, o número de transições entre programas diferentes e a quantidade de documentos gerados ao final de cada processo.

A comparação é descrita detalhadamente no Capítulo 6, do Estudo Piloto, nesta tese.

4 ANÁLISE

"Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa. Por isso, aprendemos sempre."
(FREIRE, 1989)

Neste capítulo são apresentadas, descritas e analisadas as metodologias utilizadas nas vinte dissertações em ESFTBC produzidas no Brasil, dos seus primórdios até 2010. A estrutura do capítulo segue a seguinte configuração: (4.1) são apresentadas as dissertações em ESFTBC produzidas em universidades brasileiras, cada uma tratada em uma subseção especial; (4.2.) descrevem-se os métodos aplicados nessas dissertações, buscando-se um perfil das pesquisas em ESFTBC com o apoio do quadro proposto Barnbrook (1996) e de Olohan (2004), no que tange aos elementos relacionados à organização metodológica que se devem ser considerados em uma pesquisa em corpus, ou seja, o propósito da pesquisa, o desenho do corpus, a compilação do corpus e o processamento (ver seção 2.2 desta tese); (4.3) descreve-se o desenvolvimento da metodologia das pesquisas em ESFTBC, em uma perspectiva diacrônica.

4.1. AS DISSERTAÇÕES EM ESFTBC E SEUS MÉTODOS

A *Lista ESFTBC*, apresentada no Capítulo 3, de metodologia, desta tese, exhibe 24 pesquisas em ESFTBC publicadas no Brasil até 2010, sendo 20 dissertações e 4 teses. Este capítulo se dedica à descrição e análise dessas vinte dissertações, cujas autorias são expostas nas subseções desta seção, em ordem de publicação.

Esses trabalhos apresentam algumas características metodológicas em comum, sobretudo no que tange à Metafunção investigada, ao objeto de estudo, ao tipo de corpus, ao modo como os textos se apresentam, ao tipo de alinhamento, de código de rotulação e de software, às ferramentas utilizadas no *WST*, no caso das dissertações que utilizaram este programa, e à quantificação dos dados. Além disso, as pesquisas apresentam tendências relacionadas ao ano de publicação e ao tipo de universidade a que se afiliam.

Os critérios utilizados para a descrição das dissertações são aqueles sugeridos no Capítulo 2, de Referencial Teórico, seção 2.2 desta tese para pesquisa em corpus com respeito à prática metodológica em corpus, como o desenho (Kennedy, 1998; Olohan, 2004; Fernandes, 2006), a compilação (Olohan, 2004), a anotação (Leech, 2005 e Feitosa, 2005), o processamento e os programas (Anderman & Rogers, 2008; Scott, 2007 e Anthony, 2011) e a forma de exposição dos dados. O texto foi criado a partir das informações coletadas, expostas no Quadro 10, na seção 3.2 desta tese.

Na abertura de cada subseção desta seção, as características e procedimentos metodológicos são apresentados resumidamente em tabela (conforme o Quadro 10) seguido de um comentário final. Os comentários são resgatados e reunidos nas seções 4.2 e 4.3 deste capítulo

4.1.1. Osilene Cruz (2003)

Quadro 8: Resumo das características e procedimentos metodológicos em

Informação	Significado
Informações gerais (ano/universidade) Título	(2003/UFGM). “Harry Potter and the Chamber of Secrets e sua Tradução para o Português do Brasil: uma análise dos verbos de elocução, com base na lingüística sistêmica e nos estudos de corpora”.
Objetivo geral e Metafunção	Metafunção Ideacional: investiga os verbos de elocução e os processos de transitividade em projeções paratáticas.
Objeto de análise	Os verbos de elocução em projeções paratáticas.
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escaneamento; • Separação e contabilização dos parágrafos; • Rotulação de <i>Markups</i>, • Correção do corpus • Corpus alinhado com o auxílio do <i>Viewer and Aligner (WordSmith Tools-1997)</i>

Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> • <i>WordList</i> • Tabela do <i>MS Word</i> • <i>Concord</i> • Contagem é realizada manualmente
Códigos de anotação	Não houve o uso de rótulos

A dissertação de Osilene Cruz (2003) foi a primeira pesquisa no Brasil em ESFTBC registrada no Portal da Capes⁵⁹. Foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Harry Potter and the Chamber of Secrets e sua Tradução para o Português do Brasil: uma análise dos verbos de elocução, com base na linguística sistêmica e nos estudos de corpora”. Tem como objetivo geral

“investigar os verbos de elocução e os processos de transitividade [em projeções paratáticas] que os mesmos configuram, a partir de um corpus bilíngue paralelo, (...), a fim de indagar (...) como se comportam os verbos de elocução e de que forma os mesmos evidenciam padrões de textualização no texto fonte e no texto traduzido” (Cruz, 2003, p. 14).

Desta forma seu objeto de análise é os verbos de elocução em projeções paratáticas, sob a perspectiva do aspecto lógico da Metafunção Ideacional.

Para desenvolver esta pesquisa, a Cruz utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original *Harry Potter and the chamber of secrets* (1999) e a tradução para o português *Harry Potter e a câmara secreta* (2000). Os textos encontravam-se disponíveis na versão impressa, tendo sido compilados pela própria pesquisadora. Para a compilação a investigadora passou pelas etapas de escaneamento; separação e contabilização dos parágrafos; rotulação de *Markups*, “segundo critérios do CORDIAL” (Cruz, 2003), e correção do corpus. O corpus foi alinhado com o auxílio do *Viewer and Aligner*, do programa *WordSmith Tools* (1997).

O processamento do corpus também ocorreu no *WordSmith Tools*. A ferramenta do *WordList* foi utilizada para se identificar os verbos de elocução e poder confirmar se desempenhavam tal papel no discurso direto. Uma vez confirmado, tais verbos foram copiados e colados em uma tabela do *MS Word* e, posteriormente, colocados na sua forma

⁵⁹ Fonte consultada em Julho/2012.

infinitiva. Foi elaborada uma tabela para o texto-fonte e outra para o texto-alvo. Nas parataxes, as categorias de análise dos verbos de elocução e dos elementos circunstanciais foram retiradas do *Collins Cobuild English dictionary for advanced learners* e do *Collins Cobuild English Grammar*. Do mesmo modo, os verbos foram transferidos a tabelas do *MS Word*, e categorizados nestas. Não houve o uso de rótulos para tais categorias. Para se identificar os verbos de elocução e os complementos presentes nas falas dos personagens principais do romance, foram levantadas as ocorrências dos nomes de cada personagem em questão com o auxílio da ferramenta *Concord*. As repetições computadas foram excluídas. Os Processos que realizam os verbos de elocução foram expostos em tabelas, assim como seu valor numérico.

Os dados obtidos foram expostos em tabelas, com o valor total das categorias, no texto-fonte e no texto-alvo. Cada tabela indica o total (e ou o percentual) de um determinado grupo de elementos categorizados em comparação com outros elementos ou do texto-fonte com texto-alvo. A contagem é realizada manualmente.

Observo que a opção de não adotar um sistema de anotação leva a pesquisadora a passar por diversas etapas de busca por elementos textuais e de organização manual destes elementos em tabelas, para poder realizar a contagem das categorias. A utilização do *WordSmith Tools* parece ser, em alguns casos, desnecessária, considerando o baixo retorno que lhe dá em relação ao montante de trabalho manual que a pesquisadora se vê obrigada a realizar. Especula-se que muito do trabalho realizado no *WST* poderia ter sido realizado com ferramentas de “Busca”, por exemplo, do próprio *MS Word*. Observo que a primeira pesquisa que propôs a inserção da Metodologia de Corpusaos ESFT passou por um processo longo e mais manual do que automático de contagem das categorias

4.1.2. Cristina Mauri (2003)

Quadro 9: Resumo das características e procedimentos metodológicos em Mauri (2003)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2003/UFMG), “Um estudo da tradução italiana de <i>Laços de Família</i> , de Clarice Lispector, a partir da abordagem em “ <i>corpora</i> ”: a construção da introspecção feminina através dos verbos de elocução”.

Objetivo geral e Metafunção	Analisar os verbos de elocução e suas colocações no texto de partida e de chegada (Metafunção Ideacional).
Objeto de análise	Os verbos de elocução.
O corpus Tipo de corpus Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escaneamento; • Separação e contabilização dos parágrafos; • Rotulação de <i>Markups</i>, • Correção do corpus • O corpus foi alinhado manualmente
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (2001) • <i>WordList</i> • <i>Concord</i>
Códigos de anotação	Não necessitou ser anotado.

A dissertação de Cristina Mauri (2003) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Um estudo da tradução italiana de *Laços de Família*, de Clarice Lispector, a partir da abordagem em "corpora": a construção da introspecção feminina através dos verbos de elocução”. Tem como objetivo geral analisar os verbos de elocução e suas colocações no texto de partida e de chegada. Assim, seu objeto de análise são os verbos de elocução, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Mauri utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em português *Laços de Família* (de Clarice Lispector) e a tradução para o italiano *Legami Familiari*. Os textos encontravam-se disponíveis na versão impressa, tendo sido compilados pela própria pesquisadora. Para a compilação, a investigadora passou pelas etapas de escaneamento; separação e contabilização dos parágrafos; rotulação de *Markups*, também segundo critérios do CORDIAL, e correção do corpus. O corpus foi alinhado manualmente e não necessitou ser anotado.

O processamento do corpus ocorreu no *WordSmith Tools* (2001). Semelhantemente a Cruz, a ferramenta do *WordList* foi utilizada para se identificar os verbos de elocução. Cada um dos verbos de elocução extraído do *WordList* foi processado no *Concord*. Os verbos que não

introduziam elocuições dos personagens foram excluídos. Os dados quantitativos foram obtidos através do *WordList*, a contagem foi feita manualmente, após transferir os verbos e sua quantidade a uma tabela⁶⁰, não exposta na dissertação.

Observo que o procedimento metodológico adotado por Mauri (2003) é muito semelhante ao de Cruz (2003), e igualmente mais pautada na performance manual que automática de processamento. A opção por não anotar o corpus leva à elaboração de tabelas e de contagem manual. A classificação dos verbos de elocução e a exposição do total de categorias são realizadas ao longo da análise, como uma leitura da tabela, da seguinte forma:

De um total de 394 instâncias no texto de partida, 177 são verbos neutros e 76 indicam processos mentais. Dentre as outras categorias 32 são indicadores de realização do enunciado, 39 indicadores da cronologia discursiva, 39 lexicalizam o modo que caracteriza o dizer, 10 instrumentalizam o que se diz e 21 circunstanciam o que se diz (Mauri, 2003, p.56).

Observo que esse modo de exposição de dados dificulta o acesso a eles, levando futuros pesquisadores a ter que ler todo o texto para resgatar os dados quantitativos da análise. A exposição dos dados quantitativos em tabelas facilita a visualização e a associação destes durante a análise.

4.1.3. Silvana de Jesus (2004)

Quadro 10: Resumo das características e procedimentos metodológicos em Jesus (2004)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2004/UFGM), “Representação do Discurso e Tradução: padrões de textualização em corpora paralelo e comparável”.

⁶⁰ Observou-se, nas dissertações analisadas, que os pesquisadores que não anotam corpus, organizavam em tabelas os elementos e os nomes dados a estes. A essa forma de organizar a análise para uma contagem manual em tabela, chamamos de “tabulação”.

Objetivo geral e Metafunção	Descrever e comparar os verbos de elocução <i>think</i> e pensar, enfocando nos padrões léxico-gramaticais e discursivos nas obras. (Metafunção Ideacional).
Objeto de análise	Verbos de elocução <i>think</i> e pensar
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão e comparável monolíngue de pequena dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escaneamento • Revisão; • Conversão para <i>Plain Text</i> • O corpus foi alinhado com auxílio do utilitário do <i>WST, Viewer and Aligner</i>.
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (Scott, 2001) • <i>WordList</i> • <i>Concord</i> • <i>Viewer and Aligner</i>
Códigos de anotação	O corpus foi anotado com códigos criados pela própria pesquisadora.

A dissertação de Silvana de Jesus (2004) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Representação do Discurso e Tradução: padrões de textualização em corpora paralelo e comparável”. Tem como objetivo geral descrever e comparar os verbos de elocução *think* e pensar no corpus adotado, focalizando padrões léxico-gramaticais e discursivos nas obras. Assim, seu objeto de análise são os verbos de elocução *think* e pensar, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Jesus utiliza um corpus paralelo bilíngues de pequena dimensão e comparável monolíngue de pequena dimensão, compostos pelos textos *Counter Point* (de Aldous Huxley), *Contraponto* (Trad. Érico Veríssimo) e *Caminhos cruzados* (de Érico Veríssimo), os quais se encontravam na versão impressa e foram compilados pela própria pesquisadora. A compilação consistiu em realizar escaneamento *OCR* para *Word* (Software OmniPage que contém o recurso de correção do texto *OCR*); revisão; conversão para *Plain Text*.

O corpus paralelo foi alinhado com auxílio do utilitário do *WST, Viewer and Aligner*. O corpus foi anotado com códigos criados pela própria pesquisadora. São oito etiquetas criadas que são posicionadas depois do objeto analisado, representadas das seguintes formas (Jesus, 2004, p.65):

Quadro 11: Códigos de anotação criados utilizados por Jesus (2004)

<think=pensar>	Re-textualização de processo mental realizado por THINK no original através de processo mental realizado por PENSAR no texto traduzido
<think=0>	Re-textualização de processo mental realizado por THINK no original através de verbalização sem processo mental ou com outros tipos de processo no texto traduzido
<0=pensar>	Re-textualização de passagem sem processo mental no original através de processo mental realizado por PENSAR no texto traduzido
<DVTO=pensar>	Re-textualização de processo mental realizado por verbo diferente de THINK no original através de processo mental realizado por PENSAR no texto traduzido
<DNTO=pensar>	Re-textualização de processo mental realizado por substantivo diferente de THINK no original através de processo mental realizado por PENSAR no texto traduzido
<think=DNTR>	Re-textualização de processo mental realizado por THINK no original através de processo mental realizado por substantivo diferente de THINK no texto traduzido
<think=DVTTR>	Re-textualização de processo mental realizado por THINK no original através de processo mental realizado por verbo diferente de PENSAR no texto traduzido
<think=dizer>	Re-textualização de processo mental realizado por THINK no original através de processo verbal realizado por DIZER no texto traduzido

O processamento do corpus ocorreu no *WordSmith Tools* (Scott, 2001), cujas ferramentas utilizadas foram o *WordList*, *Concord* e *Viewer and Aligner*. O *WordList* foi utilizado para se destacar a quantidade de processos mentais (*Think* e *Pensar*): a partir desta ferramenta “confirmou-se a relevância da investigação de THINK e

PENSAR no corpus”. O *Concord* foi utilizado para se buscar a concordância desses Processos e de suas etiquetas e para quantificá-las.

Na pesquisa de Jesus (2004) já se observa a utilização de um código de rotulação criado pela própria pesquisadora para cumprir os fins da pesquisa. Na fase de obtenção dos dados quantitativos não há menção da utilização de tabelas, nem da contagem manual, e sim do processamento automático pelo *Concord*. A rotulação, além de possibilitar o retorno aos dados, tornando a pesquisa mais acessível, ela evita o trânsito incontável entre os programas Microsoft *Word* ↔ *WordSmith Tools*, em uma série de comandos de *seleciona, copia e cola*, e incontáveis edições. Quanto ao código elaborado pela pesquisadora, parece servir muito bem para os propósitos da pesquisa, mas em termos de praticidade de digitação, apresenta muitos caracteres (entre letras minúsculas, símbolos, números e letras maiúsculas), dificultando o momento de inserção do código no texto e aumentando as chances de ocorrer erros de digitação, que podem acabar resultando em erros de contagem dos dados (cf. Feitosa, 2005, p.14). Outro fato que pode levar a erros de digitação é a semelhança entre as etiquetas, como <DVTO=pensar> e <DNTO=pensar>, em que um V trocado por um N (teclas muito próximas no keyboard) muda todo o resultado da pesquisa. Segundo Feitosa (idem, ibidem),

as primeiras análises realizadas pela Dra. Adriana Pagano e pelos pesquisadores membros do NET (Núcleo de Estudos da Tradução/ UFMG) mostraram que, além de se tratar de um trabalho demorado, o processo de anotação dos corpora vinha apresentando falhas devido a erros de digitação na hora da inserção de Rótulos. A anotação vinha sendo efetuada através de palavras, com o nome de cada categoria inserido por extenso em cada Rótulo. Esses Rótulos acabavam sendo longos e mais propensos a erros de digitação, o que interferia nos resultados fornecidos pelo programa de análise de corpora. Observou-se que a ausência de um padrão de anotação entre os membros do NET dificultava um diálogo mais profícuo entre os pesquisadores. Como cada pesquisador estabelecia seus Rótulos, o acesso ao corpus anotado de um outro pesquisador tornava-se mais difícil, pois implicava a necessidade de se aprender como foi efetuada a Rotulação de seu corpus.

A explicação dos códigos, disposta em formato de tabela, e os exemplos de colocação do código na frase são boas soluções, criadas por

Jesus (2004), para facilitar a leitura destes códigos por outros pesquisadores ou até mesmo sua reutilização em pesquisas futuras sobre o mesmo objeto e tema.

4.1.4. Roberto Assis (2004)

Quadro 12: Resumo das características e procedimentos metodológicos em Assis (2004)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2004/UFMG), “A Transitividade na Representação de <i>Sethe</i> no Corpus Paralelo <i>Beloved-Amada</i> ”.
Objetivo geral e Metafunção	Analisar a representação ideacional da personagem <i>Sethe</i> . (Metafunção Ideacional)
Objeto de análise	O item lexical <i>Sethe</i> e suas correspondentes traduções, Participante de Processo, no papel de Ator, Experienciador, Dizente, Portador e Comportante.
O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Textos já digitalizados na versão <i>Word Document</i>
Método de compilação	<ul style="list-style-type: none"> • Correção do corpus • Busca das ocorrências de <i>Sethe</i> • Separação dos documentos gerados • Corpus alinhado com o auxílio do <i>Viewer and Aligner (WordSmith Tools-1997)</i>
Método de processamento	<ul style="list-style-type: none"> • <i>WordList</i> • <i>WordList</i> • <i>Concord</i> • <i>MS Word</i> • Contagem se deu manualmente
Códigos de anotação	Cada oração foi etiquetada com o nome do Processo.

A dissertação de Roberto Assis (2004) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “A Transitividade na Representação de *Sethe* no Corpus Paralelo *Beloved-*

Amada”. Tem como objetivo geral analisar a representação ideacional da personagem, *Sethe*, no corpus adotado. O objeto de análise é o item lexical *Sethe* e suas correspondentes traduções, Participante de Processo, no papel de Ator, Experienciador, Dizente, Portador e Comportante, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Assis utiliza um corpus paralelo de pequena dimensão composto pelos textos *Beloved* (de T. Morrison) e *Amada* (Trad. Evelyn Kay Massaro). Os textos se encontravam já digitalizados na versão *Word Document*, mas não com os problemas de escaneamento corrigidos. Portanto, foi necessário corrigir problemas de escaneamento, tais como, palavras *eat* escaneadas como *cat*. Problemas como esses foram encontrados com o auxílio da ferramenta *WordList* utilizada no texto-original e no traduzido, conforme explicitado no Capítulo 3 – Metodologia. Durante a fase exploratória, “a busca pelas ocorrências de “*Sethe*” foi feita inicialmente através do *WordList* nos dois textos em separado” (idem, p.68). O corpus foi alinhado com auxílio do *Viewer and Aligner*, do *WST* (Scott, 2001).

No processo de alinhamento o pesquisador menciona as tarefas manuais do processo semiautomático de alinhamento, tais como: abreviações, aglutinações ou omissões. Em casos em que uma frase era traduzida em duas, três frases traduzidas em uma, frases não traduzidas, ou frases adicionadas na tradução, Assis optou por colocar o símbolo [] para as traduções omitidas e [][] para as acrescentadas, resolvendo os casos de desalinhamento causados por estes fatores. Após alinhar o corpus, o autor transformou os dois textos alinhados em um único documento (contendo a Textualização e a Retextualização), chamado de *Corpus Beloved-Amada*. A análise foi realizada sobre o par *Sethe* e sua tradução ou texto original e *Sethe*. Foram criados três documentos para salvar as unidades de alinhamento: um em que *Sethe* foi traduzida como *ela*, outro em que *she* foi traduzida como *Sethe*, e o terceiro documento em que *Sethe* foi traduzida como *Sethe*. Estes documentos foram nomeados da seguinte forma: *Setheela*, *sheSethe* e *SetheSethe*, respectivamente.

Neste corpus, realizou-se a concordância pelo *Concord* da palavra *Sethe*. As linhas de concordância foram expandidas e as “unidades de alinhamento”⁶¹ foram separadas coladas em documentos *Word*. Cada

⁶¹ “O termo Unidade de Alinhamento baseado no conceito de Unidade de Tradução proposto em (ALVES, 2000) como “um segmento do texto de partida, independente de tamanho e forma específicos, para o qual, em um dado momento, se dirige o foco de atenção do tradutor”” (Assis, 2004, p.68)

linha de concordância e sua respectiva tradução foram etiquetadas, com rótulos criados pelo próprio pesquisador. As etiquetas vinham posicionadas no início da linha de concordância. As etiquetas eram diferentes uma da outra, levando cada uma o número correspondente à ordem de ocorrência e o nome do documento em que se encontrava. Compreendo, através dos exemplos, que a classificação dos processos é adicionada nesta etiqueta. Para Processos Materiais (“mat”), por exemplo, a etiqueta pode ser representada da seguinte forma:

248mat632SetheSethe774tudo *Sethe estava apontando um bico sangrento...*

67mat149SetheSethe187tudo *that Sethe had reached the point beyond*

68mat150SetheSethe188tudo ... *que Sethe atingira um ponto além ...*

Dado modo em que o pesquisador expõe os exemplos e as etiquetas, entendo que cada “unidade de análise”⁶² é quebrada em oração, sendo que cada oração apresenta um Processo relacionado à Participante *Sethe*. A cada oração é inserida uma etiqueta, no início da oração, com a abreviação do Processo correspondente, conforme exemplificado anteriormente. Excluíram-se as ocorrências em que *Sethe* não representava um Ator, Experienciador, Dizente, Comportante ou Portador, com o objetivo de limitar a quantidade de unidades de análise (Assis, 2004, p.74). Entendo também que as linhas de concordância foram quebradas em orações em cujos Processos *Sethe* realizava o papel de Participante, e cada oração etiquetada com o nome do Processo, conforme demonstrado anteriormente e que tenha sido feita a contagem manual.

Observo que na dissertação de Assis (2004) há, em relação às outras, uma tentativa de inovar a forma de rotular e de organizar seu corpus. Quanto à organização do corpus, primeiramente os textos são digitalizados separadamente e salvos em arquivos *Plain Text*, uma para o *Corpus Beloved* e outro para o *Corpus Amada*. Esses textos são alinhados, gerando outro arquivo chamado de *Corpus Beloved-Amada*. Esse corpus é processado no *Concord*, e a busca pelo nóculo *Sethe* é realizada. As linhas de concordâncias são expandidas e são identificadas as unidades de alinhamento. Todas as unidades de alinhamento em que não há a presença de pelo menos um nóculo *Sethe* são eliminadas. O resultado gera outro documento chamado de *Corpus Sethe*. Desse documento são

⁶² “Unidade de Análise” corresponde “ao conjunto “linhas de concordância + linhas de concordância expandidas + Unidades de Alinhamento inseridas nos arquivos *SetheEla* e *SheSethe*” (Assis, 2004, p.73).

separadas, em três outros documentos, as unidades de alinhamento em que *Sethe* é traduzida como *Sethe* (chamado de *SetheSethe*), outro em que *she* é traduzida como *Sethe* (*SheSethe*) e outro em que *Sethe* é traduzida como ela (*Setheela*). A partir de então, Compreendo que cada um desses três documentos tem suas linhas quebradas em orações e cada oração etiquetada conforme a categoria do Processo, no início da linha. Houve a geração de sete documentos diferentes sobre um mesmo corpus.

Do código de rotulação (exemplo: 141mat323SetheSethe383tudo) Observo um grande carga de informação (número da linha, nome do documento, tipo de retextualização e Processo), útil na revisão e resgate da informação, mas ainda não ideal como um código de padronização, de acordo com Feitosa. Entendo que apesar de anotado o corpus, os dados foram contados manualmente. Verifico que, apesar de o corpus já ter se encontrado inicialmente digitalizado o trabalho manual, ainda é grande em relação ao processamento automático proposto pela metodologia da Linguística de Corpus.

4.1.5. Letícia Bueno (2005)

Quadro 13: Resumo das características e procedimentos metodológicos em Bueno (2005)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2005/UFMG), “Transitividade, coesão e criatividade lexical no corpus paralelo Macunaíma, de Andrade e Macunaíma, de Goodland”
Objetivo geral e Metafunção	Analisar a representação ideacional das construções criativas (Metafunção Ideacional).
Objeto de análise	Os verbos de elocução em projeções paratáticas.
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Textos já digitalizados em formato <i>Word Documents</i>
Método de compilação	<ul style="list-style-type: none"> • Processo manual de correção • Conversão para o formato de <i>Plain Text</i>, • Alinhamento manual no <i>Word</i>

Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (2001) • <i>WordList</i> • <i>Keyword</i> • <i>MS Word</i> • Contagem é realizada manualmente
Códigos de anotação	Não necessitou ser anotado.

A dissertação de Letícia Bueno (2005) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Transitividade, coesão e criatividade lexical no corpus paralelo Macunaíma, de Andrade e Macunaíma, de Goodland”. Tem como objetivo geral analisar a representação ideacional das construções criativas e como seu objeto de análise a “criatividade lexical” no corpus adotado. Centra-se na Metafunção Ideacional.

Para desenvolver esta pesquisa, Bueno utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em português chamado “Macunaíma” (de Mário de Andrade) e sua tradução para o inglês “Goodland”. O corpus já se encontrava digitalizado em formato *Word Documents*. O corpus passou por um processo manual de correção e foi convertido para o formato de *Plain Text*, para ser processado no *WordSmith Tools* (2001). Não houve a necessidade de anotar o corpus, pois a contagem foi realizada manualmente.

Para o processamento do corpus foram utilizadas as ferramentas de *WordList*, *Keyword* e *Concord*. Bueno buscou os léxicos a serem estudados no *WordList*, Lacio-Ref, *British National Corpus*, *Google*, *Altavista* e dicionários. Através do “*Hapax*” no *WordList* selecionou as palavras criativas e eliminou manualmente as não criativas. Montando uma lista de 60 processos criativos. A verificação da criatividade se deu por procuras em diferentes fontes (citadas acima). Chegou-se a um total de 13 Processos, com mais um refinamento, sobraram nove. No *Concord*, geraram-se as linhas de concordância e se realizou o alinhamento manual no *MS Word*. Através do *Keyword* analisou o emprego de verbos usuais do português e analisou a criatividade de emprego semântico dessas palavras (como brincar, rir e festejar). Não foram gerados dados quantitativos, já que restaram apenas nove Processos para serem analisados. Como a própria pesquisadora aponta, os recursos de *WST* não foi muito explorado, por uma opção metodológica. É interessante observar como Bueno escolhe realizar o alinhamento manual no *Word*, uma vez que a ferramenta de alinhamento

do *WST* exige uma intervenção prévia do pesquisador, que resulta um trabalho igual ou maior que o alinhamento manual.

Bueno utiliza mais os recursos do *WST* na fase exploratória, principalmente as ferramentas *Keyword* e *WordList*, que, mesclados ao uso de corpus de referência, serviram-lhe para filtrar os Processos “criativos” em Macunaíma. Essa foi a primeira dissertação que não realiza uma análise quantitativa, centrando-se apenas uma análise qualitativa dos Processos encontrados.

4.1.6. Roberta Rodrigues (2005)

Quadro 14: Resumo das características e procedimentos metodológicos em Rodrigues (2005)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2005/UFMG), “A Organização Temática em A Hora da Estrela e The Hour of The Star”
Objetivo geral e Metafunção	Descrever e comparar os Temas e a organização temática no corpus adotado, focalizando-se os padrões léxico-gramaticais, semânticos e discursivos das mesmas. (Metafunção Textual).
Objeto de análise	Todos os Temas e Remas.
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Textos na versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Digitalização dos textos, • Textos corrigidos e convertidos ao formato <i>Plain Text</i>. • Alinhamento manual • Anota o corpus manualmente
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (2001) • <i>WordList</i> • <i>Keyword</i> • <i>Concord</i> • <i>MS Word</i> • Etiquetas quantificadas no <i>Concord</i>

Códigos de anotação	Anota o corpus com rótulos criados pela própria pesquisadora.
----------------------------	---

A dissertação de Roberta Rodrigues (2005) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “A Organização Temática em A Hora da Estrela e The Hour of The Star”. Tem como objetivo geral descrever e comparar os Temas e a organização temática no corpus adotado,

Focalizando-se os padrões léxico-gramaticais, semânticos e discursivos das mesmas, e verificando-se também em que medida a descrição e a comparação dos Temas e da organização temática no corpus pode apontar o estilo de escrita de Clarice Lispector e do tradutor Giovanni Pontiero, no que diz respeito ao corpus analisado nesta dissertação (Rodrigues, 2005).

Seu objeto de análise são todos os Temas e Remas, sob a perspectiva da Metafunção Textual.

Para desenvolver esta pesquisa, Rodrigues utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em português, “A hora da estrela” (de Clarisse Lispector), e sua tradução para o inglês “The hour of the star” (Traduzido por Giovanni Pontiero). Os textos encontravam-se inicialmente disponível à pesquisadora em formato impresso, tendo sido compilado pela própria pesquisadora: digitalizados, corrigidos e convertidos ao formato *Plain Text*. A autora não menciona o alinhamento do corpus, apesar de os exemplos na análise serem interpretados alinhadamente e se tratar de um corpus paralelo. Por não mencionar a ferramenta *Viewer and Aligner*, Compreendo que o alinhamento tenha se dado manualmente.

Com o corpus em formato *Plain Text*, Bueno anota o corpus manualmente, com uma etiqueta criada por ela mesma para cumprir as necessidades de sua pesquisa, posicionando-os antes do objeto de análise. As etiquetas se encontravam entre parênteses angulares onde o tipo de tema e rema eram explicitados, da seguinte forma: <tema textual>; <tema ideacional participante>. Este caráter autoexplicativo das etiquetas levou a não disponibilização de uma tabela para explicitá-las. O símbolo Ø foi utilizado para indicar elipses e quando há Absoluto, o rótulo é usado. Foram classificados todos os Temas nos textos fonte e alvo. Optou-se por se classificar os Temas de orações independentes ou principais em casos de parataxe e por não se classificar os Temas das orações em casos de

hipotaxe e encaixadas houve o cuidado de se classificar os Temas até o primeiro constituinte ideacional.

O processamento do corpus ocorreu no *WordSmith Tools* (2001), a partir das ferramentas *WordList*, *Keyword*, e *Concord*. As etiquetas foram processadas no *Concord*, para sua quantificação. Os trechos que não continham etiquetas eram eliminados (através dos comandos no *WST Delete e Zap*). Todos os trechos etiquetados foram copiados e colados no *Word*. Os Temas interpessoais foram classificados e os personagens aos quais cada linha se relacionava foram identificados. Foram rastreados todos os tipos de processos a fim de se classificar os tipos de participantes que ocorreram com o narrador e a protagonista Macabéa. Compreendo que, como a busca dos Temas e dos Processos foi realizada no *Word*, o modo como esses novos itens foram contados tenha sido manual.

Observo nesta dissertação ainda uma grande carga de dedicação ao trabalho manual de preparação do corpus.

4.1.7. Eliza Morinaka (2005)

Quadro 15: Resumo das características e procedimentos metodológicos em Morinaka (2005)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2005/UFSC), “ <i>Gabriela</i> , cravo e canela and the English (re)textualization: representation through lexical relations”
Objetivo geral e Metafunção	<ul style="list-style-type: none"> • Descrever os diferentes tipos de relações coesivas lexicais usadas para construir a identidade da protagonista <i>Gabriela</i>; • Investigar se as mudanças nos padrões coesivos emergentes resultam em uma diferente representação da protagonista na (re)textualização.
Objeto de análise	O Participante <i>Gabriela</i> e as relações de coesão lexical da personagem, realizando Participante de Processo e de Circunstância. (Metafunção Ideacional e Textual).
O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.

Método de compilação	<ul style="list-style-type: none"> • Escaneamento, • Correção, • Edição, • Conversão para o formato <i>Plain Text</i>, • Adição de marcações, de acordo com o CORDIALL • Alinhamento no <i>Viewer and Aligner</i>, no <i>WordSmith Tools</i>, • Anotação o corpus manualmente.
Método de processamento	<ul style="list-style-type: none"> • <i>WordList</i> • <i>Keyword</i> • <i>Viewer and Aligner</i> • <i>Concord</i> • Contagem é realizada no <i>Concord</i>
Códigos de anotação	Anota o corpus com o CROSF11 (Feitosa, 2005)

A dissertação de Eliza Morinaka (2005) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “*Gabriela, cravo e canela and the english (re)textualization: representation through lexical relations*”. Tem como objetivo geral descrever os diferentes tipos de relações coesivas lexicais usadas para construir a identidade da protagonista *Gabriela* no corpus em questão; investigar se as mudanças nos padrões coesivos emergentes resultam em uma diferente representação da protagonista na (re)textualização. Assim, seu objeto de análise é o Participante *Gabriela* e as relações de coesão lexical da personagem, realizando Participante de Processo e de Circunstância, sob a perspectiva da Metafunção Ideacional e Textual.

Para desenvolver esta pesquisa, Morinaka utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em português *Gabriela, Cravo e Canela* (de Jorge Amado) e a tradução para o inglês *Gabriela, clove and cinnamon* (Trad. J.Taylor e W. Grossman). Encontrava-se inicialmente disponível à pesquisadora na versão impressa, sendo compilado por ela própria, passando pelas etapas de escaneamento, correção, edição, conversão para o formato *Plain Text*, adicionar mark-ups, de acordo com o CORDIALL e descrito em Souza e Morinaka (2004), em seguida seleciona manualmente as sentenças em que *Gabriela* era mencionada e alinhar, no *WordSmith Tools*, a Textualização com a Retextualização dessas sentenças. Por fim, anota o corpus com o

CROSF11⁶³ (Feitosa, 2005), posicionado o código antes dos nódulos Participantes de Processo e de Circunstância.

Em relação à anotação, Morinaka anotou os elementos lexicais *Gabriela*, e palavras que substituem esse léxico; não etiquetou pronomes e elipses destes termos. Ao CROSF foi adicionado o código “abc” ocupando os três últimos algarismos para identificar a palavra ou a referência à “*Gabriela*” não Participante de Processo nem de Circunstância e a palavra *minor* no lugar dos três últimos algarismos para indicar *minor clauses*. Etiquetou conforme os seguintes exemplos: *T. <0010abc>Gabriela no caminho / RT. <0010abc>Gabriela on the way, T. -- Mas, <0010minor>Gabriela. / RT. "But, <0010minor> Gabriela., T. A perna de <0010151>Gabriela pendia da cama, ela sorria no sono. / RT. The door to <0010500>Gabriela's little back-yard room was open.* Morinaka mantém em todas as etiquetas os quatro primeiro dígitos fixados <0010, indicando que o tema e rema não são analisados (dois primeiros dígitos), a Metafunção Ideacional é considerada (terceiro dígito), e sua ausência ou presença de marcação é desconsiderada para fins de análise (quarto dígito). Os três últimos variam entre letras e números. As letras não são consideradas por Feitosa e não são descritas no CROSF; estas foram adicionadas pela própria pesquisadora, como uma variação do CROSF, respondendo às necessidades de sua pesquisa. Os números, entretanto, se enquadram na descrição de Feitosa. Morinaka opta por apontar todos os Participantes, sem importar se sofrem ou não interpolação, com o número 1, indicando que todos os Participantes não sofrem interpolação (segundo a descrição desta posição por Feitosa), e o número 5 os que são Participantes de Circunstância. Os dois últimos dígitos indicam o tipo de Processo (o sexto dígito) e o tipo de Participante (o sétimo dígito).

O processamento ocorreu no WST (2001) através das ferramentas de *WordList*, *Keyword*, *Viewer and Aligner*, e *Concord*. Esta etapa consistiu em buscar os padrões e ocorrências entre a Textualização (T) e a Retextualização (RT) a partir do *WordList*. Conferiu-se se as palavras relacionavam-se a *Gabriela*, através do *Concord*, nas linhas de concordância. Testou-se a *keyness* das palavras relacionadas a partes do corpo da protagonista *Gabriela* nos seu corpus, com o auxílio da ferramenta *Keyword*, e em no corpus de referência Lacio-Web, justificando a relevância (numérica) de sua pesquisa. Para quantificar os

⁶³ O CROSF é explicado na seção 2.2.1.3 *Anotação de corpus* do capítulo 2 de Referencial Teórico desta tese.

dados, Morinaka gerou concordância das etiquetas no *Concord* e as contabilizou. A autora menciona que os passos seguidos encontram-se explícitos em Alves e Morinaka (2004), desta forma, não os descreve em sua dissertação.

Noto que esta é a primeira pesquisa que de fato utiliza todas as ferramentas disponíveis no *WordSmith Tools: WordList* e *Keyword* na fase exploratória; e o *Viewer and Aligner*. Outro fato interessante nessa dissertação, no ponto de vista metodológico, é que é a primeira também a organizar a descrição dos procedimentos metodológicos em dois estágios principais: um de construção do corpus, em que aponta os passos para transformar o texto impresso e corpus digitalizado, pronto para ser processado no *WST*; e outro de processamento do corpus, em que explica as ferramentas do *WST* utilizadas e detalha as etapas em que tais ferramentas são utilizadas. O estágio de processamento de corpus, é separado em duas fases: fase exploratória e fase focada. Na primeira, Morinaka investiga se o objeto escolhido apresenta, em termos de número de ocorrência, uma relevância da pesquisa. Nesta fase também se observa elementos curiosos, dignos de serem investigados em mais profundidade. Nessa fase são usadas as ferramentas *WordList* e *Keyword*, além de consultas à corpus de referência (corpus de grande dimensão), no seu caso utilizou o Lácio-Web⁶⁴. A segunda fase compreende a identificação de cada elemento a ser etiquetado, a análise sistêmico-funcional destes elementos e a inserção das etiquetas.

Observo uma constante menção ao trabalho desenvolvido por Alves e Morinaka (2004), que representa o primeiro trabalho (ainda que não publicado) de cunho metodológico em corpus direcionado aos trabalhos que vinham sendo desenvolvidos no projeto CORDIAL, em ESFTBC. Este trabalho descreve o funcionamento de leitura do texto pelas ferramenta do *WST* baseando-se nas dificuldades expressas por pesquisadores do NET. Segundo Morinaka (2004) este trabalho deveria se encontrar disponível no site web [http://www.geocities.com/xalaskero/UFMG/Metodologia/MetodologiaG](http://www.geocities.com/xalaskero/UFMG/Metodologia/MetodologiaGeral)
[eral](http://www.geocities.com/xalaskero/UFMG/Metodologia/MetodologiaG), mas ao acessá-lo atualmente (junho de 2012) este se encontra desativado. Ao contatar Daniel Alves Souza (um dos autores do trabalho) ele me respondeu o seguinte: “O trabalho de Alves e Morinaka (2004) ficou disponível no endereço <http://goo.gl/q0Aac> até 26 de agosto de 2009, quando o servidor do site que o hospedava suspendeu suas

⁶⁴ website Lácio Web (<http://www.nilc.icmc.usp.br/lacioweb>)

operações indefinidamente”⁶⁵, sendo agora acessível apenas através do contato direto com os próprios autores.

Quanto a sua forma de etiquetar, chamou-me atenção dois fatos: a sua necessidade de adaptação do CROSF e a forma de etiquetar o quinto dígito deste código. Morinaka adapta-o ao substituir os três últimos dígitos pelas letras “abc” ou pela palavra *minor*, indicando ocorrências não previstas no CROSF. Quanto ao 5º dígito do CROSF, Feitosa explica que na quinta casa o número “1” representa Participante sem interpolação; 2, Participante com interpolação; “3”, Processo; “4”, Processo-Participante; “5”, Circunstância sem interpolação; “6”, Circunstância com interpolação; “7”, “preposed attributive”; “8”, Oração e “9”, estrutura tematizada. Morinaka opta por adotar o número 1 na quinta casa representando a um Participante], apresentando ou não interpolação, ou o número “5” para indicar que *Gabriela* se encontra inserida em uma circunstância, sem explorar sua interpolação. Essa questão é observada em outros pesquisadores que utilizam o código para analisar a Metafunção Ideacional e sua forma de etiquetar o quinto dígito é enfatizada na descrição do processo de rotulação.

4.1.8. Tassiani Cançado (2005)

Quadro 16: Resumo das características e procedimentos metodológicos em Cançado(2005)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2005/UFMG), “Transitividade e Representação do Discurso no Corpus Paralelo Interview with the Vampire/ Entrevista com o Vampiro”
Objetivo geral e Metafunção	Pesquisar as escolhas feitas na retextualização dos verbos de elocução e processos de transitividade, introdutores dos discursos do entrevistador e do entrevistado. (Metafunção Ideacional)
Objeto de análise	Verbos de elocução e processos de transitividade.
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.

⁶⁵ E-mail recebido no dia 29 de maio de 2012.

Método de compilação	<ul style="list-style-type: none"> • Digitalização dos textos; • Correção dos mesmos; • Marcação conforme os critérios do CORDIALL • Converter o corpus em <i>Plain Text</i> • Arrumá-lo para o alinhamento • Alinha o corpus manualmente
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> • <i>WordList</i> • <i>Viewer and Aligner</i> • Tabela do <i>MS Word</i> • <i>Concord</i> • Contagem dos verbos de elocução é realizada no <i>WordList</i>
Códigos de anotação	Anota os interlocutores não os verbos de elocução.

A dissertação de Tassiani Cançado (2005) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Transitividade e Representação do Discurso no Corpus Paralelo Interview with the Vampire/ Entrevista com o Vampiro”. Tem como objetivo geral pesquisar as escolhas feitas na retextualização dos verbos de elocução e processos de transitividade, introdutórios dos discursos do entrevistador e do entrevistado, na representação do discurso da entrevista corpus em questão. Assim, seu objeto de análise são os verbos de elocução e processos de transitividade, dentro da perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Cançado utiliza um corpus paralelo bilíngue de pequena dimensão: *Interview with the vampire* (de Anne Rice) e a tradução *Entrevista com o vampiro* (Trad. Clarisse Linspector). Tais textos encontravam-se inicialmente disponíveis na versão impressa, tendo sido compilado pela própria pesquisadora. A compilação consistiu em realizar a digitalização dos textos e correção dos mesmos, a marcação (adição de mark-ups⁶⁶) conforme os critérios do CORDIALL, converter o corpus em *Plain Text* e arrumá-lo para o

⁶⁶ A diferenciação entre Marcação (markups) e Anotação é realizada no capítulo 2 de Referencial Teórico desta tese, na seção 2.2.1.2 *Compilação de Corpus*, a partir da explicação de Olohan (2004, p.48-56).

alinhamento. O alinhamento foi realizado manualmente (apesar de mencionar o uso do *Viewer and Aligner*, a autora não detalha esse procedimento, explicitando apenas o alinhamento manual). Inseriu-se no alinhamento os símbolo [#] para indicar elipses na tradução (ou seja, os casos de redução).

Entendo que não anotou os verbos de elocução, apenas os interlocutores. Apesar de não mencionar haver anotado os verbos de elocução ou os Processos, no capítulo de análise encontra-se a descrição quantitativa dos verbos de elocução, classificados de acordo com (THOMPSON, 2004), e as seguintes abreviações representando cada tipo de verbo (Cançado, 2005, p.91):

VN = verbos neutros;

PF = verbos que indicam o propósito do falante;

MD = verbos que mostram o modo do dizer;

VD = verbos que indicam o que foi dito;

AM = verbos que indicam como a mensagem se ajusta com o resto da oração;

VCI = verbos que podem indicar uma crença ou ideologia; e

VNE = verbos não considerados em categoria de elocução por Thompson.

Observo apenas a anotação dos interlocutores do corpus, sendo etiquetado com o nome do entrevistado em primeira pessoa <Louis>, em terceira pessoa <vampire> e do entrevistador <BOY>, posicionada antes do verbo de elocução.

O processamento do corpus ocorreu no *WordSmith Tools* (Scott, 1997), através das ferramentas do *WordList* e *Concord*. O levantamento dos Processos foi realizado no *WordList*. O *Concord* foi utilizado para buscar por aspas, marcadores de discurso direto, localizando os verbos à esquerda; ao radical de cada verbo. No *Concord*, buscaram-se também todos os verbos da *WordList* (aos verbos regulares foi colocado um asterisco no fim da forma finita, por exemplo, <cant*>, para englobar todas suas conjugações). Os verbos de elocução são quantificados no *WordList*. Recorreu-se ao *Concord* para analisar esses verbos de elocução na fala de cada personagem.

Não fica claro se os dados coletados no *Concord* e *Viewer and Aligner* são frutos de uma rotulação ou de outra forma de classificação⁶⁷.

⁶⁷ “Os dados quantitativos para cada categoria de verbos de elocução, começando pelos dados estatísticos fornecidos pela ferramenta *WordList*, e complementando com os dados oferecidos pelas ferramentas *Concord* e *Viewer and Aligner*, a fim de reunir todos os argumentos relevantes para uma análise

A partir das tabelas que Cançado (2005) expõe no seu Capítulo de Análise, especulo que o processamento da quantificação dos dados é gerado a partir de uma lista de verbos de elocução oferecida por Thompson (1994) e esses verbos são buscados e quantificados no *WordList* (i.e. Cançado, 2005, p. 82). A tabela com os dados quantitativos que Cançado expõe no Capítulo de Análise é a seguinte:

Quadro 17: Verbos de elocução encontrados no *WordList* (Cançado, 2005)

Lista dos verbos de elocução em projeção paratática em *Interview with the vampire*

Verbo de Elocução	Ocorrências	Verbo de Elocução	Ocorrências
Added	1	Nodded	5
Answered	19	Offered	1
Asked	98	Pleaded	4
Assured	3	Pointed	2
Began	5	Protested	3
Begged	2	Raised	1
Blurted	1	Repeated	4
Burst	2	Replied	1
Called	3	Rose	2
Came	1	Said, Say, Saying	317
Cautioned	1	Sat	3
Confessed	1	Shook	8
Cried	1	Shouted, Shouting	7
Cursing	1	Shrugged	2
Demanded	5	Sighed	7
Draw	1	Smiled	10
Explained	2	Speak	1
Felt	2	Stammered	2
Fought	1	Stared	1
Gasped	3	Started	2
Gestured	1	Suggested	1
Groaned	1	Thinking, Thought	8
Grounded	1	Told	6
Guessed	1	Threw	1
Insisted, Insisting	5	Turned	2
Knelt	1	Urged	2
Laugh	1	Uttered	1
Looked	5	Went	7
Murmur, Mumbled	7	Whisper, Whispered	51
Mused	1		

Em seguida, os verbos de elocução são classificados e quantificados segundo sua projeção (paratática ou hipotática). Esse processo de quantificação é realizado a partir da busca das etiquetas dos nomes dos personagens no *Concord*, a partir da classificação baseada em Thompson (1994). O Quadro com os dados quantitativos que Cançado expõe no Capítulo de Análise é o seguinte:

discursiva e baseada nas teorias abordadas nesta dissertação” (Cançado, 2005, p.81).

Quadro 18: Quadro de análise em Cançado (2005)

Representação da entrevista na narrativa em 1ª pessoa, nas quatro partes do livro

Tipos de verbos	Textualização (TE)				Retextualização (RT)			
	1a	2ª	3a	4a	1a	2a	3a	4a
Neutros	66,9%	55,5%	61,8%	78,6%	60,6%	46,1%	42,6%	55,9%
Propósito do falante	5,8%	11,1%	3,8%	0	4,5%	7,7%	8,7%	0
Modo de dizer	20,5%	24,1%	16,7%	7,1%	21,2%	25,0%	17,4%	2,9%
Indicam o conteúdo dito	0,7%	3,7%	0	0	0,7%	1,9%	0	0
Ajuste da mensagem	5,8%	5,5%	9,9%	10,7%	11,4%	19,2%	21,7%	11,8
Verbos de crença ou ideologia	1,4%	0	2,9%	0	1,5%	0	5,0%	8,8%
Verbos não elocutórios	0	0	8,1%	3,6%	0,7%	0	7,6%	2,9%
TOTAL EM NÚMEROS	138	54	135	28	133	52	119	28

Compreendo que tais resultados numéricos foram calculados manualmente, pois não há menção de anotação dos os verbos de elocução com as nomenclaturas de Thompson (1994) e nem do modo de processamento de etiquetas. Possivelmente a autora parte do *Concord* para selecionar os fragmentos de textos que servem para sua análise e os organiza no Processador de Texto (*MS Word*). Possivelmente tal organização no *Word* (pressupõe-se que tenha sido em tabelas) é tão particular e não sistematizada que se torna muito complexa para ser descrita detalhadamente nas etapas metodológicas. Certas particularidades, como o modo de obtenção dos dados quantitativos, não podem ser resgatadas com a leitura do texto da dissertação. Quanto ao processo de alinhamento, por também não ter sido detalhado, considero que tenha sido preparado manualmente, mas processado automaticamente no *Viewer and Aligner* (já que a autora menciona esta ferramenta).

4.1.9. Marcos Feitosa (2005)

Quadro 19: Resumo das características e procedimentos metodológicos em Feitosa (2005)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2005/UFGM), “Uma proposta de anotação de corpora paralelos com base na Linguística Sistêmico-Funcional”
Objetivo geral e Metafunção	Elaborar um código numérico para a classificação das categorias da Gramática Sistêmico-Funcional. O teste do código é realizado na Metafunção Textual.
Objeto de análise	O CROSF e a organização temática no Estudo Piloto.
O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escaneamento dos textos • Correção “manual” dos erros de leitura do escâner • Alinhamento semi-automaticamente com o auxílio do <i>Viewer and Aligner</i> • Anotação manual
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> • <i>WordList</i> • Tabela do <i>MS Word</i> • <i>Concord</i> • Contagem é realizada manualmente
Códigos de anotação	O corpus do Estudo Piloto foi anotado com o CROSF14.

A dissertação de Feitosa (2005) foi defendida na Universidade Federal de Minas Gerais (UFGM), sob o título “Uma proposta de anotação de corpora paralelos com base na Linguística Sistêmico-Funcional”. Tem como objetivo geral

Elaborar um código numérico para a classificação das categorias da Gramática Sistemico-Funcional de M. A. K. Halliday (GSF), visando a agilizar e aumentar a precisão de pesquisas em corpora de pequenas dimensões que se utilizam dessas categorias, além de uniformizar as anotações dos corpora manipulados pelos pesquisadores do Núcleo de Estudos da Tradução (NET) da Faculdade de Letras (FALE) da Universidade Federal de Minas Gerais (UFMG). Objetiva-se um código aberto suficientemente maleável para pesquisas futuras no NET e, possivelmente, também, de outros grupos de pesquisadores que se utilizem da GSF (Feitosa, 2005)

Assim, seu objeto de análise é o código criado e a organização temática no Estudo Piloto, na Metafunção Textual.

O teste é realizado em um corpus paralelo bilíngue de pequena dimensão (um conto e sua tradução) que se encontrava inicialmente disponível ao pesquisador na versão impressa. A compilação consistiu em realizar escaneamento dos textos com o software *OmniPage*, permitindo gravá-los em *Rich Text Format*. Foi realizada uma correção “manual” dos erros de leitura do escâner. O corpus foi alinhado semi-automaticamente com o auxílio do *Viewer and Aligner*.

O texto foi anotado com o CROSF 14, posicionando-o depois do objeto analisado. Algumas correções no código foram feitas antes mesmo de se terminar a anotação e outras depois de extrair os dados. Ao não se detectar falhas o protótipo foi considerado pronto. Quanto ao modo de anotação da Metafunção Textual: “Todos os Temas e absolutos do corpus selecionado foram anotados com o CROSF. No caso de períodos compostos, cada uma das orações em relação paratática teve o Tema anotado. Por exemplo: “Abri <1111411> a gaveta, vi <1111421> os cadernos.” Quando a relação entre as orações era de hipotaxe ou existem orações encaixadas, apenas o Tema da oração principal foi anotado, a menos que alguma “força paratática” fosse encontrada, como, por exemplo, nas orações causais. Nesses casos, ambas as orações tiveram seus Temas anotados. Por exemplo, nas orações que se seguem: “Guardou-a <1111411> num envelope pardo no qual escreveu *Recortes* e <2130321> guardou <2211411> o envelope entre as páginas de *Sempre-vivas para o jardim de pedras*, onde ninguém jamais olharia.” As orações encaixadas foram desconsideradas para a classificação temática. Também foram desconsideradas as orações hipotáticas. Entretanto, segundo

Halliday (1994: 235), certas orações a priori hipotáticas têm uma certa “força paratática” e foram por isso consideradas nesta análise temática. Exemplo: “Depois <2130331> destruíram <2211411> a si mesmos porque <2130333> eram <2211431> espertos demais.” (Feitosa, 2005).

O corpus do estudo piloto foi processado no *WordSmith* (2001), utilizando-se o *WordList* e o *Concord*, ambos para auxiliar na obtenção dos dados quantitativos.

Esta pesquisa é de cunho metodológico (ver detalhes de sua pesquisa no Capítulo 2, de Referencial Teórico), o que a faz diferente das outras. Feitosa visa mais à criação de um código de rotulação do que à análise discursiva de textos em relação tradutória. O CROSF é discutido no Capítulo 2, de Referencial Teórico, seção 2.2.1.3 desta tese. A iniciativa de Feitosa foi uma inovação no campo dos ESFTBC. Em termos práticos, esse código exige do pesquisador a familiaridade com os números e seus significados, inicialmente demandando um momento de estudo e a fixação da tabela ao lado do computador (cf. Fleuri, 2006). Por focalizar a aplicação do código de rotulação, não se detalha o processo de construção do corpus principalmente quanto ao processo de alinhamento. Como o pesquisador menciona o uso do *Viewer and Aligner*, há margem para compreender que foi necessário passar pelo processo de edição de texto para ser alinhado antes de realizar o alinhamento na ferramenta *Viewer and Aligner* do WST.

4.1.10. Viviane Paquilin (2005)

Quadro 20: Resumo das características e procedimentos metodológicos em Paquilin (2005)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2005/UFSC), “The various facets of a message: an analysis of the thematic structure in Bridget Jones’s diary in the light of the systemic functional grammar, corpus linguistics and translation studies interface”
Objetivo geral e Metafunção	Identificar tendências da tradução em relação às estruturas temáticas e analisar o impacto dessas estruturas no corpus estudado (Metafunção Textual).
Objeto de análise	As estruturas temáticas do corpus
O corpus • Tipo de corpus	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão

• Versão inicial do texto	• Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escanear os textos, • Corrigi-los, • Convertê-los <i>Plain Text</i>, • Anotá-lo e • Alinhá-lo com o auxílio do <i>Viewer and Aligner</i>.
Método de processamento	<ul style="list-style-type: none"> • <i>Viewer and Aligner</i> • <i>WordSmith Tools (1997)</i> • <i>WordList</i> • Contagem dos dados é realizada no <i>WordList</i>
Códigos de anotação	O corpus é anotado com o CROSF.

A dissertação de Paquilin (2005) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “The various facets of a message: an analysis of the thematic structure in Bridget Jones’s diary in the light of the systemic functional grammar, corpus linguistics and translation studies interface”. Tem como objetivo geral identificar tendências da tradução em relação às estruturas temáticas e analisar o impacto dessas estruturas no corpus estudado. Assim, seu objeto de análise são as estruturas temáticas do corpus, sob a perspectiva da Metafunção Textual.

Para desenvolver esta pesquisa, Paquilin utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em inglês *The Bridget Jones’s Diary* (Helen Fielding) e sua tradução para o português brasileiro *O Diário de Bridget Jones* (Trad. Beatriz Horta). O corpus encontrava-se inicialmente disponível à pesquisadora na versão impressa. A compilação deste corpus consistiu em escaneá-lo, corrigi-lo, convertê-lo em documento *Word* em *Plain Text*, anotá-lo e alinhá-lo com o auxílio do *Viewer and Aligner*.

O corpus é anotado com o CROSF, posicionando-o antes do objeto de análise e os elementos não considerados são marcados com um hífen, conforme o exemplo:

<2130--> <2211---> And she told him everything we’d said in a reconciliatory confession [first tag: 2 stands for multiple Theme, 1 stands for first position, 3 stands for Textual and 0 stands for the absence of subcategories; second tag: 2 stands for

Multiple Theme, 2 stands for second position, 1 stands for Ideational and 1 stands for Unmarked Theme] (Paquilin, 2005, p.64).

O processamento do corpus ocorreu no *WordSmith Tools* (1997) através da utilização do *WordList* para identificar a quantidade das etiquetas. Muitos detalhes dos processos metodológicos não foram descritos ao longo da dissertação, principalmente em relação ao modo de lidar com o *WST* no processamento do corpus e de resolução de problemas encontrados.

É interessante observar que no momento em que Paquilin e Morinaka utilizaram o CROSF, este estava ainda em fase de criação e a dissertação de Feitosa não havia sido publicada. Esse fato pode ter gerado duas formas de interpretar o modo de anotação: uma como interpreta Morinaka, em que marca com zero os algarismos relacionados a um elemento não analisado, por exemplo, como a função temática não é em alguns casos analisada, ela marca como zero o primeiro, o segundo e o quarto algarismos, que são relacionados a tal função: <0010161>. Já Paquilin marca com hífen (não necessariamente correspondendo ao número de algarismos que faltam) os algarismos não analisados: <2130--> ou <2211-->. Os casos de etiquetas múltiplas são marcados por Paquilin um seguida do outro, conforme realiza Feitosa em sua dissertação. Morinaka não apresenta exemplos de casos de etiquetas múltiplas.

4.1.11. Gleimara Zuniga (2006)

Quadro 21: Resumo das características e procedimentos metodológicos em Zuniga (2006)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2006/UFSC), “Construing the translator: a meta-reflection grounded in corpus-based translation studies and systemic functional linguistics”
Objetivo geral e Metafunção	Descrever e comparar o perfil de transitividade no modelamento do item lexical “translator” e “tradutor” (Metafunção Ideacional).
Objeto de análise	Os verbos de elocução em projeções paratáticas.

O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escaneamento • Correção dos erros, • Conversão para o formato <i>Plain Text</i>, • Marcação do corpus (por mark-up) • Alinhamento manual apenas das linhas de concordância. • Anotação manual do corpus
Método de processamento	<ul style="list-style-type: none"> • <i>MS Word</i> • <i>Concord</i> • <i>WordSmith Tools</i> (1997) • <i>WordList</i> • A contagem dos dados é realizada no <i>Concord</i>
Códigos de anotação	O corpus é anotado com o CROSF 14.

A dissertação de Zuniga (2006) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “Construing the translator: a meta-reflection grounded in corpus-based translation studies and systemic functional linguistics”. Tem como objetivo geral descrever e comparar o perfil de transitividade no modelamento do item lexical “translator” e “tradutor” na textualização e retextualização. Assim, seu objeto de análise são os itens lexicais “tradutor” e “translator”, Participante de Processo, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Zuniga utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em inglês *Becoming a Translator* (Douglas Robinson) e a tradução para o português *Construindo Tradutor* (Trad. Jussara Simões). A tradução encontrava-se inicialmente disponível à pesquisadora em formato impresso. Conseguiu com o autor a versão digitalizada, da textualização, ainda assim necessitou passar pela etapa de revisão em que foram corrigidas as diferenças entre a versão enviada pelo autor e a presente no livro editado. O texto-original foi compilado pela própria pesquisadora, passando-se pela etapa de escaneamento, correção dos erros, conversão para o formato *Plain Text*, marcação do corpus (por *markup*)—esse processo foi agilizado com o auxílio da ferramenta de *Find & Replace* do

MS Word e o alinhamento manual apenas das linhas de concordância. O corpus foi anotado com o CROSF14.

O alinhamento se dá manualmente no *MS Word*. No *Concord*, buscou-se pelo nódulo “*translator**” e “*tradut**”, na Textualização e Retextualização respectivamente. Os nódulos analisados são deixados e as ocorrências que não são consideradas para fins de análise são removidas (com tecla “zap” do *Concord*). Isso é realizado com a Textualização e a Retextualização. Os complexos oracionais que contêm tais nódulos são transferidos para o *Word*, os complexos oracionais da Textualização são colocados lado-a-lado aos da Retextualização, realizando assim o alinhamento manual apenas dos objetos de análise. A pesquisadora enumera cada complexo oracional selecionado com um número de ordem de aparecimento do complexo no texto.

Após o alinhamento manual, passou-se a fase de inserção do CROSF14 e da análise de Transitividade: cada oração relacionada ao “*tradut**” ou “*translator**”, dentro do complexo oracional, é etiquetado com o CROSF 14. Esse documento é salvo em formato Rich Text File. O Processo e o Participante são anotados em uma só etiqueta, posicionada depois do objeto analisado. Como não era de interesse da pesquisa apontar sua ausência ou presença de “interpolação” a autora optou em marcar a quinta casa do CROSF com o dígito zero (que corresponde a um item não considerado para análise ou ausente de análise). Como o sexto dígito já indica o tipo de Processo e o sétimo o tipo de Participante, os únicos dígitos que variaram durante a análise e a rotulação foram os dois últimos, permanecendo iguais em todas as etiquetas os 5 primeiros dígitos, conforme ilustra o exemplo: *Then the TRANSLATOR has to decide <0010021> whether s/he is willing to undertake <0010011> the job at all.*

O processamento do corpus ocorreu no *WordSmith Tools* (1997) através das ferramentas *WordList* e *Concord*. O corpus foi processado inicialmente no *WordList* para verificar a frequência do nódulo “*translator**”, confirmando o objeto de estudo—a lista mostrou 1078 ocorrências, sendo “*translator**” a palavra mais frequente. A quantificação dos dados é realizada no *Concord*, que fornece o total da etiqueta procurada.

Observo que a organização do capítulo de metodologia de Zuniga segue a ordem proposta por Lincoln Fernandes (2004) em sua tese, que aconselha que tal capítulo contenha em três seguintes sessões: desenho do corpus (*corpus design*), construção do corpus (*Corpus Building*) e processamento do corpus (*Corpus Processing*). Essa forma de organização de fato já vinha sendo seguida nas dissertações anteriores

de modo menos sistematizado e explícito (Morinaka, 2005). Tal organização fez com que ficasse mais fácil rastrear, pela presente tese, detalhes metodológicos da dissertação de Zuniga.

Quanto à forma de etiquetar com o CROSF, observo um modo diferente do aplicado por Morinaka, que também realiza a análise da Metafunção Ideacional, em relação à marcação da quinta casa da etiqueta: Zuniga optar por não apontar se há ou não interpolação no Participante, marcando essa casa com o algarismo zero. Desta forma, mantêm-se como fixos os 5 primeiros números, variando apenas os dois últimos <0010031>, dos quais o sexto corresponde ao tipo de Processo e o sétimo ao tipo de Participante. Essa forma pareceu bastante prática, visto que não houve a necessidade de se usar duas etiquetas: uma para o Processo e outra para o Participante. Em Fleuri (2006), essa economia de etiquetas já não acontece, como será visto em seguida.

4.1.12. Lilian Fleuri (2006)

Quadro 22: Resumo das características e procedimentos metodológicos em Fleuri (2006)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2006/UFSC), “O perfil ideacional dos itens lexicais translator/tradutor em “Translators Through History” e os “Tradutores na História”
Objetivo geral e Metafunção	Analisar a representação ideacional do item lexical “tradutor” e “translator” (Metafunção Ideacional).
Objeto de análise	Itens lexicais “tradutor” e “translator”, núcleo de grupo nominal e Participante de Processo,
O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.

Método de compilação	<ul style="list-style-type: none"> • Escaneamento • Correção dos erros de leitura do escâner • Edição do texto para realizar o alinhamento • Alinhamento com o auxílio do <i>Viewer and Aligner</i> • Transferência do alinhamento a uma tabela no <i>MS Word</i> • Eliminação das “unidades de alinhamento” • Anotação do corpus alinhado • Correção das etiquetas
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (1999) • <i>Concord</i>, • Contagem é realizada no <i>Concord</i>
Códigos de anotação	O corpus foi anotado com o CROSF14

A dissertação de Fleuri (2006) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “O perfil ideacional dos itens lexicais tradutor/tradutor em “Translators Through History” e os “Tradutores na História”. Tem como objetivo geral analisar a representação ideacional do item lexical “tradutor” e “translator” no corpus adotado. Assim, seu objeto de análise são itens lexicais “tradutor” e “translator”, núcleo de grupo nominal e Participante de Processo, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Fleuri utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em inglês *Translator Thought History* (Deslile e WoodsWorth) e sua tradução para o português *Os Tradutores na História* (Trad. Sérgio Bath), que se encontravam inicialmente disponíveis na versão impressa, tendo sido compilados pela própria pesquisadora. A compilação consistiu em realizar o escaneamento (com *OCR*), correção dos erros de leitura do escâner (menciona a utilização da ferramenta *Find & Replace* do *MS Word* para auxiliar esta tarefa), edição do texto para realizar o alinhamento (também menciona o auxílio de *Find & Replace*), conversão para o formato *Plain Text*, alinhamento com o auxílio do *Viewer and Aligner*, transferência do alinhamento a uma tabela no *MS Word*, editadas para melhor visualização e análise, eliminação das “unidades de alinhamento” (cf. Assis, 2004) que não apresentavam os nódulos que seriam analisados, anotação do corpus alinhado com o CROSF14 e correção das etiquetas.

O CROSF14 é anotado antes do objeto analisado. Para realizar a anotação, Fleuri seguiu duas etapas fundamentais: a de identificação dos Participantes e Processos, inserindo o código comum a todos (<0010) e a de classificação dos Processos e Participantes e de reconhecimento da Circunstância, acrescentando os outros códigos (yyy>). Foi utilizada uma etiqueta para anotar o Processo e outra para anotar o Participante. Os casos de elipses do objeto analisado foram marcados com o símbolo Ø, conforme os exemplos: *Miniatures usually <0010330> represent <0010132> <0010151> TRANSLATORS humbly <0010350> kneeling at the feet of a monarch, or some other royal personage, and <0010141> Ø <0010340> presenting <0010> their work to him” (p.79-80). Fleuri em alguns casos utilizou o CROSF com uma variação: como o CROSF não oferece o detalhamento do tipo de léxico em relação ao seu grupo (nominal, verbal, adverbial etc.), Fleuri adotou a barra transversal depois do 7º dígito como indicativo de que havia uma classificação adicional referente ao seu “grupo”, indicando se era modificador de um núcleo nominal ou se era modificador de um modificador. Sugeriu ainda outras indicações, como sendo um “elemento referencial (r), um elemento elíptico (e) ou se é ele próprio (t)” (Fleuri, 2006, p.78). Para as demais etiquetas a pesquisadora fixou os 4 primeiros dígitos <0010114> e variou segundo a classificação os três últimos números.*

O processamento do corpus ocorreu no *WordSmith Tools* (1999), a partir da ferramenta do *Concord*, onde os dados são quantificados. Cada texto foi processado separadamente no *Concord*, buscando-se por cada etiqueta. O resultado final do número de etiquetas é anotado e convertido em tabela e gráficos para a análise.

Observo que Fleuri segue a mesma estrutura de organização do capítulo de metodologia proposta por Fernandes (2004), assim como realizou Zuniga. Há um grande nível de detalhamento e ilustração em cada passo na construção do corpus, separando essa subseção em quatro fases (cujos passos também são ilustrados com figuras da tela de computador): de digitalização, correção e formatação do texto-fonte e do texto-alvo; de alinhamento; de rotulação; e de correção da análise e preparação para o processamento e contagem dos dados. Há uma preocupação especial em detalhar o modo de preparação do texto para a realização de correções de classificação, que não se observa nas outras pesquisas. Quanto à forma de etiquetar o corpus, Fleuri também utiliza o CROSF 14, marcando a quinta casa com o número 1, independente do Participante ser ou não interpolado (visto que esse não fazia parte do escopo de sua pesquisa); etiqueta tanto o Participante quanto o Processo em que este se encontra envolvido, cada um com uma etiqueta diferente.

Além disso, a pesquisadora menciona ter necessitado criar variações na etiqueta do CROSF para anotar o tipo de grupo nominal realizado pelo Participante. Essa necessidade é sinalizada no capítulo de metodologia, mas não é seguida até o capítulo de análise, uma vez que essas estruturas não são quantificadas ou mencionadas.

4.1.13. Ariana Zanella (2006)

Quadro 23: Resumo das características e procedimentos metodológicos em Zanella (2006)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2006/UFSC), “Mapeamento macro e micro estrutural da retextualização de resumos on-line: estudo da transitividade de abstracts biométricos”.
Objetivo geral e Metafunção	Descrever a Retextualização dos Processos e Participantes em resumos/abstracts em relação a “macroestrutura” do gênero resumo.
Objeto de análise	Processos e Participantes do corpus. (Metafunção Ideacional).
O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão digitalizada, online.
Método de compilação	<ul style="list-style-type: none"> • Alinhamento manual • O corpus foi anotado com auxílio do <i>Syscoder</i>
Método de processamento	<ul style="list-style-type: none"> • <i>Syscoder 4.6</i> (Michael O’Donnell, 2004) • Contagem é realizada manualmente
Códigos de anotação	A anotação foi realizada através do software <i>Syscoder</i>

A dissertação de Zanella (2006) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “Mapeamento macro e micro estrutural da retextualização de resumos on-line: estudo da transitividade de abstracts biométricos”. Tem como objetivo geral descrever a Retextualização dos Processos e Participantes em resumos/abstracts em relação a “macroestrutura” do gênero resumo.

Assim, seu objeto de análise são os Processos e Participantes do corpus, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Zanella utiliza um corpus paralelo bilíngue de pequena dimensão. O corpus é composto por quinze resumos na área médica, de pesquisas classificadas como Qualis A nos critérios da Capes. Encontrava-se inicialmente disponível à pesquisadora online nos e-journals veiculados na MEDLINE, ou seja, na versão digitalizada. A compilação consistiu em realizar apenas o alinhamento manual conforme sua organização retórica. O corpus foi anotado no software de processamento.

O processamento do corpus ocorreu no *software Syscoder 4.6* (O'DONNELL, 2004). Hoje foi substituído por UAMCorpus Tool⁶⁸. Após serem salvos em formato *Plain Text*, os corpora foram importados ao *software Syscoder 4.6* (idem), utilizando a ferramenta *import file*. Foi criada uma taxonomia no software. Os textos alinhados manualmente foram segmentados automaticamente em orações pelo software (para a análise dos Processos e Participantes) e em sentenças e parágrafos (para a análise dos movimentos retóricos). Após a codificação realizada no próprio *Syscoder*, alguns casos foram revisados e filtrados para serem utilizados de exemplo, seus resultados estatísticos foram descritos e comparados (no próprio software) para posterior análise. A anotação foi feita através do software *Syscoder*, da seguinte forma: criou-se um esquema de codificação para a análise da macroestrutura e da microestrutura; elaborou-se uma taxonomia de rede de sistemas; e desenhou-se um sistema para a organização retórica; criou-se uma taxonomia na sessão *scheme*.

Com auxílio da ferramenta *coding* (do *Syscoder*) cada oração, sentença e parágrafo segmentado podem ser visualizados em caixas de textos com as categorias de análise (fruto da taxonomia criada), para a realização da codificação semiautomática. As dúvidas podem ser anotadas na caixa de texto *comment*. “Estes comentários poderão ser impressos junto com os resultados obtidos através de filtragem posterior” (Zanella, 2006, p.32). Os segmentos foram codificados por participantes,

⁶⁸ “NOTE: Systemic Coder is no longer supported. It has been replaced by a much more advanced tool, called UAM CorpusTool, which does the same thing as Coder, but allows multiple texts to be annotated with the same scheme, and each text to be annotated at multiple levels”.

Fonte: <http://www.wagsoft.com/Coder/> (acesso em maio de 2012, Brisbane, Austrália)

e por processos a cada oração de cada estágio retórico, mas as circunstâncias foram segmentadas sem codificação detalhada somente para efeito estatístico, e não para análise (idem, ibidem). As estatísticas foram geradas automaticamente pelo software *Syscoder*.

Esse software permite visualizar os resultados de dois textos comparativamente, isto é, lado-a-lado. O *Systemic Coder*, que pode ser baixado gratuitamente online⁶⁹ em plataformas Windows, Macintosh e Linux, é

Uma ferramenta que facilita a codificação linguística de textos. O usuário é levado a categorizar cada segmento textual dentro de uma rede de sistemas, providenciadas pelo usuário. A ferramenta apresenta uma interface que facilita a segmentação que permite uma construção/edição gráfica da rede de sistemas. A interface de análise permite que se recupere todos os segmentos que apresentam uma determinada característica ou uma construção lógica de características. A interface de estatística oferece uma descrição estatística dos códigos, ou comparações estatísticas entre subconjuntos dos códigos. Tais códigos podem também ser exportados em um formato legível por pacotes estatísticos.⁷⁰

Este parece ser um software interessante, digno de ter suas funções mais exploradas em pesquisas em ESFTBC. Conforme declaração da pesquisadora em uma troca de e-mails (ver anexo), esse software é de fácil manuseio, sendo que a fase que mais lhe tomou tempo e atenção foi a de elaboração da taxonomia, que deve ser realizada pelo próprio pesquisador segundo seus propósitos de pesquisa:

O *Syscoder* é fácil de baixar e manipular. Você precisa saber bem fazer a matriz sistêmico-funcional, é como na linguística de corpus, depois *tagging* o seu corpora. Manualmente seria impossível ou bem difícil. O melhor é que ele te dará um percentual de

⁶⁹ Download gratuito está disponível no site:
<http://www.wagsoft.com/Coder/index.html>

⁷⁰ Fonte:
http://ccl.pku.edu.cn/doubtfire/NLP/Text_Generation/Systemic%20Software.htm

confiança utilizando *student T test* e estatísticas tratadas para cada processo do sistema de transitividade (Zanella, e-mail trocado dia 28 de maio de 2012).

Esse programa apresenta a vantagem de que as etapas de segmentação textual, a codificação sistêmico-funcional e quantificação dos dados se dão em um só programa, no próprio *Systemic Coder*, evitando a transição entre Microsoft *Word* e *WordSmith Tools*: texto editado no *Word* para a segmentação, conversão em txt, *WordSmith Tools* para a concordância, copia e cola para *Word* para a codificação, converte em *Plain Text* e processo novamente no *WordSmith Tools* para a quantificação. Entretanto, ele exige a elaboração e o teste de uma taxonomia para que essa codificação seja realizada no próprio programa. A criação dessa taxonomia pode levar o mesmo tempo que toma a realização da anotação textual, no caso dos que optam por usar um programa de análise lexical, como o *WST*. Esse programa também não oferece uma ferramenta de alinhamento (semi-)automático ou de visualização de textos em paralelo. Mas, dado que quase a metade dos pesquisadores optaram por realizar um alinhamento manual, isto não representa necessariamente uma desvantagem. Para quem quiser optar por realizar sua pesquisa neste programa, os detalhes de uso se encontram disponíveis no site <http://www.wagsoft.com/Coder/index.html>.

A dissertação de Zanella é a única aqui analisada que não pertence ao projeto CORDIALL. Foi analisada por se encontrar entre as pesquisas em ESFTBC. É a primeira e única, dentro desta tradição a utilizar um programa pronto de análise textual baseada na Gramática Sistêmico-Funcional e a não utilizar o *WordSmith Tools*.

4.1.14. Daniel Alves (2007)

Quadro 24: Resumo das características e procedimentos metodológicos em Alves (2007)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2007/UFMG), “Aspectos da representação do discurso em textos traduzidos: os verbos de elocução neutros”.
Objetivo geral e Metafunção	Analisar a representação do discurso em citações paratáticas (Metafunção Ideacional).
Objeto de análise	Os verbos de elocução.

O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão digitalizada.
Método de compilação	<ul style="list-style-type: none"> • Revisão e correção dos erros de digitalização, • Organização do corpus conforme as diretrizes de preparação de corpus do projeto CORDIALL • Marcação do corpus conforme tais diretrizes • Anotação das aspas, travessões e dois-pontos. • Classificação dos discurso no Openoffice; • Identificação dos Processos Verbais • Alinhamento manual em tabelas software Openoffice.org Writer.
Método de processamento	<ul style="list-style-type: none"> • Openoffice.org Writer • <i>WordSmith Tools</i> (Scott, 2005) • <i>Concord</i> • <i>Openoffice.org Calc</i> • Contagem é realizada no openoffice.org Calc
Códigos de anotação	O corpus é anotado com rótulos criados pelo próprio pesquisador.

A dissertação de Alves (2007) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Aspectos da representação do discurso em textos traduzidos: os verbos de elocução neutros”. Tem como objetivo geral analisar a representação do discurso em citações paratáticas no corpus adotado. Assim, seu objeto de análise são os verbos de elocução, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Alves utiliza um corpus paralelo bilíngue de pequena dimensão (duas traduções de um mesmo original). Os textos encontravam-se disponíveis já na versão digitalizada. Assim, a compilação consistiu em realizar a conferência do corpus digitalizado com os textos impressos, a revisão e correção dos possíveis erros de digitalização, a organização do corpus conforme as diretrizes de preparação de corpus do projeto CORDIALL—os pontos devem ser seguidos de espaço em branco e entre os parágrafos deve haver “uma

linha em branco”. O corpus foi marcado (mark-ups) conforme tais diretrizes. Anotaram-se as aspas, os travessões e os dois-pontos. As linhas de concordância dessas etiquetas foram passadas a uma planilha do *OpenOffice*, classificando cada linha como sendo um caso de discurso direto (DD) ou de discurso direto livre (DDL). Os Processos Verbais foram identificados e escritos na sua forma infinitiva em uma coluna ao lado das linhas de concordância. Alinhamento dos seis primeiros capítulos foi realizado manualmente em tabelas no mesmo software (*OpenOffice.org Writer*).

Alves cria seu próprio código de rotulação. Rotula antes do elemento analisado, com as seguintes etiquetas: <trav>, após a ocorrência do travessão; <dois pontos>, após a ocorrência dos dois pontos; <asp> e </asp>, para indicar o início e final de aspas respectivamente. Essas etiquetas foram inseridas e imediatamente processadas no *Concord* e sua quantidade e aplicações conferidas. Concluiu-se que as aspas e travessões são os símbolos textuais que mais representam presença de discurso no texto. A concordância de ambas as etiquetas foi levantada e cada ocorrência, separada entre casos de discurso direto e discurso direto livre. Com as linhas de ocorrência de tais rótulos em tabela, classifica-se cada linha em DD (Discurso Direto) e DDL (Discurso Direto Livre). Essas linhas de concordância foram organizadas em uma planilha do *OpenOffice.org* conforme ilustra a figura (Alves, 2007, p.54). "Neste ponto da pesquisa, foram também verificados erros de marcação e casos de linhas de concordância em duplicidade" (idem, ibidem). Cada linha foi classificada como sendo discurso direto (DD), discurso direto livre (DDL) ou “excluir”, que apontou casos de duplicidades.

Quadro 25: codificação adotada por Alves (2007)

	Discurso	Concordance	File
1	DD	uck Finn, he hain't got no family -- what you going to do 'bout him? " <Asp>"Well, hain't he got a father? " says Tom Sawyer. "Yes, he's	c:\twain.txt
2	excluir	im? " "Well, hain't he got a father? " says Tom Sawyer. <Asp>"Yes, he's got a father, but you can't never find him, these days. He used	c:\twain.txt
3	DDL	them till they're ransomed. " "Ransomed? What's that? " <Asp>"I don't know. But that's what they do. I've seen it in books; and so of co	c:\twain.txt
4	DD	ichões pelo corpo todo. Ao cabo de certo tempo Jim rompeu o silêncio. --<Trav> Quem está aí? repetiu ele. É alguém, bem sei. Meus ouvidos não me enganam	c:\lobato.txt
5	DDL	--<Trav> Cale essa boca! Que é que você entende de bruxas? e o herói metia-se nas	c:\lobato.txt
6	DD	alguma amplitude, onde podíamos ficar de pé. Era ali o ponto da reunião. --<Trav> Muito bem, disse Tom. Vamos agora lançar as bases da nossa quadrilha -- o	c:\lobato.txt
7	DD	ponta do dedo para tirar sangue e assinar, e eu fiz a minha marca no papel. <Asp>"Agora", disse Ben Rogers, "o que é que a nossa quadrilha vai fa	c:\flaksman.txt
8	excluir	nar, e eu fiz a minha marca no papel. "Agora", disse Ben Rogers, <Asp>"o que é que a nossa quadrilha vai fazer? " "Só assaltar e ma	c:\flaksman.txt
9	DDL	<Asp>"Mas assaltar quem? Casas? Ou gado, ou. . ." "Imagine! Rouba	c:\flaksman.txt

Os três textos foram processados ao mesmo tempo no *WST*, o que levou o autor da pesquisa a anotar na quarta coluna o arquivo de onde foi extraída a linha de concordância. Essa planilha também permitiu que se detectassem erros de anotação do corpus. Uma vez identificados, os erros eram corrigidos no corpus e este novamente processado. A partir desta tabela, foram identificados os verbos de elocução que representavam discurso (marcados em negrito) e os discursos representados (marcado em itálico). Nos casos de discurso direto livre (DDL), houve a necessidade de retornar ao corpus e verificar se realmente não havia a presença de orações verbais projetantes (de discurso).

O processamento do corpus ocorreu no *WordSmith Tools* (Scott, 2005), com a utilização da ferramenta *Concord*, para a identificação das etiquetas anotadas no corpus, e a organização do alinhamento e a obtenção dos dados quantitativos no *OpenOffice.org Calc*, pois, segundo o autor, este software facilita a organização de textos em tabelas, assim como a contagem e a sistematização das linhas de concordância levantadas.

O *OpenOffice.org* é um programa gratuito, possível de ser baixado online através do site <http://www.openoffice.org>. Por este programa apresentar uma interface semelhante à da *Microsoft Excel* (cf. <http://www.openoffice.org/why/>) e pelo fato de todos os pesquisadores terem mencionado obter acesso ao *Microsoft Office*, toma-se como entendido que a descrição detalhada deste programa seria redundante para os leitores desta tese. O que nos interessa é apontar a possibilidade de explorar programas como o *OpenOffice.org Calc* ou o *MS Excel* para os

ESFTBC, no que refere principalmente à quantificação dos dados (conforme realizado por Alves, 2007), à visualização dos textos em paralelo e, por que não, à anotação do corpus. Essa questão foi explorada em mais detalhe no capítulo de Referencial Teórico ao se explicar o funcionamento do *MS Excel*.

Observo que a pesquisa de Alves é a única, até então, a utilizar o OpenOffice.org—versão gratuita do *MS Excel*. A vantagem em relação ao *WST* é que permite elaborar cálculos que realizam automaticamente a contagem de determinadas células e elementos dentro destas. Quando o trabalho com tabelas é realizado no Microsoft *Word*, o pesquisador enfrenta problemas de configuração e de formatação, o que lhe toma tempo em organiza-la e voltar a organizá-la sempre que tal documento é aberto e reaberto. Tal desconfiguração não ocorre no OpenOffice.org ou no *MS Excel*. A forma de organizar a classificação dos objetos analisados e o alinhamento em planilhas do OpenOffice.org pareceu uma opção prática e de fácil visualização das linhas de concordância, classificações e grupos verbais realizados pelos Processos Verbais, em um só documento. A análise e organização dos dados são realizadas simultaneamente ao processamento do corpus, economizando tempo e tipos de arquivos. O *WST* é utilizado apenas para identificar as linhas de *Concordância*, o resto é manipulado no OpenOffice.org.

4.1.15. Giacomo Figueredo (2007)

Quadro 26: Resumo das características e procedimentos metodológicos em Figueredo (2007)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2007/UFMG), “Uma descrição sistêmico-funcional da estrutura do grupo nominal em português orientada para os estudos linguísticos da tradução”.
Objetivo geral e Metafunção	Descrever o sistêmico-funcional da ordem do grupo em português a partir de seus elementos estruturais.
Objeto de análise	Classes de grupos e itens na ordem da palavra que serviram de elementos na estrutura dos grupos. (Não explorou uma Metafunção em especial).

O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus monolíngue em português • Corpus compilado e não-compilado
Método de compilação	<ul style="list-style-type: none"> • Inseriu marcações ao longo do corpus • Fez anotações em tabelas.
Método de processamento	<ul style="list-style-type: none"> • <i>Não menciona uso de programas durante o processamento</i> • Não quantifica os dados.
Códigos de anotação	Criou seu próprio código para anotar informações gerais do corpus

A dissertação de Figueredo (2007) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “Uma descrição sistêmico-funcional da estrutura do grupo nominal em português orientada para os estudos linguísticos da tradução”. Tem como objetivo geral realizar uma descrição sistêmico-funcional da ordem do grupo em português a partir de seus elementos estruturais, em outras palavras, promover a identificação da ordem do grupo em português e a descrição dos sistemas componentes das estruturas experiencial e lógica da classe do grupo nominal. Assim, seu objeto de análise são classes de grupos e itens na ordem da palavra que serviram de elementos na estrutura dos grupos. Não explorou uma Metafunção em especial; trabalhou com as noções de grupos, registro Interpessoal e Experiencial.

Para desenvolver esta pesquisa, Figueredo utiliza um corpus monolíngue em português, contendo 12000 palavras, composto por textos em português de registros diversos. Compreendo que o pesquisador recorreu a um corpus compilado por ele e outro já compilado pelo NILC, o Lácio-Web⁷¹, para a identificação de estruturas não presentes no corpus compilado. O pesquisador não detalha os procedimentos de compilação.

Figueredo criou seu próprio código para anotar informações gerais do corpus—não rotulou elementos linguísticos. Inseriu marcações ao longo do corpus e as anotações fez em tabelas. Montou um corpus anotado total de aproximadamente 4.800 palavras. A anotação foi manual, da seguinte forma: segmenta as orações separando-as em grupos; identifica as funções experimentais e interpessoais do grupo nas orações e as classes que esse formavam; classifica os grupos com base nas categorias propostas pela LSF (Figueredo, 2007, p.127).

⁷¹ www.nilc.icmc.usp.br/lacioweb/downloads.htm

As marcações levaram em consideração cada categoria teórica da gramática de categoria em escala, o que aumentou consideravelmente o tempo tomado no processo de anotação. Segundo o pesquisador, o tempo de anotação de um item teve de ser multiplicado por seis, pois foi analisado em termos de (i) unidade, (ii) estrutura, (iii) classe, (iv) sistema, além de sua função na oração (v) ideacional e (vi) interpessoal. As informações anotadas do texto foram: <número do texto>, <processo sócio-semiótico>, <modo de interação>, <modo de veiculação>, <tipo de texto (registro)>, <título>, <autor>, <data de produção do texto>, <número de palavras>, <Nota>. Exemplo: <TEXTO 01>, <Compartilhar> <Monólogo> <Escrito>, <Carta>, <De volta>, <Autor: Bruno Profeta> <Data: 23/03/2006 > <Palavras: 316 >, <Nota: <http://mail.google.com/mail/>> etc.

Considerando a citação a seguir, compreendo que não foram utilizados programas de processamento de corpus:

Não houve extração de dados quantitativos, pois, pela forma como esta análise foi conduzida – a partir da teoria geral de descrição sistêmica, levando em conta ambientes amplos e restritos, as dimensões e as categorias teóricas – cada dado teve de ser analisado individualmente de forma qualitativa. Desta forma, a noção de probabilidade de ocorrência tal como está entendida nesta dissertação não está relacionada diretamente ao volume absoluto de ocorrências para uma determinada instância, mas sim com a probabilidade de o sistema linguístico do português ter os recursos necessários para produzi-la (idem, p.128).

Observo que Figueredo (2007) concentra a atenção na descrição da metodologia de análise, muito mais que na de construção e processamento do corpus. Desta forma, coletar os itens estipulados pela presente tese não se constituiu como uma tarefa simples. A dissertação de Figueredo segue um método de pesquisa e um objetivo diferente das demais pesquisas, no sentido de adotar uma visão linguístico-funcional da tradução, “na qual a tradução é entendida a partir da dimensão da instanciação como uma relação entre sistemas linguísticos e a descrição linguística se constitui como uma das etapas dos estudos tradutórios no contexto desta visão” (Figueredo, 2007, p.5). Figueredo realiza uma “descrição sistêmico-funcional do português como instrumento de análise e estudo da tradução” (idem, *ibidem*). Isso o leva a compilar e focar seu

estudo em um corpus monolíngue em português, sendo a primeira pesquisa que se contextualiza nos ESFTBC e que trabalha com esse tipo de corpus.

Sua pesquisa não tem cunho quantitativo; não usa diretamente o processamento de corpus por ferramentas de concordância ou de quantificação (como o *WordList* ou *Keyword*). A noção de trabalho com corpus parte do manuseio de diferentes tipos de textos e do acesso a corpus de grande dimensão, disponibilizado online.

4.1.16. Cristiano Araújo (2007)

Quadro 27: Resumo das características e procedimentos metodológicos em Araújo (2007)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2007/UFMG), “O sistema semântico de PROJEÇÃO e sua dispersão gramatical em português brasileiro: uma descrição sistêmico-funcional orientada para os estudos linguísticos da tradução”.
Objetivo geral e Metafunção	Descrever o sistema semântico de PROJEÇÃO em português brasileiro, assim como sua dispersão gramatical nesta língua, no escopo da teoria sistêmico-funcional da linguagem. (Metafunção Ideacional).
Objeto de análise	Projeções de orações projetadas.
O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus monolíngue em português • Corpus compilado (de referência)
Método de compilação	<ul style="list-style-type: none"> • O corpus foi anotado manualmente • O corpus foi marcado (<i>Markups</i>),
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> • <i>Concord</i> • Os dados não foram quantificados
Códigos de anotação	Anota o corpus com símbolos para indicar projeção.

A dissertação de Araújo (2007) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “O sistema semântico de PROJEÇÃO e sua dispersão gramatical em português brasileiro: uma descrição sistêmico-funcional orientada para os estudos linguísticos da tradução”. Tem como objetivo geral descrever o sistema semântico de PROJEÇÃO em português brasileiro, assim como sua dispersão gramatical nesta língua, no escopo da teoria sistêmico-funcional da linguagem. Seu objeto de análise são as projeções de orações projetadas, sob a perspectiva da Metafunção Ideacional.

Compreendo que o corpus seja monolíngue em português: “corpus com cerca de 30.000 *tokens*, (...) seu conteúdo é composto por amostras textuais”, abrangendo diferentes tipos textuais (p.65). Foi também realizada a consulta ao corpus de referência Lácio-Ref. O corpus já se encontrava digitalizados. Foram coletadas amostras de textos. Para o procedimento desta coleta, o pesquisador seguiu os procedimentos: “1) manter excertos inteiros (i.e. capítulos inteiros, seções dentro de capítulos, ou pelo menos, parágrafos que concluam certo assunto); 2) limitar linguisticamente e temporalmente as amostras incluídas no corpus, e com efeito, incluir apenas extratos textuais em português brasileiro produzidos nos séculos XX e XXI; 3) não se retirar apenas trechos do início ou fim dos textos, embora introduções e conclusões também estejam representadas no corpus; e, por fim, 4) incluir preferencialmente extratos de textos já disponíveis em formato eletrônico (i.e. já digitalizados ou disponíveis na internet ou em CD-ROM)” (idem, p.65).

O corpus foi anotado manualmente utilizando os símbolos |||, ||, |, [[]], <<>>, para indicar projeção, números, para identificar relações paratáticas e letras gregas para relações hipotáticas. Exemplo da anotação com símbolos: *Quando é submetida a uma revista especializada, | a pesquisa é analisada por cientistas renomados | que revisam o trabalho | em busca de possíveis falhas*”, || conta <projeção> Átila da Rosa, paleontólogo da Universidade Federal de Santa Maria. |||

Quadro 28: símbolos de análise adotados por Araújo (2007)
 Anotação das relações de interdependência entre as orações

Símbolo	Função
	Separar as sentenças
	Separar as orações em relação paratática
	Separar as orações em relação hipotática
[[]]	Separar as orações encaixadas (“embedded clauses”)
<< >>	Separar as orações intercaladas (“interrupting clauses”)

Fonte: Araújo, 2007.

O corpus foi também marcado (*Markups*), posicionando as etiquetas antes da projeção. Exemplo: <Processo sócio-semiótico> <modo> <tipo de relação> <tipo textual> <título> <referência: veículo – data ou ano de publicação– edição – outras informações relevantes> <projeção>. Araújo marcou os Títulos e referências segundo especificações sugeridas pelo CORDIALL.

O processamento se deu no *WordSmith Tools* (o pesquisador não especifica o ano do programa), através da ferramenta do *Concord*. Apesar de não haver, no capítulo de análise, a exposição e discussão dos dados quantitativos, uma observação desses dados através do *Concord* é mencionada:

Os dados quantitativos e linhas de concordância foram geradas com o auxílio das ferramenta *Concord*. A partir destes dados, procedeu-se à descrição do sistema de PROJEÇÃO, observando-se como as opções do estrato semântico deste sistema são realizadas pela lexicogramática. Em casos de dúvida, ou em que as ocorrências do corpus eram demasiado reduzidas para traçar considerações mais abrangentes sobre determinados pontos observados, procedeu-se a consultas ao corpus Lácio-Ref, utilizando a ferramenta de concordância de seu servidor. Este procedimento buscou ampliar a confiabilidade das considerações traçadas (Araújo, 2007, p.76).

Observo que as pesquisas de Figueredo e Araújo são pesquisas orientadas para os estudos linguísticos da tradução, ou seja, analisam a partir da Gramática Sistemico-Funcional certos elementos linguísticos em corpora monolíngues em português. Os textos destes corpora não se encontram em relação tradutória. Na citação abaixo o pesquisador justifica sua afiliação com os Estudos da Tradução:

O trabalho considerou a linguística sistêmico-funcional como parte integrante de um modelo de estudo da tradução, e buscou contribuir com este modelo não por meio da análise de textos em relação tradutória, mas por meio da descrição de um sistema semântico do português que pudesse subsidiar futuros trabalhos descritivos e/ou discursivos de tradução, e de outros tipos de produção textual (Araújo, 2007, p.122).

Assim como em Figueredo, não há um enfoque na discussão dos dados quantitativos, ou seja, os excertos anotados e observados são descritos em sua forma de ocorrência, e não em sua quantidade, justificando que “O presente estudo tem por objetivo descrever o sistema semântico de PROJEÇÃO e sua dispersão gramatical em português brasileiro” (idem, p.77) Além disso, ambos os autores realizam consultas a um corpus de referência chamado Lácio-Web.

Araújo (2007) menciona⁷² o programa *Syscoder*, como uma alternativa ao *WordSmith Tools*, mas não relata tê-lo utilizado em sua pesquisa. Esta é a segunda vez que o *Syscoder* é mencionada em pesquisas em ESFTBC.

4.1.17. Joanna Filgueiras (2007)

Quadro 29: Resumo das características e procedimentos metodológicos em Filgueiras (2007)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2007/UFSC), “Capoeira em tradução: representações discursivas em um corpus paralelo bilíngüe”.
Objetivo geral e Metafunção	Analisar o perfil da prosódia semântica de <i>capoeir*</i> , <i>malícia*</i> , <i>malandr*</i> , e <i>mandinga*</i> (não analisa nenhuma Metafunção)
Objeto de análise	Os verbos de elocução em projeções paratáticas.

⁷² “Há aplicativos que integram, em um mesmo ambiente de trabalho, um sistema semi-automático de anotação, a partir de categorias pré-definidas, com um sistema de levantamento automático de dados. Menciona-se, por exemplo, o Coder, desenvolvido por Mike O’Donnel . Contudo, estes não foram utilizados porque para os propósitos da pesquisa pareceu ser mais adequado a análise de padrões a partir de linhas de concordância”. (Araújo, 2007).

O corpus <ul style="list-style-type: none"> • Tipo de corpus • Versão inicial do texto 	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escanear os textos • Convertê-los a Word Document, • Corrigir erros de escaneamento • Convertê-los a <i>Plain Text</i>; • O corpus não foi anotado ou marcado • O corpus não foi alinhado
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (Scott, 1999) • <i>Concord</i> • <i>WordList</i> • <i>Keyword</i> • A contagem dos dados se deu através do <i>WordList</i> e <i>Concord</i>
Códigos de anotação	Não anota o corpus.

A dissertação de Filgueiras (2007) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “Capoeira em tradução: representações discursivas em um corpus paralelo bilíngue”. Tem como objetivo geral analisar o perfil da prosódia semântica de *capoeir**, *malícia**, *malandr**, e *mandinga** no corpus adotado. Assim, seu objeto de análise são as palavras-chaves *capoeir**, *malícia*, *malandr**, e *mandinga*. Não analisa nenhuma Metafunção, centrando sua discussão em Prosódia Semântica.

Para desenvolver esta pesquisa, Filgueiras utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em português *Capoeira. Pequeno manual do jogador* (Nestor Capoeira) e a tradução para o inglês *The Little Capoeira Book* (Trad. Alex Ladd), os quais se encontravam disponíveis na versão impressa. A compilação consistiu em escanear os textos e convertê-los a Word Document, corrigir erros de escaneamento e convertê-los a *Plain Text*. Não houve necessidade de anotar ou marcar o corpus. Pelo fato de o texto-original e o traduzido não corresponderem em estrutura, o corpus não foi alinhado. Quando possível e necessário, o alinhamento era realizado manualmente.

O processamento do corpus ocorreu no *WordSmith Tools* (Scott, 1999) através das ferramentas do *Concord*, *WordList*, *Keyword*. A ferramenta do *Keyword* foi utilizada para ver a frequência da palavra no corpus, mas de fato serviu para “contextualizar o ambiente colocacional

dos itens lexicais investigados” (Filgueiras, 2007, p.71). A quantificação dos nódulos estudados foi realizada através do *WordList* e do *Concord*.

Observe que, em comparação às dissertações anteriores, o trabalho de Filgueiras foi o primeiro a tratar de coesão lexical, partindo de uma visão mais contextual, explicando o objeto de análise pelo contexto, mais do que pela função deste na oração, no texto. Em outras palavras, sua análise partiu da perspectiva Macro textual para explicar as evidências Micro textuais. Não houve a geração de etiquetas; apesar de ser considerado um corpus paralelo, este não foi alinhado. As concordâncias das palavras analisadas foram feitas diretamente no *Concord* e o cálculo destas extraídas do *WordList*.

4.1.18. Alinne Fernandes (2009)

Quadro 30: Resumo das características e procedimentos metodológicos em Fernandes (2009)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2009/UFSC), “Black into white and preto no branco: can you tell one’s colour by the company one keeps?”
Objetivo geral e Metafunção	Analisar a representação do perfil ideacional de mestiço (Metafunção Ideacional).
Objeto de análise	Item lexical “mestiço” e a cadeia coesiva.
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Corpus compilado de grande dimensão • Versão impressa.
Método de compilação	<ul style="list-style-type: none"> • Escaneamento • Conversão em .rtf; • Organização de cada capítulo em um documento; • Correção de leitura do escâner • Formatação dos textos para o alinhamento • Marcação do corpus de acordo com os critérios da <i>TEI</i> • Alinhamento do corpus no <i>Viewer and Aligner</i>, • Anotação dos objetos de análise com o CROSF 14.

Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (Scott, 1999) • <i>WordList</i> • <i>Concord</i> • <i>Keyword</i> • Contagem é realizada no <i>Concord</i>
Códigos de anotação	Anota o corpus com o CROSF 14

A dissertação de Fernandes (2009) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “Black into white and preto no branco: can you tell one’s colour by the company one keeps?”. Tem como objetivo geral analisar a representação do perfil ideacional de mestiço no corpus em questão. Assim, seu objeto de análise é o item lexical “mestiço” e a cadeia coesiva (color*, mixed blood*, mulatto*, non-white*), sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Fernandes utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em inglês “*Black into White*” e sua tradução para o português “Preto no Branco”, os quais se encontravam originalmente disponíveis na versão impressa. Consultou também o corpus compilado de grande dimensão *The British National Corpus*. A compilação do corpus paralelo consistiu em realizar escaneamento em *OCR*, convertendo-o em .rtf; organização de cada capítulo em um documento; correção de leitura do escâner comparando a versão digitalizada com a cópia impressa; formatação dos textos para o alinhamento (retirou os pontos não marcadores de final de sentença, realizou a quebra de linha, separou palavras juntas etc., contou em grande parte do trabalho com o auxílio da ferramenta de substituir do *Word*); marcação do corpus de acordo com os critérios da *TEI* (*Text Encoding Initiative*); alinhamento do corpus no *Viewer and Aligner*, cada capítulo separadamente, pois, de acordo com a autora, o *WST* “não salva o trabalho e uma vez começado tem que ser terminado” (Fernandes, 2009, p.57) e anotação dos objetos de análise com o CROSF 14.

Como analisou o perfil ideacional, manteve os quatro primeiros números invariáveis e os três últimos variáveis: “all data has been labelled according to the following pattern: <0010efg>. This means that “efg” are the variables under analysis, and the first four digits (“0010”) remain constant” (idem, p.65). Etiquetou os Participantes e os Processos relacionados ao objeto de análise, da seguinte forma: <0010141> *Luís Gama, o explosivo advogado mulato de São Paulo, pioneiro da Abolição,*

<0010340> recomendava <0010142> à atenção de seu filho <0010143> dois livros: a Bíblia e A Vida de Jesus, de Renan.

O processamento do corpus ocorreu no *WordSmith Tools* (Scott, 1999) através das ferramentas do *WordList*, *Concord* e *Keyword*. Fernandes utilizou o *Keywords* para encontrar os nódulos a serem analisados e justificar sua análise; o *WordList Tools* para encontrar a frequência das palavras e identificar um possível nódulo e o *Concord* para realizar a concordância das etiquetas.

Observo uma organização do capítulo de acordo com a proposta de organização de Fernandes (2004), do mesmo modo que é realizado por Fleuri (2006) e Zuniga (2006), descrevendo o *corpus design*, *corpus building* e *corpus processing*. Apesar de outras pesquisas mencionarem ter anotado o corpus, tornando suas partes explícitas, como o <título> ou o início <paragrafo> e o final de um </paragrafo>, esta é a primeira vez que o *TEI (Text Encoding Initiative)* (cf. Leech, 2005) foi mencionado. Sua forma de rotulação dos objetos de análise se assemelha ao aplicado por Fleuri (2006).

4.1.19. Thiago Pires (2009)

Quadro 31: Resumo das características e procedimentos metodológicos em Pires (2009)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2009/UFSC), “The construal of bishop’s ideational profile in flores raras e banalíssimas and rare and commonplace flowers: a corpus-based translation study”
Objetivo geral e Metafunção	Analisar a representação do perfil ideacional de Bishop (Metafunção Ideacional).
Objeto de análise	<i>Bishop</i> , núcleo de grupo nominal e Participante de Processo
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue de pequena dimensão • Versão digitalizada e corrigida.
Método de compilação	<ul style="list-style-type: none"> • Converter o texto digitalizado a <i>Plain Text</i>; • Preparar o corpus para ser alinhado • Alinhar o corpus com o auxílio do <i>Viewer and Aligner</i> • Anotar os objetos de análise com CROSF14

Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> (Scott, 1999) • <i>Concord</i> • <i>Viewer and Aligner</i> • Contagem é realizada no <i>Concord</i>
Códigos de anotação	Anotou o corpus com o CROSF14

A dissertação de Pires (2009) foi defendida na Universidade Federal de Santa Catarina (UFSC), sob o título “The construal of bishop’s ideational profile in flores raras e banalíssimas and rare and commonplace flowers: a corpus-based translation study”. Tem como objetivo geral analisar a representação do perfil ideacional de Bishop no corpus adotado. Assim, seu objeto de análise é Bishop, núcleo de grupo nominal e Participante de Processo, sob a perspectiva da Metafunção Ideacional.

Para desenvolver esta pesquisa, Pires utiliza um corpus paralelo bilíngue de pequena dimensão, composto pelo texto-original em português *Flores Raras e Banalíssimas* (L.C. Oliveira) e sua tradução para o inglês *Rare and commonplace flowers* (Trad. N. K. Besner), os quais se encontravam inicialmente disponível ao pesquisador no formato digitalizado e corrigido. A compilação consistiu em converter o texto digitalizado a *Plain Text*, preparar o corpus para ser alinhado (eliminando pontos não marcadores de final de sentença e juntou e separou frases que estavam juntas ou separadas no texto-original em relação a sua tradução), alinhar o corpus com o auxílio do *Viewer and Aligner* e anotar os objetos de análise com CROSF14, posicionando-o antes do objeto anotado.

Pires não rotulou apenas os Participantes, como também os Processos, posicionando os rótulos antes do objeto. Manteve os quatro primeiros números invariáveis e os três últimos variando conforme a análise ia sendo realizada (como em Fleuri, 2006 e Fernandes, 2009). Exemplo: *Com gosto, <0010111> BISHOP <0010310> escreveu no alto da folha suas duas primeiras palavras em português, duas sólidas e sonoras palavras que <0010161> a <0010360> situavam no planeta.* Anotou com o símbolo de vazio Ø as elipses consideradas objetos de análise (Participante elíptico, parte da corrente coesiva do complexo oracional): *BISHOP estava deliciada, Ø queria parar, Ø saltar do carro, mas Ø era acanhada demais para Ø pedir.*

O processamento do corpus ocorreu no *WordSmith Tools* (Scott, 1999) através da ferramenta do *Concord* para realizar a concordância das etiquetas e obter a quantificação dos dados. Após a concordância dos rótulos, os processos da textualização foram

organizados em tabela, cada tipo colorido de uma cor diferente, ao lado dos verbos, também coloridos, da retextualização. Isso foi feito para uma última revisão da análise e visualização das novas configurações.

Observo que Pires segue a organização de Fernandes (2004) para a descrição do capítulo de metodologia, a ela é incluída uma quarta seção chamada “Manual Corpus Analysis”, em que delimita detalhes do objeto e a unidade de análise em relação ao processo de rotulação. Em razão dos objetivos de pesquisa serem semelhantes, a metodologia empregada em Pires (2009) se parece muito a de Morinaka (2005), Zuniga (2006), Fleuri (2006) e Fernandes (2009).

4.1.20. Leonardo Nunes (2010)

Quadro 32: Resumo das características e procedimentos metodológicos em Nunes (2010)

Categorias	Características e procedimentos Metodológicos
Informações gerais (ano/universidade) Título	(2010/UFGM), “As conjunções <i>but</i> e <i>mas</i> em textos ficcionais originais e traduzidos: uma análise tridimensional com base na linguística sistêmico-funcional”
Objetivo geral e Metafunção	Examinar um elemento do sistema de COESÃO do português brasileiro por uma perspectiva de comparação entre sistemas linguísticos (Não específica Metafunção).
Objeto de análise	Conjunções <i>but</i> e <i>mas</i> e um total de 26 categorias de conjunções internas textuais e internas interpessoais
O corpus • Tipo de corpus • Versão inicial do texto	<ul style="list-style-type: none"> • Corpus paralelo bilíngue e comparável monolíngue • Não especifica a versão original dos textos
Método de compilação	<ul style="list-style-type: none"> • Apesar de mencionar que a compilação foi realizada conforme alguns autores, o pesquisador não descreve os processos de compilação. • O corpus paralelo foi alinhamento manualmente por parágrafos. • Anotou o corpus manualmente.
Método de processamento	<ul style="list-style-type: none"> • <i>WordSmith Tools</i> • <i>WordList</i> • <i>Concord</i> • Contagem é realizada no <i>WordList</i> e <i>Concord</i>.

Códigos de anotação	Anotou o corpus com o tagset Susanne (SAMPSON, 1995)
----------------------------	--

A dissertação de Nunes (2010) foi defendida na Universidade Federal de Minas Gerais (UFMG), sob o título “As conjunções *but* e *mas* em textos ficcionais originais e traduzidos: uma análise tridimensional com base na linguística sistêmico-funcional”. Tem como objetivo geral “examinar um elemento do sistema de COESÃO do português brasileiro por uma perspectiva de comparação entre sistemas linguísticos, com subsídios da linguística de *corpus* e embasamento sistêmico-funcional” (Nunes, 2010, p.12). Assim, seu objeto de análise são as conjunções *but* e *mas* e um total de 26 categorias de conjunções internas textuais (Martin e Rose, 2003) e internas interpessoais (Thompson, 2005). Nunes identifica a qual Metafunção tais categorias pertencem, englobando as três Metafunções: Ideacional, Textual e Interpessoal.

Para desenvolver esta pesquisa, Nunes utiliza um corpus paralelo bilíngue e comparável monolíngue. O corpus paralelo é composto por 10 obras ficcionais originais em inglês e suas respectivas traduções para o português. E o comparável monolíngue por 10 obras traduzidas e 10 originais para/em português. A produção e tradução das obras datam de 1990 a 2008. Não especifica em que versão se encontrava os textos inicialmente, se impressos ou digitalizados. Segundo Nunes, o corpus “foi compilado seguindo um dos registros presentes no *Corpus Brasileiro de Língua Portuguesa - KLAPT!* - desenvolvido a partir do projeto CroCo, idealizado na Universidade do Sarre, Alemanha” e “a compilação de ambos os *corpora* foram baseados nos critérios de compilação de (BIBER, 1990) e (BIBER, 1993)” (Nunes, 2010, p.60-62). O corpus paralelo foi alinhamento manualmente por parágrafos. A partir deste alinhamento, selecionou-se manualmente as orações e complexos oracionais contendo as conjunções *but* e *mas*. Inseriram-se manualmente as etiquetas de classificação nestas conjunções do tagset Susanne (SAMPSON, 1995), posicionada depois do objeto. Exemplo: *I could not lift that much cash, Aringarosa thought, closing the case. “Bonds are negotiable as cash. You said so yourself.” The cardinals exchanged uneasy looks, and finally one said, “Yes, but <CCI> these bonds are traceable directly to the Vatican Bank.”*

O processamento do corpus ocorreu no *WordSmith Tools* (não especifica a versão usada) através do *WordList*, e *Concord*. A quantificação dos objetos de análise se deu em ambas ferramentas.

Observo a menção, original em relação às outras pesquisas, dos critérios de compilação do *Corpus Brasileiro de Língua Portuguesa* -

KLAPT! e dos critérios de compilação de Douglas Biber (1990, 1993). Entretanto tais critérios não são explicados ou detalhados, cabendo ao leitor buscá-los na fonte, em caso de buscar uma forma de reproduzir as etapas metodológicas destes critérios. Pela primeira vez é mencionado a utilização do código de rotulação do tagset Susanne (Sampon, 1995), descrito por Leech (2005), citado no Capítulo 2 de Referencial Teórico desta tese.

4.2. ANÁLISE GERAL DOS MÉTODOS DAS DISSERTAÇÕES EM ESFTBC

Conforme visto no Capítulo 2, de Referencial Teórico, seção 2.2 desta tese, Barnbrook (1996, p. 5-20) argumenta que o desenho de um corpus, sua compilação e processamento se respaldam no projeto de pesquisa e em seus objetivos. A análise da metodologia de cada dissertação foi pautada neste argumento, desta forma, foram destacados nas seções anteriores deste capítulo elementos referentes ao desenho do corpus, à compilação (como os procedimentos de digitalização, correção, alinhamento e anotação e/ou marcação do corpus) e ao processamento (como os programas e ferramentas envolvidos em cada etapa de processamento do corpus). Nesta seção, oferece-se um balanço geral das dissertações, buscando traçar um perfil metodológico das dissertações em ESFTBC publicadas no Brasil até 2010, com o intuito de auxiliar a construção da proposta metodológica, exposta no próximo capítulo desta tese. Considerando que o objetivo da pesquisa influencia o desenho e o processamento do corpus (Barnbrook, 1996), observam-se os elementos relacionados ao (i) **objetivo das pesquisas em ESFTBC** que estão relacionados com aspectos da LSF, como a Metafunção: 4.2.1 as Metafunções analisadas em relação ao objetivo geral da pesquisa e 4.2.2 os objetos de análise. Em seguida, passa-se aos detalhes do (ii) **desenho do corpus**, como: 4.2.3 a versão original dos textos que compõem o corpus e 4.2.4 o tipo de corpus. Conforme Barnbrook (idem), o (iii) **método de compilação** e o (iv) **método de processamento** do corpus determinam o tipo de programa adotado e devem ser definidos na fase de projeto da pesquisa. Referentes ao método de compilação, as seguintes subseções tratam de 4.2.5 o método de alinhamento dos corpora paralelos e 4.2.6 o código de anotação e marcação. Referentes ao método de processamento, as subseções tratam de 4.2.7 os programas utilizados, 4.2.8 as ferramentas do *WST*, em caso de pesquisas que utilizaram este programa, e 4.2.9 o método de quantificação dos dados, no caso das

pesquisas que realizaram análise quantitativa. Em cada subseção são expostos em Quadros uma tabela criada no *Excel* e os gráficos gerados a partir desta tabela, seguidos de um comentário.

4.2.1. Objetivo da pesquisa: a Metafunção

São quantificadas as pesquisas que investigam apenas a Metafunção Ideacional, apenas a Textual, a Ideacional e Textual e outros aspectos. De um total de 20 respostas, uma por pesquisa, observou-se o seguinte perfil:

Quadro 33: Balanço das Metafunções exploradas nas dissertações em ESFTBC.

	METAFUNÇÃO				Total
	IDEACIONAL	TEXTUAL	IDEACIONAL E TEXTUAL	OUTRAS	
2003, Osilene Cruz (UFMG); Harry Potter	1				20
2003, Cristina Mauri (UFMG); Um estudo da	1				
2004, Silvana de Jesus (UFMG);	1				
2004, Roberto Assis (UFMG); A	1				
2005, Leticia Bueno (UFMG); Transitividade,	1				
2005, Roberta Rodrigues (UFMG); A		1			
2005, Eliza Morinaka (UFSC); Gabriela,			1		
2005, Tassiani Cançado (UFMG);	1				
2005, Marcos Feitosa (UFMG); Uma proposta		1			
2005, Viviane Paquillin (UFSC); The various		1			
2006, Gleimara Zaniga (UFSC); Construing	1				
2006, Lillian Fleuri (UFSC); O perfil	1				
2006, Ariana Zanella (UFSC); Mapeamento	1				
2007, Daniel Alves (UFMG); Aspectos da	1				
2007, Giacomo Figueredo (UFMG); Uma				1	
2007, Cristiano Araújo (UFMG); O sistema	1				
2007, Joanna Filgueiras (UFSC); Capocira				1	
2009, Alinne Fernandes (UFSC); Black into	1				
2009, Thiago Pires (UFSC); The construal of	1				
2010, Leonardo Nunes (UFMG); As				1	
total	13	3	1	3	20

Gráfico 2: Balanço das Metafunções exploradas nas dissertações em ESFTBC



A maioria (65%) das dissertações em ESFTBC explora a Metafunção Ideacional em suas análises, corroborando à análise realizada por Alves e Vasconcellos (no prelo). Aquelas que analisam a Metafunção Textual representam 15% das pesquisas (Rodrigues, 2005; Feitosa, 2005; Paquilin, 2005) e as que não analisam nenhuma Metafunções, ou seja, lidam com a análise de outros aspectos da Gramática Sistêmico-Funcional, representam 15% (Figueredo, 2007; Filgueiras, 2007; Nunes, 2010). Apenas uma dissertação (5%) (Morinaka, 2005) se propõe a analisar duas Metafunções simultaneamente (ideacional e textual).

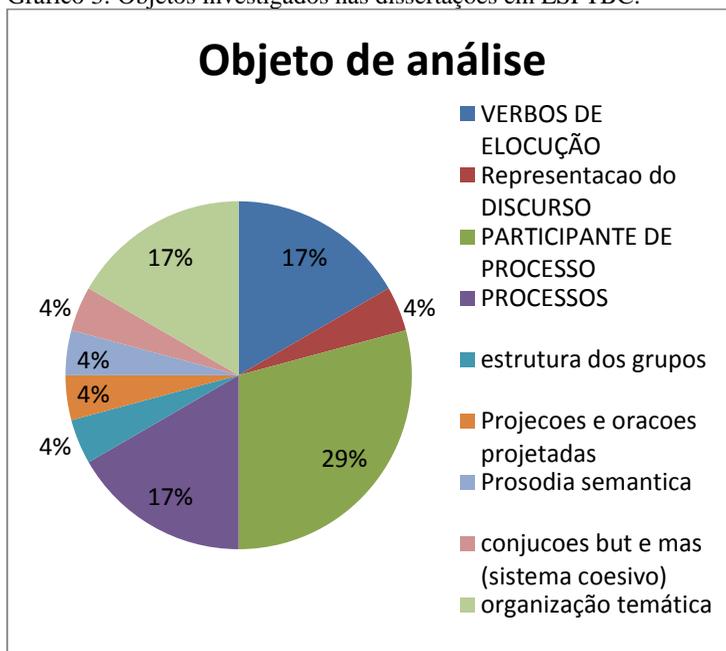
4.2.2. Objeto de análise

Esta subseção busca descrever o foco de estudo das pesquisas, o elemento analisado, anotado (em caso de ter sido rotulado) e/ou quantificado. O percentual foi calculado sobre um total de 24 postostas, pois houve pesquisas que estudaram mais de um dos objetos de análise apontados no gráfico.

Quadro 34: Objetos investigados nas dissertações em ESFTBC.

	OBJETO									Total
	VERBOS DE ELOCUÇÃO	Representação do DISCURSO	PARTICIPANTE DE PROCESSO	PROCESSOS	estrutura dos grupos	Projecoes e oracoes projetadas	Prosodia semantica	conjucoes but e mas (sistema coesivo)	organização temática	
2003, Osilene Cruz (UFMG);	1									
2003, Cristina Mauri (UFMG);	1									
2004, Silvana de Jesus (UFMG);	1									
2004, Roberto Assis (UFMG); A			1							
2005, Letícia Bueno (UFMG);				1						
2005, Roberta Rodrigues										1
2005, Eliza Morinaka (UFSC);			1							1
2005, Tassiani Cancado (UFMG);	1									
2005, Marcos Feitosa (UFMG);	1									1
2005, Viviane Paquillin (UFSC);										1
2006, Gleimara Zuniga (UFSC);			1							
2006, Lillian Fleuri (UFSC); O			1	1						
2006, Ariana Zanella (UFSC);			1	1						
2007, Daniel Alves (UFMG);		1								
2007, Giacomo Figueredo					1					
2007, Cristiano Araújo (UFMG);						1				
2007, Joanna Figueiras (UFSC);							1			
2009, Aline Fernandes (UFSC);			1							
2009, Thiago Pires (UFSC); The			1	1						
2010, Leonardo Nunes (UFMG);								1		
total	4	1	7	4	1	1	1	1	1	4

Gráfico 3: Objetos investigados nas dissertações em ESFTBC.



Os objetos de análise mais estudados foram a análise de: Participantes de Processo (29%), verbos de elocução (17%), Processos

(17%) e organização temática (17%). O restante dos objetos foram cada um estudado em uma pesquisa. Grande parte das pesquisas (29%) analisaram a representação e a construção de Participantes de Processos, como é o caso de: Assis (2004) que analisa a construção da personagem *Sethe*; Morinaka (2005) com a personagem *Gabriela*; Zuniga (2006) e Fleuri (2006) com o item lexical *tradutor* e *translator*; Zanella (2006) com os Participantes e Processos na estrutura de resumos e abstracts; Fernandes (2009) com as configurações de *mestiço*; e Pires (2009) com a personagem *Bishop*. Das pesquisas que trabalharam com tal objeto de análise, apenas Roberto Assis afiliava-se a UFMG, os restantes vinham da UFSC. Das quatro pesquisas (17%) que analisaram Processos, três analisaram também Participantes de Processo (sendo já citados), apenas Bueno (2005) analisa somente Processos. Das 4 pesquisas que analisam a organização Temática, as que se dedicam exclusivamente a este objeto de análise são 3: as pesquisas de Rodrigues (2005), Feitosa (2005) e Paquilin (2005). Os verbos de elocução são foco de outros 17% das pesquisas (Cruz, 2003; Mauri, 2003; Jesus, 2003; e Cançado, 2005).

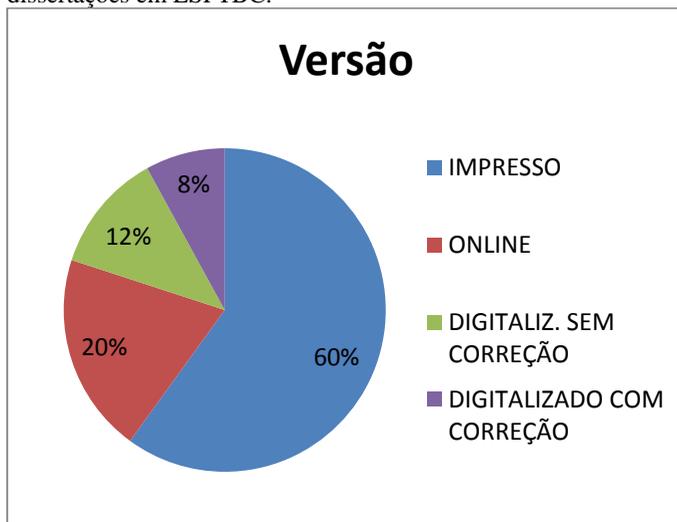
4.2.3. O corpus: versão original dos textos

Os textos que compõem os corpora das pesquisas chegam inicialmente às mãos do pesquisador nas seguintes versões: impressa, digitalizada não corrigida ou digitalizada corrigida. Não há menção explícita por parte dos pesquisadores quanto ao fato de terem trabalhado ou não com corpus previamente anotado ou com a versão em PDF. O percentual é calculado sobre um total de 25 respostas, visto que houve pesquisadores que obtiveram parte do corpus impresso e outra parte digitalizada ou online.

Quadro 35: A versão original dos textos que compuseram os corpora nas dissertações em ESFTBC.

	VERSÃO				total
	IMPRESSO	ONLINE	DIGITALIZ. SEM CORREÇÃO	DIGITALIZAD O COM CORREÇÃO	
2003, Osilene Cruz (UFMG); Harry Potter and the	1				25
2003, Cristina Mauri (UFMG); Um estudo da	1				
2004, Silvana de Jesus (UFMG); Representação do	1				
2004, Roberto Assis (UFMG); A Transitividade na			1		
2005, Letícia Bueno (UFMG); Transitividade,		1	1		
2005, Roberta Rodrigues (UFMG); A Organização	1				
2005, Eliza Morinaka (UFSC); Gabriela, cravo e	1	1			
2005, Tassiani Cançado (UFMG); Transitividade e	1				
2005, Marcos Feitosa (UFMG); Uma proposta de	1				
2005, Viviane Paquilin (UFSC); The various facets	1				
2006, Gleimara Zuniga (UFSC); Construing the	1		1		
2006, Lilian Fleuri (UFSC); O perfil ideacional dos	1				
2006, Ariana Zanella (UFSC); Mapeamento macro		1			
2007, Daniel Alves (UFMG); Aspectos da				1	
2007, Giacomo Figueredo (UFMG); Uma descrição	1	1			
2007, Cristiano Araújo (UFMG); O sistema	1	1			
2007, Joanna Filgueiras (UFSC); Capoeira em	1				
2009, Alinne Fernandes (UFSC); Black into white	1				
2009, Thiago Pires (UFSC); The construal of				1	
2010, Leonardo Nunes (UFMG); As conjunções but	1				
total	15	5	3	2	25

Gráfico 4: A versão original dos textos que compuseram os corpora nas dissertações em ESFTBC.



A maioria dos pesquisadores (60%) lidou com textos impressos durante a compilação do corpus. Cinco pesquisadores (representando 20%) encontraram seu corpus online. Neste caso, considerou-se corpus online, textos disponíveis online que foram copiados e colados em documento *Plain Text* ou *Document Word* e os corpora de referência, como foi o caso do *Lácio-Web* entre outros. Apenas Zanella (2006) construiu seu corpus a partir de textos online, o restante consulta o corpus de referência online como fonte alternativa e complementar de pesquisa (Bueno, 2005; Morinaka, 2005; Figueredo, 2007; Araújo, 2007). Três pesquisadores, ou seja, 12%, tiveram seus textos já digitalizados, mas não corrigidos, foi o caso de: de Bueno (2005), Assis (2004) e Zuniga (2006). Os corpora de Bueno e de Assis se encontravam disponibilizados pelo projeto CORDIALL, mas ainda apresentavam erros de leitura do escâner que teve que ser corrigido. Zuniga (2006) conseguiu apenas a tradução com o próprio tradutor em formato digitalizado, como era uma versão não publicada, uma leitura cuidadosa teve que ser feita para identificar possíveis diferenças textuais com a versão publicada e impressa. Apenas 8% das pesquisas não tiveram que passar pelo processo de construção do corpus, resumindo suas etapas a conversão do texto em *Plain Text*, preparação para alinhamento e anotação (Alves, 2007 e Pires, 2009).

Os métodos acessados para a compilação do corpus variava de acordo com a versão inicial dos textos. Textos impressos demandam mais tempo e trabalho de correção dos erros de escaneamento que textos em PDF e estes demandam mais que textos em *Word* ou *Plain Text* (ver seção 2.2.1.2 desta tese). Todos os 60% dos pesquisadores citaram alguma forma de trabalho e tempo de pesquisa gasto na digitalização e compilação do corpus⁷³.

4.2.4. O Corpus: tipo

Tomando como parâmetro as classificações discutidas no capítulo de Referencial Teórico, seção 2.2.1.1 desta tese (cf. Olohan, 2004

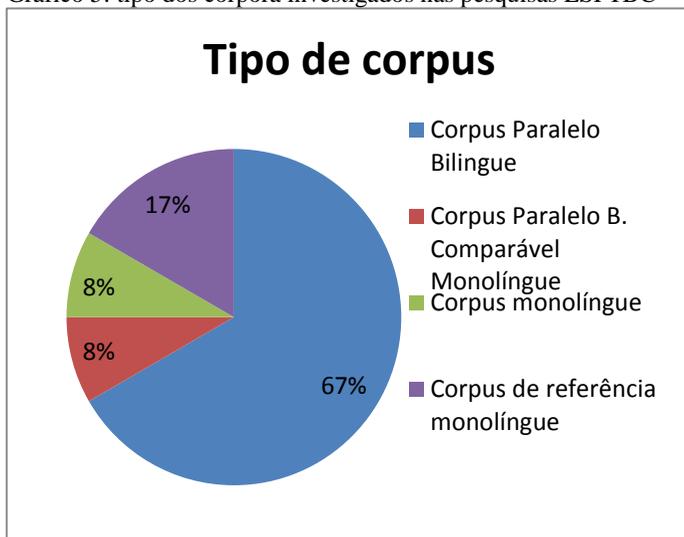
⁷³ A exemplo, expomos a fala de Cruz (2003): “Esta primeira etapa consistiu em digitalizar os romances através de “scanner” e salvá-los em formato “txt”. Muita atenção foi demandada na conferência da digitalização porque, durante o processo, ocorrem muitas alterações no corpus, dentre elas, omissão de palavras, orações e até mesmo parágrafos; apresentação dos parágrafos em formato de colunas; palavras com letras trocadas. A revisão do processo de digitalização foi fundamental e, para maior confiabilidade, uma releitura dos romances, em sua totalidade, foi realizada”.

e Fernandes, 2006), um corpus, em ETBC, pode ser classificado entre corpus paralelo e corpus comparável e se relacionar com aspectos referentes ao tamanho, número de idiomas envolvidos, etc. Observou-se a ocorrência tanto de corpora paralelos como comparáveis. Quanto ao tamanho emergiram dois tipos de corpus, de pequena e de grande dimensão, com predominância do primeiro. Dos corpora de grande dimensão, foram acessados corpora de referência monolíngues, tais como o *Lácio-Web* (em português) e o *The British National Corpus* (em inglês) —entre outros. Quanto ao número de línguas envolvidas, houve a ocorrência dos seguintes tipos: paralelo bilíngue, paralelo bilíngue e comparável monolíngue e comparável bilíngue. Os idiomas envolvidos foram: português, inglês e italiano. O percentual foi calculado sobre um total de 24 respostas, pois houve trabalhos que lidaram com mais de um tipo de corpus – no caso dos que consultaram também um ou mais corpora de referência, por exemplo.

Quadro 36: tipo dos corpora investigados nas pesquisas ESFTBC

	TIPO DE CORPUS				total
	Corpus Paralelo Bilingue	Corpus Paralelo B. Comparável Monolíngue	Corpus comparável Monolíngue	Corpus de referência monolíngue	
2003, Osilene Cruz (UFMG); Harry	1				
2003, Cristina Mauri (UFMG); Um estudo	1				
2004, Silvana de Jesus (UFMG);			1		
2004, Roberto Assis (UFMG); A	1				
2005, Leticia Bueno (UFMG);	1				1
2005, Roberta Rodrigues (UFMG); A	1				
2005, Eliza Morinaka (UFS C); Gabriela,	1				1
2005, Tassiani Caçado (UFMG);	1				
2005, Marcos Feitosa (UFMG); Uma	1				
2005, Viviane Paquiliin (UFS C); The	1				
2006, Gleimara Zuniga (UFS C);	1				
2006, Lilian Fleuri (UFS C); O perfil	1				
2006, Ariana Zanella (UFS C);	1				
2007, Daniel Alves (UFMG); Aspectos da	1				
2007, Giacomo Figueredo (UFMG); Uma				1	1
2007, Cristiano Araújo (UFMG); O				1	1
2007, Joanna Filgueiras (UFS C);	1				
2009, Alinne Fernandes (UFS C); Black	1				
2009, Thiago Pires (UFS C); The construal	1				
2010, Leonardo Nunes (UFMG); As			1		
total	16		2	2	4
					24

Gráfico 5: tipo dos corpora investigados nas pesquisas ESFTBC



A maioria (67%) realiza a análise em corpus paralelo bilíngue. Corpora de referência são usados por 17% dos pesquisadores. Em Bueno (2005) e Morinaka (2005) esse corpus de referência foi utilizado na fase exploratória da pesquisa, enquanto em Araújo (2007) e Figueredo (2007) esse corpus foi utilizado para complementar o corpus compilado, fornecendo mais informações as suas pesquisas. Como adotaram um direcionamento aos estudos linguísticos da tradução, as pesquisas de Araújo (2007) e de Figueredo (2007) analisam também corpora monolíngues, representando 8% das pesquisas em ESFTBC. Apenas Jesus (2003) e Nunes (2010) realizam pesquisas em dois tipos de corpus, paralelo bilíngue e comparável monolíngue, representando os outros 8%.

A escolha por trabalhar em um corpus de pequena dimensão, característico nos trabalhos brasileiros em ESFTBC é justificada por maior parte das pesquisas pelo fato de optarem em explorar os aspectos discursivos dos textos, como argumenta Cançado:

Optamos por corpus de pequenas dimensões, em vez de grandes corpora, devido, principalmente, à necessidade levantada pelos pesquisadores do CORDIAL de uma maior investigação dos aspectos discursivos e cognitivos dos textos de menor escala, onde a prioridade encontra-se não na busca da objetividade e exaustividade, mas de

crítica e questionamento na análise dos fenômenos linguísticos (Cançado, 2005).

Entretanto, a tarefa de compilação do corpus e de obtenção dos dados se tornou tão “árdua” (Cruz, 2003) que em muitos casos a profundidade da “crítica e questionamento na análise dos fenômenos linguísticos” (idem, ibidem) foi afetada pelo elemento tempo, cedendo espaço à “objetividade e exaustividade” (idem, ibidem).

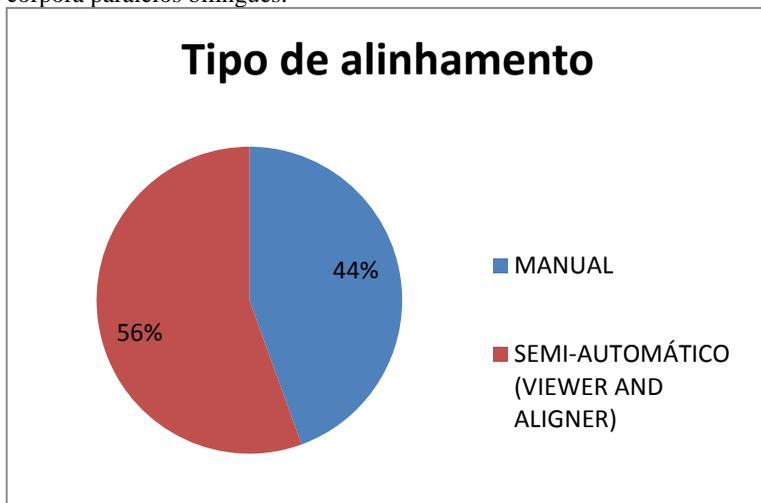
4.2.5. Método de Compilação: alinhamento

A maioria das pesquisas operaram com corpus paralelo bilíngue e, portanto, optaram por alinhar o corpus. Foram 18 dissertação que trabalharam com corpus alinhados e tiveram que passar por esse processo metodológico (somente Figueiredo (2007) e Araújo (2007) não alinharam seus corpora).

Quadro 37: o método de alinhamento nas pesquisas em ESFTBC que utilizaram corpora paralelos bilíngues.

	TIPO DE ALINHAMENTO		total
	MANUAL	SEMI-	
2003, Osilene Cruz (UFMG); Harry Potter and the			1
2003, Cristina Mauri (UFMG); Um estudo da	1	0	
2004, Silvana de Jesus (UFMG); Representação do			1
2004, Roberto Assis (UFMG); A Transitividade na			1
2005, Leticia Bueno (UFMG); Transitividade,	1		
2005, Roberta Rodrigues (UFMG); A Organização	1	0	
2005, Eliza Morinaka (UFSC); Gabriela, cravo e			1
2005, Tassiani Cançado (UFMG); Transitividade e			1
2005, Marcos Feitosa (UFMG); Uma proposta de	0	1	
2005, Viviane Paquillin (UFSC); The various facets			1
2006, Gleimara Zuniga (UFSC); Construing the	1		
2006, Lilian Fleuri (UFSC); O perfil ideacional dos			1
2006, Ariana Zanella (UFSC); Mapeamento macro	1		
2007, Daniel Alves (UFMG); Aspectos da	1	0	
2007, Giacomo Figueiredo (UFMG); Uma descrição	0	0	
2007, Cristiano Araújo (UFMG); O sistema	0	0	
2007, Joanna Filgueiras (UFSC); Capoeira em	1	0	
2009, Alinne Fernandes (UFSC); Black into white			1
2009, Thiago Pires (UFSC); The construal of			1
2010, Leonardo Nunes (UFMG); As conjunções but	1		
total	8	10	18

Gráfico 6: o método de alinhamento nas pesquisas em ESFTBC que utilizaram corpora paralelos bilíngues.



Pelo fato de praticamente todos os pesquisadores demonstrarem ter familiaridade com o *WordSmith Tools*, esperar-se-ia que muito mais da metade dos pesquisadores optasse por alinhar seu corpus a partir da ferramenta *Viewer and Aligner*. No entanto, o grande esforço aplicado no processo de preparação do corpus para o alinhamento automático levou a um grande número de pesquisas optar por realizar o alinhamento manual. A diferença percentual de pesquisas que optaram pelo alinhamento semiautomático variou pouco das que optaram pelo alinhamento manual, sendo aquele representado por 56% das pesquisas que alinharam e este por 44%, com a diferença numérica de uma pesquisa a mais que realizou o alinhamento semiautomático.

De fato o esforço empregado no alinhamento não parece diferenciar muito entre o manual do semiautomático. Quem optou por alinhar no *WST* relata que a intervenção do pesquisador é ainda grande, transformando essa tarefa em um trabalho “árduo” e longe de ser completamente automático, como relata Cruz:

Apesar de tal ferramenta [*Viewer and Aligner*] realizar o alinhamento entre os textos fonte e traduzido, emparelhando as orações ou períodos, tal tarefa precisa da interferência do pesquisador porque, diante de diferentes decisões da autora e da tradutora ou de aspectos estritamente linguísticos, como pontuação diferente, tamanho distinto de

orações, omissões ou acréscimos de palavras ou locuções, o tamanho dos períodos em ambos os textos diverge. (...) Esta etapa é árdua porque, conforme explicado acima, devido à disparidade no tamanho das orações, a diferença no alinhamento é certa, demandando do pesquisador o ajuste manual do texto traduzido ao texto fonte. (Cruz, 2003)

Já os que optaram pelo o alinhamento manual alegaram que o trabalho de correção dos textos para serem alinhados automaticamente era tão grande quanto ou maior que alinhar os textos manualmente. Os que optaram realizar o alinhamento manual em geral escolheram alinhar por parágrafo—na tentativa de diminuir o trabalho, enquanto os que utilizaram o *WST* para o alinhamento alinharam por sentenças. Bueno (2005) relata sua tarefa e sua escolha pelo alinhamento manual:

optei por fazer um alinhamento dos dois textos com o programa *Word*. O esforço para alinhá-los foi grande, principalmente devido a questões inerentes a qualquer retextualização, em que não se espera uma “correspondência” exata de orações e períodos entre os dois textos. Assim, muitas vezes, orações separadas foram aglutinadas em uma só, ou o contrário. Optei por um alinhamento mais estruturado em função dos parágrafos, uma unidade maior de tradução. Ainda assim, em diversos momentos, houve uma aglutinação de mais de um parágrafo ou uma fragmentação de um mesmo parágrafo em dois ou mais. A ferramenta *Aligner* disponibilizada pelo software, embora poupe muito do esforço que o alinhamento manual exige, também oferece este mesmo problema de correspondência, que deve ser resolvido pela intervenção manual do pesquisador (Bueno, 2005)

Conforme Bueno (idem) o alinhamento semiautomático no *Viewer and Aligner* poupa esforços no que se refere à tarefa de “copiar e colar” do alinhamento manual, mas exige um esforço extra no que abrange questões de preparação do texto para o alinhamento automático (como é o caso da eliminação dos pontos não marcadores de final de sentença e da falta de correspondência entre as sentenças do texto original com o texto traduzido).

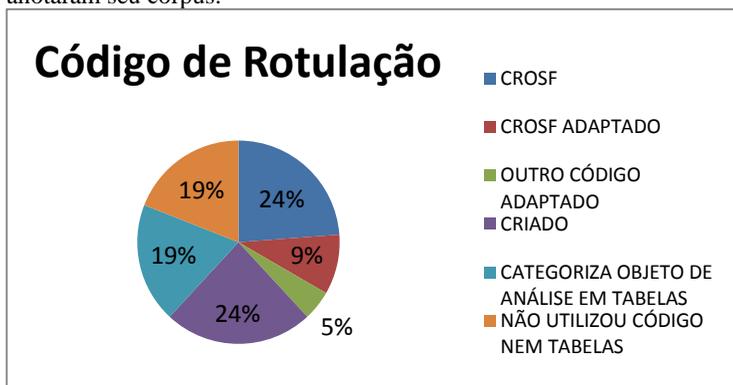
4.2.6. Código de anotação

Foram identificadas pesquisas que etiquetaram utilizando apenas o CROSF; o CROSF adaptado; outro código adaptado; um código criado pelo próprio pesquisador; tabelas organizando os itens classificados e a classificação; e os que não utilizaram códigos nem tabelas. O percentual é apresentado sobre um total de 21 respostas, pois Alves (2007) rotula o corpus com um código criado por ele para cobrir um propósito de sua análise e classifica em tabela, cobrindo outro propósito.

Quadro 38: códigos de anotação utilizados nas pesquisas em ESFTBC que anotaram seu corpus.

	CÓDIGO DE ROTULAÇÃO						total
	CROSF	CROSF ADAPTADO	OUTRO CÓDIGO ADAPTADO	CRIADO	CATEGORIZA OBJETO DE ANÁLISE EM TABELAS	NÃO UTILIZOU CÓDIGO	
2003, Osilene Cruz (UFMG); Harry Potter and the					1		
2003, Cristina Mauri (UFMG); Um estudo da					1		
2004, Silvana de Jesus (UFMG); Representação do				1			
2004, Roberto Assis (UFMG); A Transitividade na						1	
2005, Letícia Bueno (UFMG); Transitividade,						1	
2005, Roberta Rodrigues (UFMG); A Organização				1			
2005, Eliza Morinaka (UFSC); Gabriela, cravo e			1				
2005, Tassiani Cançado (UFMG); Transitividade e					1		
2005, Marcos Feitosa (UFMG); Uma proposta de		1					
2005, Viviane Paquillin (UFSC); The various facets		1					
2006, Gleimara Zuniga (UFSC); Construing the		1					
2006, Lilian Fleuri (UFSC); O perfil ideacional dos			1				
2006, Ariana Zanella (UFSC); Mapeamento macro						1	
2007, Daniel Alves (UFMG); Aspectos da				1	1		
2007, Giacomo Figueredo (UFMG); Uma descrição				1			
2007, Cristiano Araújo (UFMG); O sistema				1			
2007, Joanna Filgueiras (UFSC); Capoeira em						1	
2009, Alinne Fernandes (UFSC); Black into white		1					
2009, Thiago Pires (UFSC); The construal of		1					
2010, Leonardo Nunes (UFMG); As conjunções but				1			
total	5	2	1	5	4	4	21

Gráfico 7: códigos de anotação utilizados nas pesquisas em ESFTBC que anotaram seu corpus.



O Gráfico acima mostra que 19% das pesquisas optaram por usar de tabelas e outros 19% por não anotar nem usar tabelas, somando um total de 38% que não anotaram o corpus. A maioria, ou seja, 62% das pesquisas optaram por anotar o corpus: 24% utiliza o CROSF, 9% utiliza o CROSF adaptado, em geral acrescentando letras ou palavras após o sétimo dígito ou no lugar dos últimos 3 dígitos; 5% adapta outro código, de fato isso representa uma única pesquisa, a de Nunes (2010) que adapta o código *Tagset Susanne* proposto por Sampson (1995); e 24% cria seu próprio código.

Não importa a forma de codificação utilizada, todas as pesquisas provaram eficiência no modo de codificar e de classificar seu objeto de análise. Entretanto, do ponto de vista prático, Feitosa (2005) ressalta alguns aspectos que devem ser considerados no momento de elaborar um código, que são:

- O número de toques dados para se escrever o código (uma letra maiúscula, por exemplo, representa dois toques: Shift+letra), a extensão do código e a variação de caracteres (letras, números, símbolos, pontos, acentos, cedilha, etc.): muitos toques, diferentes caracteres e palavras grandes são mais propensas à erros de digitação e conseqüente alteração dos resultados.
- A explicação do código: para poder ser reutilizado, um código criado tem que ser bem detalhado no capítulo de metodologia, ou seja, todos os códigos teriam que ser explicados e expostos (a utilização de tabelas ou de uma lista de códigos de rotulação parece uma boa ideia) e o modo como foi utilizado deveria ser exemplificado, caso a caso. Deste modo, é possível resgatar a informação ou as etiquetas no corpus, citá-la e/ou reaplicá-la em outras pesquisas.
- A possibilidade de reutilização do código em outras pesquisas: um código mais abrangente e flexível possibilita a continuação de uma pesquisa por outros pesquisadores ou a utilização do mesmo código em outros trabalhos acadêmicos.

Acrescento às ressalvas de Feitosa dois outros aspectos para tornar o código reutilizável: os códigos devem ser autoexplicativos ou fáceis de serem aprendidos. O CROSF, com o apoio da tabela ao lado do computador, torna-se muito fácil de ser aprendido, apesar de não ser autoexplicativo. Um código autoexplicativo seria o utilizado por (EGGINS, 1994) (p.337), por exemplo, que escreve uma abreviação (curta, para obedecer a questão dos números de toques) de categorias da

LSF, por exemplo, <pm> para Processo Material, <sy> para Sayer, <C> para circunstância, que mesmo sem ler a explicação de seu código é possível adivinhar a classificação à qual o código se refere dentro do contexto.

Um total de cinco pesquisas, ou seja, 24% das pesquisas, criaram seu próprio código. A seguir vamos expor como se apresentam cada código e sua aplicabilidade segundo as ressalvas mencionadas de Feitosa (2005). Os pesquisadores que optaram por criar seus próprios códigos de anotação foram: Jesus (2004), Rodrigues (2005), Alves (2007), Figueredo (2007) e Araújo (2007).

Jesus (2004) optou por utilizar letras maiúsculas símbolo de igual e uma palavra escrita em letras minúsculas entre parênteses angulares, conforme o exemplo: <DVTO=pensar>. Como apontado, etiquetas como as criadas por Jesus (2004) são autoexplicativas, possibilitando que rapidamente alguém que desconheça o código possa entendê-lo e reutilizá-lo. Entretanto, a presença de letras maiúsculas, abreviação em maiúscula com símbolos e palavras em letra minúscula, demanda uma grande quantidade de toques (incluindo *Shift* e/ou *Caps Lock*), tornando-a muito propensa a sofrer erros de digitação e consequente alteração dos resultados durante sua coleta.

Rodrigues (2005) prefere escrever o nome da categoria por extenso em letras minúsculas, <tema textual> <tema ideacional participante>. Esta forma é autoexplicativa e fácil de ser identificada a olho nu no texto, não mistura símbolos ou maiúsculas e minúscula, facilitando a digitação da etiqueta. Entretanto segue havendo muitos toques, que aumentam as chances de erros de digitação.

Alves (2007) cria um modo de identificar as marcas de discurso através de etiquetas ao longo do texto, para, depois de encontrá-los automaticamente, poder separar as categorias em tabela. As etiquetas criadas foram: <trav>, <dois pontos>, <asp> e </asp>. Estas correspondem, com exceção de <dois pontos>, abreviações com máximo de 4 toques, escritas em letras minúsculas, facilitando a digitação e diminuindo as chances de erros de digitação.

Figueredo (2007) não rotula propriamente dito, ele anota o texto com informações, desta forma uma etiqueta se diferencia da outra, pois se trata de textos, autores, anos etc. diferentes. Exemplo: *As palmas estrugiram; e, dentro do grande salão iluminado, pareceu-lhe que recebia as palmas da Terra toda. [Texto <Sua Excelência> <Lima Barreto> <1998> <nota: data da segunda edição do livro Histórias Fantásticas].*

Araújo (2007) marca o texto com símbolos e rotula o objeto analisado com uma palavra referente à categoria, da seguinte forma: *Quando é submetida a uma revista especializada, | a pesquisa é analisada por cientistas renomados | que revisam o trabalho | em busca de possíveis falhas”, || conta <projeção> Átila da Rosa, paleontólogo da Universidade Federal de Santa Maria. |||*. Estes símbolos é uma forma de assinalar secções da frase importantes para a análise do autor (semelhante ao que faz Eggins, 1994, p.337), mas difíceis, se não impossíveis, de serem rastreados no WST. Como sua pesquisa, assim como a de Figueredo, não foram de cunho quantitativo, a forma de etiquetar era mesmo uma marca textual para auxiliar a visualização das secções dos textos e suas análises.

Quanto ao modo de explicação dos códigos criados Jesus (2004) e Morinaka (2005) nos oferecem um bom exemplo. A forma como Jesus (2004) explica seu código é bastante didática (apresenta todos em tabela, de acordo com o exemplo abaixo) e detalhada.

Quadro 39: código criado por Jesus (2004)

<think=pensar>	Re-textualização de processo mental realizado por THINK no original através de processo mental realizado por PENSAR no texto traduzido
<think=0>	Re-textualização de processo mental realizado por THINK no original através de verbalização sem processo mental ou com outros tipos de processo no texto traduzido
<0=pensar>	Re-textualização de passagem sem processo mental no original através de processo mental realizado por PENSAR no texto traduzido
<DVTO=pensar>	Re-textualização de processo mental realizado por verbo diferente de THINK no original através de processo mental realizado por PENSAR no texto traduzido
<DNTO=pensar>	Re-textualização de processo mental realizado por substantivo diferente de THINK no original através de processo mental realizado por PENSAR no texto traduzido
<think=DNTR>	Re-textualização de processo mental realizado por THINK no original através de processo mental realizado por substantivo diferente de THINK no texto traduzido
<think=DVTR>	Re-textualização de processo mental realizado por THINK no original através de processo

	mental realizado por verbo diferente de PENSAR no texto traduzido
<think=dizer>	Re-textualização de processo mental realizado por THINK no original através de processo verbal realizado por DIZER no texto traduzido

Fonte: Jesus (2004)

Este detalhamento das etiquetas promove a reutilização desta etiqueta por ser de rápida aprendizagem e por oferecer ao leitor detalhes de cada código. Fleuri (2006), Fernandes (2009) e Pires (2009), apesar de utilizar o CROSF, um código extensamente explicado na dissertação de Feitosa (2005), esse pesquisadores dedicam uma tabela para mostrar como o interpretaram e todas as categorias anotadas.

Morinaka (2005) expõe em anexo todos os exemplos ao final da dissertação, possibilitando a compreensão de detalhes de uso do código, como: o posicionamento da etiqueta em relação ao objeto analisado, os elementos rotulados e os não rotulados, a forma que interpretou cada elemento, etc.

Sete pesquisadores utilizaram o CROSF (adaptado ou não) para anotar o corpus, representando 33% dos casos observados. A proposta do CROSF foi criar um código que visasse a

agilizar o processo de anotação dos corpora dos pesquisadores do NET, a dirimir os erros de digitação durante esse processo e também a possibilitar uma maior interface entre as diversas pesquisas em tradução (...) [viabilizando] uma melhor anotação dos corpora nos termos da Gramática Sistemico-Funcional de Halliday e que permita um diálogo maior entre os que realizam este tipo de pesquisa, através de uma uniformidade de anotação que seja de conhecimento de todos os pesquisadores (Feitosa, 2005, p.14-15).

Entretanto as diferentes interpretações do código e necessidades emergentes não previstas pelo código levaram a um comprometimento da “uniformidade” na forma de anotação entre essas sete pesquisas. Houve dois casos que uma adaptação do código se viu necessária para responder a questões não previstas no código, é o caso de Fleuri (2006) e Morinaka (2005). Fleuri (2006) adiciona uma barra transversal “/” depois do 7º dígito para indicar se o elemento etiquetado (*tradutor* e *translator*) é modificador de um núcleo nominal ou se é modificador de um modificador. Essa adaptação aponta a necessidade de se etiquetar elementos de outra instância da Gramática Sistemico-

Funcional, do nível “de baixo”, representado pelos grupos nominais. Morinaka (2005) substitui os três últimos dígitos do CROSF pelas letras “abc” para identificar a palavra ou a referência à “*Gabriela*” não Participante de Processo nem de circunstância e substitui esses três algarismos pela palavra “minor” para indicar “minor clauses”: <0010abs> *Gabriela no caminho / Mas*, <0010minor> *Gabriela*. Sua adaptação aponta para a necessidade de se nomear outros elementos em que o objeto de análise não se apresenta como Participante.

Diferenças também ficaram evidenciadas na utilização do CROSF entre os pesquisadores que o utilizaram, principalmente no que se refere ao 5º dígito e aos três últimos. A quinta casa do CROSF pode ser anotado o número 1 para representar participante sem interpolação; 2, participante com interpolação; 3, processo; 4, processo-participante; 5, circunstância sem interpolação; 6, circunstância com interpolação; 7, “preposed attributive”; 8, oração; 9, estrutura tematizada ou 0 para nenhum dos itens mencionados. Nenhum dos pesquisadores que analisa Participantes de Processo considera para análise o fato de o Participante sofrer ou não interpolação. No entanto Morinaka (2005), Fleuri (2006), Pires (2009) e Fernandes (2009) marcam a 5ª casa com o número 1 para indicar simplesmente que se etiqueta um Participante (independente deste sofrer ou não interpolação). Exemplo: <0010121> Zuniga (2006) a marca com o número 0, pois considera que nenhum dos elementos mencionados por Feitosa no CROSF são considerados para sua análise, já que não lhe interessa saber se o Participante sofre ou não interpolação. Portanto ela mantém fixa em sua análise as cinco primeiras casas e modifica apenas as duas últimas, referentes ao par Processo-Participante, indicando apenas a tipologia do Participante rotulado. Exemplo: <0010021>. Paquilin (2005) é a única que utiliza o CROSF para etiquetar elementos da Metafunção Textual, além do próprio Feitosa. Diferentemente de Feitosa (2005) que enumera todas as casas do código—*Abri* <1111411> *a gaveta*, *vi* <1111421> *os cadernos*, Paquilin não anota os Participantes ou Processos, apenas tema e rema, optando por marcar com hífen essas três últimas casa, da seguinte forma: <1011---> **Hate the new Year*.

Não importa qual seja a forma de anotar um corpus ou qual seja o código adotado pelo pesquisador, o objetivo aqui é mostrar que para agilizar a etapa de anotação e promover sua reprodução e/ou continuação da sua pesquisa por outros pesquisadores é importante considerar os aspectos apontados por Feitosa (2005)—e discutidos acima, ao aplicar ou elaborar um código. Se a opção for criar um código próprio ou adaptar um código já criado, aconselha-se então explicar cada etiqueta, exemplificando a aplicação desta no corpus. É importante considerar o

número de toques (de preferência usar somente números ou letras minúsculas, para agilizar a digitação), de caracteres (por volta de cinco caracteres sem espaço parece ser mais fácil de digitar e tornando-se menos propenso a erros), e evitar códigos ambíguos ou de fácil confusão, como por exemplo <DNTR> e <DVTR>, cuja semelhança torna difícil a identificação imediata da diferença dos códigos gerada pela troca da letra N pela V. De acordo com Simone Bittencourt (s/d)⁷⁴ o mais importante na identificação de uma palavra é que a primeira e a última letra estejam no seu lugar, sendo essa uma palavra de menos de 8 letras, o resto pode estar fora de ordem que mesmo assim o cérebro vai identificá-la. Veja como funciona no exemplo:

De acordo com uma pesquisa de uma universidade inglesa, não importa em qual ordem as letras de uma palavra estão, a única coisa importante é que as primeiras e as últimas letras estejam no lugar certo. O resto pode ser uma bagunça que você pode ainda ler sem problema. Isso é porque nós não lemos cada letra isolada, mas a palavra como um todo

Portanto, no caso da elaboração do código é importante estar atento para evitar códigos semelhantes, como o caso mencionado do <DNTR> e <DVTR>, e/ou com muito toques (com mais 8 dígitos), pois o computador pode lê-las corretamente, mas a revisão da rotulação os erros podem passar despercebidos aos olhos do pesquisador, podendo alterar os resultados finais da pesquisa.

4.2.7. Método de Processamento: programa adotado

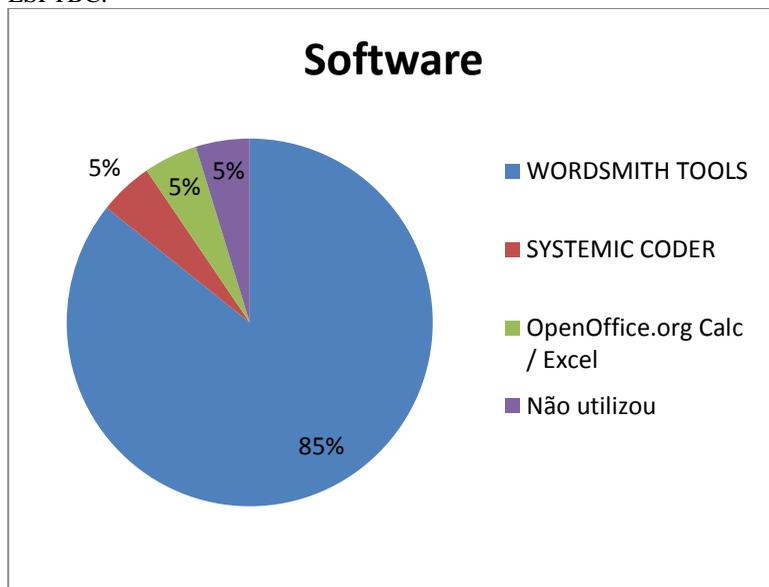
Além do uso do *WordSmith Tools*, houve também o uso de *Systemic Coder*, de Aplicativos de Planilha (MS *Excel*) ou de nenhum *software*, processando os dados manualmente. Os percentuais no gráfico abaixo são calculados sobre um total de 21 respostas, visto que Alves (2007) utiliza dois *software*.

⁷⁴ Fonte: <http://www.univesp.ensinosuperior.sp.gov.br/preunivesp/503/como-nosso-c-rebro-l-.html> (data 27/06/2012). Esta matéria cita os resultados de uma pesquisa desenvolvida na Universidade de Nottingham na Inglaterra.

Quadro 40: Programa utilizado no processamento dos corpora em pesquisas em ESFTBC.

	SOFTWARE				total
	WORDSMITH	SYSTEMIC	OpenOffice.	Não utilizou	
2003, Osilene Cruz (UFMG); Harry Potter and the	1				21
2003, Cristina Mauri (UFMG); Um estudo da	1				
2004, Silvana de Jesus (UFMG); Representação do	1				
2004, Roberto Assis (UFMG); A Transitividade na	1				
2005, Leticia Bueno (UFMG); Transitividade,	1				
2005, Roberta Rodrigues (UFMG); A Organização	1				
2005, Eliza Morinaka (UFSC); Gabriela, cravo e	1				
2005, Tassiani Cançado (UFMG); Transitividade e	1				
2005, Marcos Feitosa (UFMG); Uma proposta de	1				
2005, Viviane Paquillin (UFSC); The various facets	1				
2006, Gleimara Zuniga (UFSC); Construing the	1				
2006, Lilian Fleuri (UFSC); O perfil ideacional dos	1				
2006, Ariana Zanella (UFSC); Mapeamento macro		1			
2007, Daniel Alves (UFMG); Aspectos da	1		1		
2007, Giacomo Figueredo (UFMG); Uma descrição	0	0		1	
2007, Cristiano Araújo (UFMG); O sistema	1				
2007, Joanna Filgueiras (UFSC); Capoeira em	1				
2009, Alinne Fernandes (UFSC); Black into white	1				
2009, Thiago Pires (UFSC); The construal of	1				
2010, Leonardo Nunes (UFMG); As conjunções but	1				
total	18	1	1	1	

Gráfico 8: Programa utilizado no processamento dos corpora em pesquisas em ESFTBC.



A grande maioria dos pesquisadores optaram por utilizar o *WordSmith Tool* (Scott, 1999 a 2004); Zanella (2006) foi a única que usou o *Systemic Coder* (Mick O'Donell, 2002); Figueredo (2007) não processou seu corpus em *software* e utilizou o corpus de referência Lácio-Web, que não exige lidar com programa de análise lexicais como o *WST*, e Alves (2007) processou seu corpus inicialmente no *Concord*, do *WordSmith Tools*, mas para a quantificação dos dados utiliza os recursos do *OpenOffice.org Calc*, um Aplicativo de Planilha.

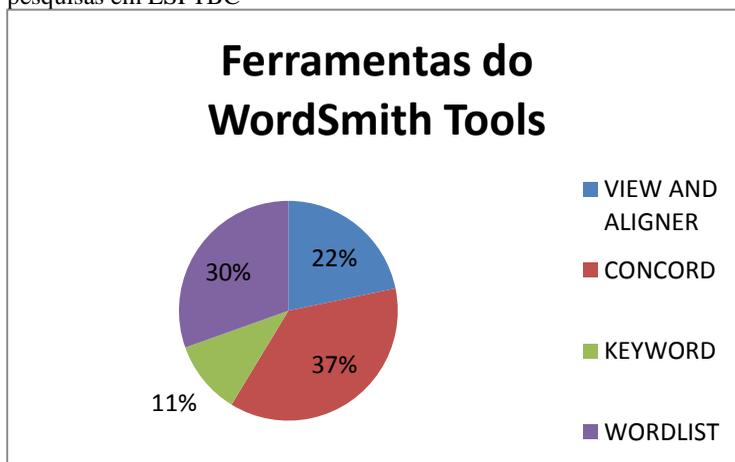
4.2.8. Método de Processamento: ferramentas do *WST*

No caso dos trabalhos que utilizaram o *software WordSmith Tools*, todos utilizaram inclusive ou exclusivamente as seguintes ferramentas: *Viewer and Aligner*, *Concord*, *WordList* e *Keyword*. Os dados percentuais no gráfico abaixo são calculados sobre um total de 46 respostas, por causa do uso de mais de uma ferramenta por muitos pesquisadores.

Quadro 41: Ferramentas do *WST* utilizadas no processamento dos corpora na pesquisas em ESFTBC

	FERRAMENTAS WORDSMITH TOOL				Total
	VIEW AND	CONCORD	KEYWORD	WORDLIST	
2003, Osilene Cruz (UFMG); <i>Harry Potter and the</i>	1	1	0	1	
2003, Cristina Mauri (UFMG); <i>Um estudo da</i>	0	1	0	1	
2004, Silvana de Jesus (UFMG); <i>Representação do</i>	1	1	0	1	
2004, Roberto Assis (UFMG); <i>A Transitividade na</i>	1	1	0	1	
2005, Letícia Bueno (UFMG); <i>Transitividade,</i>		1	1	1	
2005, Roberta Rodrigues (UFMG); <i>A Organização</i>	0	1	1	1	
2005, Eliza Morinaka (UFSC); <i>Gabriela, cravo e</i>	1	1	1	1	
2005, Tassiani Cançado (UFMG); <i>Transitividade e</i>	1	1	0	1	
2005, Marcos Feitosa (UFMG); <i>Uma proposta de</i>	1	1	0	1	
2005, Viviane Paquillin (UFSC); <i>The various facets</i>	1			1	
2006, Gleimara Zuniga (UFSC); <i>Construing the</i>		1		1	
2006, Lilian Fleuri (UFSC); <i>O perfil ideacional dos</i>	1	1			
2006, Ariana Zanella (UFSC); <i>Mapeamento macro</i>	0	0	0	0	
2007, Daniel Alves (UFMG); <i>Aspectos da</i>		1			
2007, Giacomo Figueredo (UFMG); <i>Uma descrição</i>	0	0	0	0	
2007, Cristiano Araújo (UFMG); <i>O sistema</i>		1			
2007, Joanna Filgueiras (UFSC); <i>Capoeira em</i>		1	1	1	
2009, Aline Fernandes (UFSC); <i>Black into white</i>	1	1	1	1	
2009, Thiago Pires (UFSC); <i>The construal of</i>	1	1			
2010, Leonardo Nunes (UFMG); <i>As conjunções but</i>		1		1	
total	10	17	5	14	46

Gráfico 9: Ferramentas do WST utilizadas no processamento dos corpora na pesquisas em ESFTBC

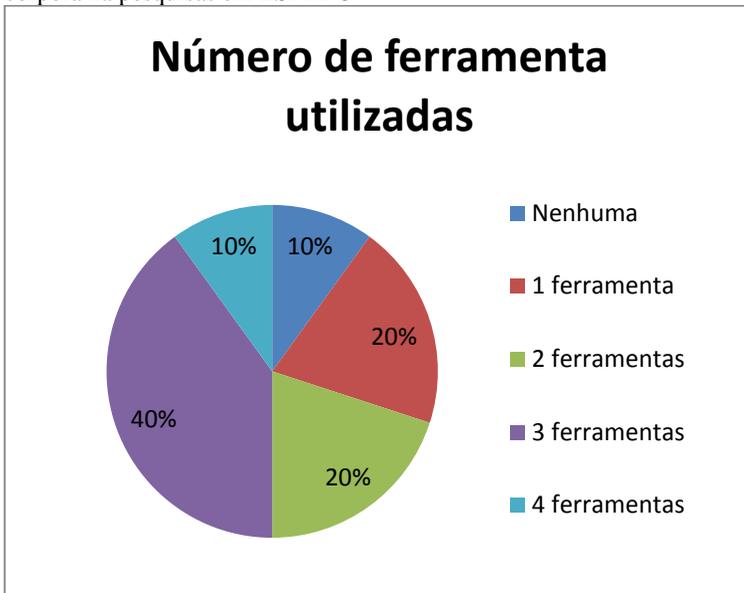


Na ordem de maior para menor utilização está o *Concord* como sendo a ferramenta mais utilizada pelos pesquisadores (37%), em seguida vem o *WordList* (30%), o *Viewer and Aligner* (22%) e o *Keyword* (11%). Todos os pesquisadores, com exceção de Zanella (2006) que não utiliza o WST e Figueredo (2007) que não processa seu corpus, utilizaram pelo menos uma dessas ferramentas. A distribuição de ferramentas do WST utilizadas entre os 20 pesquisadores é ilustrada no Quadro e Gráfico seguintes:

Quadro 42: Quantidade de Ferramentas utilizadas no processamento dos corpora na pesquisas em ESFTBC

	Número de ferramenta utilizadas					total
	Nenhuma	1 ferramenta	2 ferramentas	3 ferramentas	4 ferramentas	
2003, Osilene Cruz (UFMG); Harry				1		
2003, Cristina Mauri (UFMG); Um			1			
2004, Silvana de Jesus (UFMG);				1		
2004, Roberto Assis (UFMG); A				1		
2005, Leticia Bueno (UFMG);				1		
2005, Roberta Rodrigues (UFMG); A				1		
2005, Eliza Morinaka (UFSC); Gabriela,					1	
2005, Tassiani Cançado (UFMG);				1		
2005, Marcos Feitosa (UFMG); Uma				1		
2005, Viviane Paquillin (UFSC); The			1			
2006, Gleimara Zuniga (UFSC);			1			
2006, Lilian Fleuri (UFSC); O perfil			1			
2006, Ariana Zanella (UFSC);	1					
2007, Daniel Alves (UFMG); Aspectos		1				
2007, Giacomo Figueredo (UFMG); Uma	1					
2007, Cristiano Araújo (UFMG); O		1				
2007, Joanna Filgueiras (UFSC);				1		
2009, Alinne Fernandes (UFSC); Black					1	
2009, Thiago Pires (UFSC); The		1				
2010, Leonardo Nunes (UFMG); As			1			
total	2	3	5	8	2	20

Gráfico 10: Quantidade de Ferramentas utilizadas no processamento dos corpora na pesquisas em ESFTBC



Poucos foram os pesquisadores que utilizaram todas as ferramentas, de fato, apenas Morinaka (2005) e Fernandes (2009) relatam ter explorado as quatro ferramentas na fase de compilação do corpus e de quantificação dos dados.

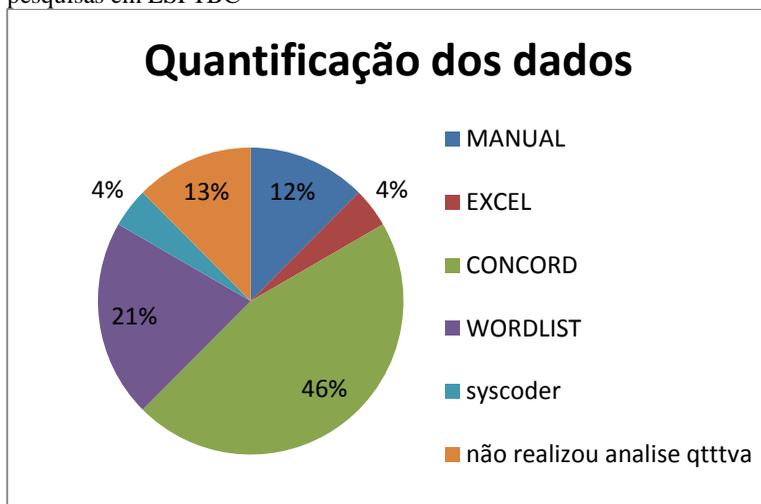
4.2.9. Método de Processamento: quantificação dos dados

Para a fase de quantificação dos dados, além de diferentes ferramentas do *WST* terem sido exploradas, como o *Concord* e *WordList*, houve quem utilizou outros programas (*OpenOffice.org Calc* e *Syscoder*), conforme ilustrado no gráfico abaixo. Os percentuais deste gráfico são calculados sobre um total de 24 respostas, pois 4 pesquisadores relataram ter obtido a quantificação dos dados através do *Concord* e do *WordList* simultaneamente (i.e. Caçado, 2005; Feitosa, 2005; Filgueiras, 2007 e Pires, 2009), o restante dos pesquisadores utilizaram apenas um recurso.

Quadro 43: método e ferramenta do *WST* para a quantificação dos dados nas pesquisas em ESFTBC

	QUANTIFICAÇÃO DOS DADOS						total
	MANUAL	EXCEL	CONCORD	WORDLIST	syscoder	não realizou análise	
2003, Osilene Cruz (UFMG); Harry Potter and the	1						
2003, Cristina Mauri (UFMG); Um estudo da	1						
2004, Silvana de Jesus (UFMG); Representação do			1				
2004, Roberto Assis (UFMG); A Transitividade na	1						
2005, Letícia Bueno (UFMG); Transitividade,						1	
2005, Roberta Rodrigues (UFMG); A Organização			1				
2005, Eliza Morinaka (UFSC); Gabriela, cravo e			1				
2005, Tassiani Cancado (UFMG); Transitividade e			1	1			
2005, Marcos Feitosa (UFMG); Uma proposta de			1	1			
2005, Viviane Paquillin (UFSC); The various facets				1			
2006, Gleimara Zuniga (UFSC); Construing the			1				
2006, Lilian Fleuri (UFSC); O perfil ideacional dos			1				
2006, Ariana Zanella (UFSC); Mapeamento macro					1		
2007, Daniel Alves (UFMG); Aspectos da		1					
2007, Giacomo Figueredo (UFMG); Uma descrição						1	
2007, Cristiano Araújo (UFMG); O sistema						1	
2007, Joanna Filgueiras (UFSC); Capoeira em			1	1			
2009, Alinne Fernandes (UFSC); Black into white			1				
2009, Thiago Pires (UFSC); The construal of			1				
2010, Leonardo Nunes (UFMG); As conjunções but			1	1			
total	3	1	11	5	1	3	24

Gráfico 11: método e ferramenta do *WST* para a quantificação dos dados nas pesquisas em ESFTBC



Como Bueno (2005), Figueredo (2007) e Araújo (2007) não enfocam sua análise quantitativa do objeto de análise, eles representam o percentual de 13% que não realiza análise quantitativa. Apenas Alves (2007) utiliza o *OpenOffice.org Calc* para a contabilização de seus dados e Zanella (2006) o *SysCoder* (Systemic Coder). Cruz (2003), Mauri

(2003) e Assis (2003), que optam por não anotar o corpus, também não os processam para quantificação automática, representando os 12% que contabilizam manualmente seus dados. O restante das pesquisas (67%) utilizam inclusive ou exclusivamente o *Concord* e/ou o *WordList* (do *WST*) para gerar a quantificação dos dados automaticamente.

4.2.10. Conclusão: perfil geral das dissertações em ESFTBC

Foram analisadas 20 dissertações brasileiras em ESFTBC—todas investigando temas relativos a traduções a partir do viés teórico da Linguística Sistemico-Funcional, buscando aplicar a metodologia de corpus. Cada pesquisa apresentou de modo peculiar seus procedimentos metodológicos. Observo que após a publicação da tese de Lincoln Fernandes (2004), este modo de apresentação passou a ser mais sistematizado em algumas pesquisas (i.e. Fleuri, 2006; Fernandes, 2009; Pires, 2009; Filgueiras, 2007, entre outras). Baseada na descrição de Olohan (2004), que se pauta na descrição do desenho, compilação e processamento do corpus para detalhar a metodologia de corpus aplicada na pesquisa, essa forma de organização fez com que informações metodológicas estivessem mais acessíveis e explícitas ao leitor, permitindo que fossem mais replicáveis em futuros estudos.

A partir das informações coletadas dessas dissertações, elaboraram-se dez gráficos, em nove seções, que expõem os percentuais das pesquisas em ESFTBC no Brasil até 2010, em relação à Metafunção explorada, ao objeto de estudo, ao tipo de corpus estudado, à versão dos textos compilados, ao tipo de processo de alinhamento, à anotação, aos programas adotados, às ferramentas do *WST*, à quantidade de ferramentas utilizadas e ao modo de processar a quantificação dos dados.

Em relação cada item mencionado, a maioria (65%) estuda a Metafunção Ideacional e a minoria associa a Metafunção Textual com a Ideacional (5%); a maioria tem como objeto de análise os Participantes de Processos (29%); a maior parte dos corpora utilizados eram paralelos bilíngues de pequena dimensão (75%)⁷⁵; grande parte das pesquisas tiveram que lidar com a compilação de textos impressos (60%) e a minoria obteve o corpus preparado para ser processado, ou seja, digitalizado e corrigido (8%) ou acessou corpora já compilados; dos

⁷⁵ 67% analisa exclusivamente um corpus paralelo bilíngue e 8% (Jesus, 2004; Nunes, 2010) analisa corpus paralelo bilíngue e corpus comparável. Este é o percentual sobre o número de corpus utilizado, seis pesquisas utilizaram mais de um corpus.

trabalhos que alinham o corpus, a maioria acessou o utilitário *Viewer and Aligner* do *WST* (56%); empatou a proporção majoritária que optou por rotular o corpus com o *CROSF* (24%) e com um código criado (24%); a grande maioria das dissertações acessou o *WST* (85%); dos que acessaram o *WST* a maioria utilizou a ferramenta de *Concord* (37%) e optaram por utilizar três ferramentas deste programa simultaneamente (40%); dos trabalhos que realizaram análise quantitativa, a maioria optou por contabilizar os dados no *Concord* (40%). Este é o perfil geral das dissertações em ESFTBC publicadas no Brasil de 2003 a 2010.

4.3. DESENVOLVIMENTO DIACRÔNICO DA METODOLOGIA DAS DISSERTAÇÕES EM ESFTBC

Esta seção expõe a evolução diacrônica da metodologia de corpus das pesquisas em ESFTBC analisadas nesta tese.

O trabalho pioneiro em ESFTBC foi a dissertação de Cruz (2003), seguido pelo de Mauri (2003). Os trabalhos produzidos em 2003 organizam as categorias em tabelas e realizam a contagem manualmente. As pesquisas defendidas neste ano se caracterizam também por analisarem verbos de elocução, descritos a partir da Metafunção Ideacional.

Em 2004 há uma transição entre o estudo de verbos de elocução para o estudo de Participantes de Processo, com a dissertação de Assis (2004). Nesse mesmo ano, Jesus (2004) cria pela primeira vez códigos de anotação para anotar o corpus e os quantifica automaticamente no *WST*. Os dois primeiros anos de ESFTBC permanecem concentrados na UFGM.

O ano de 2005, que coincide com a realização do convênio PROCAD UFGM-UFSC⁷⁶ (cf. Capítulo 1, de Introdução, nesta tese), representa o período de maior produção em ESFTBC, com 6 defesas neste ano (ao invés de duas como vinha sendo até então), envolvendo pela primeira vez pesquisas afiliadas à UFSC (Morinaka, 2005 e Paquilin, 2005). Em 2005 novas perspectivas de investigação são exploradas, como é o caso da análise da Metafunção Textual por Rodrigues (2005), Paquilin (2005) e Feitosa (2005) e da descrição de duas Metafunções simultaneamente, a Ideacional e a Textual, por Morinaka (2005). Em

⁷⁶ O convênio intensificou e consolidou as relações acadêmicas entre os pesquisadores das duas universidades, tendo efeito sinérgico: a produção de conhecimento compartilhado se intensificou durante o período de vigência do convênio.

2005 constatou-se a necessidade da elaboração de um código de rotulação, com a defesa da dissertação de Feitosa (2005). Esse código foi testado (além do teste realizado pelo próprio criador) em 2005, enquanto se encontrava ainda em fase de elaboração, por Paquilin (2005) e por Morinaka (2005).

Em 2006 é defendida a primeira tese em ESFTBC (Rodrigues Junior, 2006). Neste ano foram defendidas 3 dissertações afiliadas à UFSC: duas afiliadas ao projeto CORDIAL (Fleuri, 2006 e Zuniga, 2006) e, pela primeira vez, uma não afiliada ao projeto (Zanella, 2006). As dissertações de Fleuri e de Zuniga mantiveram características semelhantes: analisar a representação e construção do perfil ideacional do item lexical *tradutor* e *translator*, em corpora acadêmicos dos Estudos da Tradução. Ambas consolidaram o uso do CROSF rotulando categorias da Metafunção Ideacional.

Em 2007 quatro dissertações são defendidas. Figueredo (2007) e Araújo (2007) introduzem um modo diferente de conduzir esse tipo de pesquisa, utilizando pela primeira vez corpora monolíngues e realizando um estudo voltado para a descrição linguística da tradução. Filgueiras (2007) também aborda pela primeira vez a questão de prosódia semântica, trabalhando com corpus em relação tradutória, mas não alinhados. Alves (2007) apresenta-nos um novo modo de processar a quantificação dos dados e organização do corpus por meio de um aplicativo de planilha, o OpenOffice.org.

O ano de 2008 ocorreu a defesa de uma tese de doutorado (Jesus, 2008) e nenhuma defesa de dissertação em ESFTBC. Em 2009, foram defendidas duas dissertações afiliadas à UFSC (Fernandes, 2009 e Pires, 2009), sendo que ambas analisavam a representação e a construção do perfil ideacional de Participantes de Processos e ambas utilizaram o CROSF como código de rotulação. Neste mesmo ano, na UFMG, são defendidas duas teses de doutorado em ESFTBC (Mauri, 2009 e Assis, 2009).

Nunes (2010) representa a única dissertação defendida no ano de 2010. Por não analisar categorias de Metafunção, Nunes não utiliza o CROSF e sim uma adaptação do *tagset Susanne* (Sampson, 1995). Duas teses afiliadas a UFSC (Espindola, 2010 e Souza, 2010) e uma afiliada à UFMG (Rodrigues, 2010) são defendidas nesse ano, aumentando o número de defesas de teses em relação ao número de defesas de dissertações.

No que tange às metodologias das dissertações em ESFTBC, ao longo desses sete anos, observo que códigos foram criados e testados, diferentes formas de alinhamento foram aplicadas e modos variados de

quantificar dos dados foram utilizados. No entanto, não se observam formas diferentes de construir o corpus, de modo a se gastar menos tempo na compilação e mais no estudo e análise das categorias pesquisadas. As etapas de escaneamento, revisão e correção dos erros de leitura do escâner, de preparação do corpus para um alinhamento automático ou manual, de rotulação, de quantificação dados e a etapa de revisão desta rotulação continuam sendo relatados como tarefas: “demandantes de atenção” e “árduas” (Cruz, 2003), “*supostamente* objetivas” (Jesus, 2004, grifo meu), “custosas” (Assis, 2004), pois os programas são pagos, “esforçosas” (Bueno, 2005), “*not simple*” e “*requires manual work and patience*”, (Morinaka, 2005), “requerem um trabalho mecânico e lento” (Fleuri, 2006), “exigem minuciosa atenção” (Filgueiras, 2007), “*takes several hours of work*” (Fernandes, 2009), etc.

Em dois anos de estudo de mestrado, muitas vezes sendo introduzidos à teoria da LSF e dos ET e às ferramentas metodológicas da LC, os pesquisadores não arriscam gastar seu tempo em explorar os pormenores do Suíte de Aplicativos para Escritório, com a qual lida durante a redação de sua pesquisa. Ao invés, aplicam seu tempo no aprendizado do programa *WST* e em realizações de tarefas manuais e cumprimento destas, muito mais que em possíveis frustrantes estudos de programação das ferramentas do *MS Office*, por exemplo. Isso porque a exploração de programas e ferramentas básicas do Suíte de aplicativos para escritório demanda um estudo, criações de hipóteses, familiarização de seu corpus e suas necessidades e inúmeros testes, muitas vezes frustrantes. Com o intuito de oferecer alternativas de compilação de corpus que visem a uma agilização desse processo, apresenta-se nas próximas seções o funcionamento de um possível método de compilação e de processamento de corpus, buscando tornar essa etapa mais rápida e economicamente acessível a todos os pesquisadores neste campo.

5. PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE CORPUS

Baseando-se nas desideratas propostas por Garretson (2008)⁷⁷ e na análise das dissertações em ESFTBC, este capítulo apresenta uma proposta metodológica de procedimentos de compilação de corpus, relativos (i) à preparação do corpus para alinhamento; (ii) ao alinhamento; (iii) à anotação do corpus ou classificação dos objetos analisados; (iv) à revisão dos elementos anotados/classificados e (v) quantificação dos dados em gráficos. Para tanto são apresentados alguns pontos da conclusão do Capítulo 4, de Análise, que conduziram à elaboração da presente proposta.

No Capítulo 4 foi destacado o perfil das dissertações em ESTBC publicadas no Brasil até 2010. Consto que:

- Essas pesquisas lidam não exclusivamente, com exceção de duas pesquisas (Figueredo, 2007; Araújo, 2007), com corpus paralelo bilíngue de pequena dimensão;
- Das pesquisas que alinham textos, a maioria utiliza o *Viewer and Aligner* e mencionam que esta fase demanda muito tempo, organização e atenção;
- Das pesquisas que anotaram o corpus, variaram muito o modo de rotular. Exatamente a mesma quantidade (24% cada) optou por criar seu próprio rótulo e por utilizar o CROSF tal qual é descrito por Feitosa (2005).
- A maioria das pesquisas utiliza o Programa *WordSmith Tools* para processar o corpus;
- Quanto ao modo de quantificação dos dados, a maioria das pesquisas utilizou o *Concord* para realizar a contagem automática das etiquetas ou dos nódulos que representavam o objeto de análise.

Ainda que não tenha quantificado estas informações, observei durante nas dissertações os pesquisadores mencionam ou insinuam (i) terem utilizado um Processador de Texto (*MS Word*) para, ao menos, a redação da pesquisa, demonstrando familiaridade com o programa e (ii)

⁷⁷ cf. seção 2.2 desta tese.

terem formulado um método próprio para a revisão do corpus pós-escaneamento, pré-alinhamento, durante o alinhamento, pós-classificações dos objetos de análise e de suas rotulações (i.e. Fleuri, 2006; Mauri, 2003; Cruz, 2003; Bueno, 2005; Rodrigues, 2005; Morinaka, 2005; entre outros). Em geral esse método não é detalhado. As correções e revisões são feitas aleatoriamente, conforme a necessidade de cada pesquisa. Quando descritas na metodologia, observa-se um relato do que fizeram, mais que um método desenvolvido e passível de ser reproduzido. O mesmo ocorre com a anotação do corpus. Apesar de haver um código sistematizado desde 2005 (Feitosa, 2005), a forma de utilização deste código por alguns pesquisadores (Fleuri, 2006; Zuniga, 2006; Paquilin, 2005; Morinaka, 2005; Fernandes, 2009; Pires, 2009) visou a cumprir apenas com os propósitos da própria pesquisa e apresentaram pequenas variações entre si, não sendo possível criar um único corpus anotado composto por cada corpus. Observo ainda que a partir da publicação da dissertação de Zuniga (2006) ocorre uma padronização da estrutura do capítulo de metodologia, principalmente nas dissertações publicadas na UFSC, baseada na estrutura da tese de Fernandes (2004), que aconselha que tal capítulo seja organizado segundo (i) o desenho do corpus; (ii) a construção do corpus e (iii) o processamento do corpus. Essa estrutura já vinha sendo praticada a seu modo por Morinaka (2005), sem se referir a Fernandes (2004).

Concluo, ao final do Capítulo 4, que apesar de muito progresso na metodologia dessas dissertações ao longo de sete anos de tradição em pesquisa em ESFTBC, ocorre ainda um (i) grande gasto de tempo de pesquisa na fase de compilação do corpus, comprometendo o aprofundamento da análise das categorias textuais; (ii) inúmeras transições de um programa a outro durante a fase de compilação do corpus e, conseqüente, a geração de inúmeros documentos gravados em *Document Word* e *Plain Text*, que pode levar à perda de documentos e de dados. A fase de compilação em geral implica os processos de:

- (i) Escaneamento;
- (ii) Revisão e correção dos erros de leitura do escâner;
- (iii) Preparação do corpus para um alinhamento semiautomático;
- (iv) Alinhamento das unidades desalinhasdas;
- (v) Anotação do corpus;
- (vi) Revisão da anotação;
- (vii) Quantificação dos dados
- (viii) Exposição dos dados na redação da pesquisa.

Considerando, então, o perfil das pesquisas em ESFTB, as observações feitas e as etapas de compilação aqui listadas, identifico que as pesquisas em ESFTBC:

- (i) Necessitam recursos financeiros para a aquisição do programa de compilação e processamento, pois a maioria utiliza o *WST*;
- (ii) Relatam gastar muito tempo no processo de compilação do corpus paralelo bilíngue de pequena dimensão;
- (iii) Anotam seu corpus para cumprir apenas com o interesse de sua pesquisa e não para criar um grande corpus anotado;
- (iv) Não apresentam um método que vise a agilizar e padronizar as etapas de compilação.

Tendo essas observações em vista, exponho neste capítulo uma proposta que visa a oferecer um método econômico que agilize e padronize os processos de compilação, que diminua o número das etapas em cada processo, dos programas e da geração de diferentes documentos e que flexibilize o modo de anotação do corpus para satisfazer as necessidades próprias de cada pesquisa. Para isso, proponho utilizar um programa que mostra-se familiar a todos os pesquisadores—o *MS Office*, e que permite que seja adaptado pelo pesquisador para se adequar ao perfil de sua própria pesquisa. A correção pós-escaneamento e a agilização da preparação do corpus para o alinhamento são desenvolvidas na seção (5.1) através do *Template Editor* e o alinhamento, a flexibilização de anotação do corpus⁷⁸, a organização para a revisão da anotação e para a visualização automática da quantificação de dados é desenvolvida na seção (5.2) *Template Corpus Paralelo*. A forma de desenvolvimento desses *Templates* é explicada no Capítulo 3, de Metodologia. Esse capítulo e os anexos referidos nele, ao oferecerem as fórmulas e *Macros* criados, permitem a reprodução e adaptação de tais *Templates*. Este

⁷⁸ Veremos que o que agora chamamos de “flexibilização de anotação” não se trata realmente de ‘anotação’ (segundo a definição de Leech, 2005), mas de um modo de organizar a classificação dos objetos analisados. A anotação, segundo Leech, é realizada no corpo do texto, com marcações textuais (como traço baixo ou parênteses angulares) que permitem programas prontos, como o *WST*, buscar e realizar concordâncias. A forma proposta aqui é um modo de anotar em tabela a quantidade de elementos presentes em uma determinada categoria, escrita na primeira linha de cada coluna. Essa forma de “anotação flexibilizada” é exemplificada na seção 5.2. Também denominamos de “classificação do objetos de análise”.

capítulo, de Proposta, explica a organização e funcionamento dos *Templates* para a pesquisa.

Os *Templates* criados são disponibilizados online para download gratuito no site: <https://sites.google.com/site/lilianjfleuri/>. Caso o *Template* oferecido no site não funcione ou necessite adaptações para a pesquisa em que é utilizado, no Capítulo 3 e nos anexos 9 e 10 ofereço a descrição detalhada de cada *Macro* gravado e programado e de cada fórmula criada, possibilitando sua reprodução. Conforme mencionado no Capítulo 3, os *Templates* são criados considerando corpora compostos por textos em língua portuguesa e/ou inglesa. É possível que os *Templates* funcionem em textos em idiomas de origem latina, como francês, italiano, espanhol e anglo-saxões como o alemão, mas seguramente não funcionam em corpora compostos por textos em árabe, chinês, japonês, coreano, russo, grego e outras línguas que possuem caracteres não latinos e outro sistema numérico que não seja o algorítmico.

Lembro que a utilização dos *Templates* não significa, entretanto, abandonar completamente o uso do *WST* ou de outros programas prontos. Pois, tanto o *WordSmith Tools* como o *AntConc* oferecem as ferramentas de *WordList* e de *Keyword* que não são substituíveis em nenhum dos *Templates*⁷⁹. Também não significa abandonar códigos de anotação, como o CROSF⁸⁰, pois a proposta de montar corpus anotado não é o foco destes *Templates*.

⁷⁹ Entretanto, se o objetivo é apenas saber o número de palavras ou de toques, essa informação pode também ser adquirida no “contar palavras” dos Processadores de Textos.

⁸⁰ Enquanto a forma de rotulação proposta nesta tese visa simplesmente flexibilizar e dinamizar o processo de classificação do pesquisador em sua própria pesquisa, o CROSF propõe uma sistematização de codificação (cf. Feitosa, 2005).

5.1. *TEMPLATE EDITOR*: PREPARAÇÃO DO CORPUS PRÉ-ALINHAMENTO

Esta seção é dedicada a explicar e ilustrar o funcionamento do *Template Editor*, criado no Processador de Texto *MS Word*. As instruções se encontram resumidamente expostas na própria página de cada *Template*. Opto por ilustrar cada passo, deixando a tela do *Template* exposta no corpo do texto ao invés de no anexo, pois considero que tais ilustrações representam os resultados visíveis do trabalho aqui realizado. O *Template Editor* realiza a preparação de cada texto (Texto-Fonte e Texto-Alvo separadamente) para seu alinhamento, visando a obter menos ocorrências de desalinhamento (causados pela presença de pontos finais em abreviações, por exemplo). Realiza isso eliminando em um único toque a maioria dos pontos não marcadores de final de sentença e quebrando todo o texto em sentenças.

As etapas de utilização do *Template Editor* são:

- (i) **Colar o texto no *Template***: os textos colados no *Template Editor* podem ser copiados de um arquivo em *Plain Text*, *Document Word*, PDF ou *online*.
- (ii) **Formatar o texto se necessário**: Se o texto apresentar muitas quebras de linhas ou de páginas é necessário formatá-lo, transformando-o em um só bloco:
 - a. Na ferramenta *Find & Replace*, eliminando as marcas de parágrafos:

Equação 6: Eliminar parágrafos com o *Find & Replace*

Find what	Replace with
^p	

- b. Nesta ferramenta, eliminando quebras manuais de página:

Equação 7: Eliminar quebras manuais de página com o *Find & Replace*

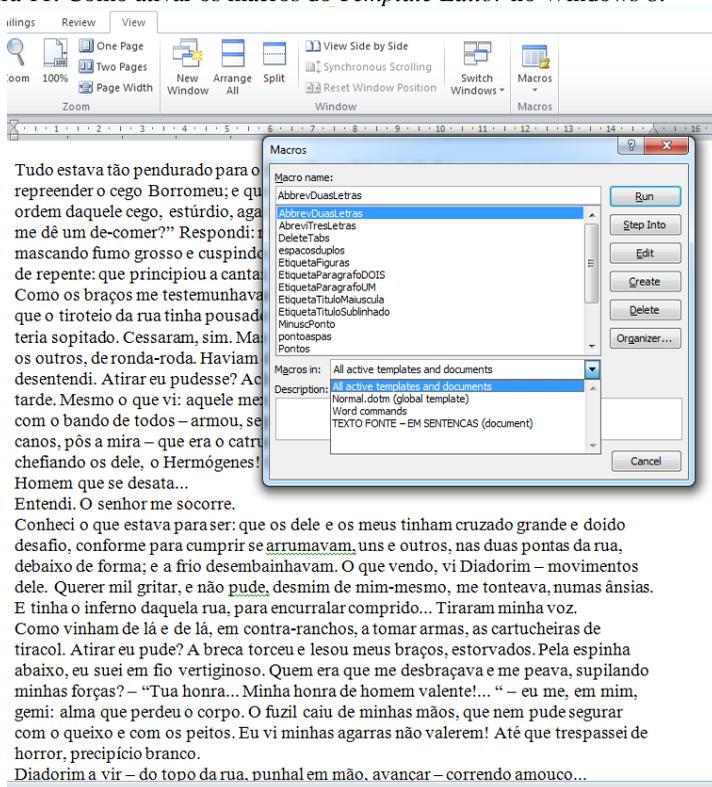
Find what	Replace with
^m	

- (iii) **Ativar os Macros criados**: uma vez transformados em um só bloco de texto, passa-se a etapa de ativação dos Macros do *Template Editor* para se eliminar pontos não marcadores de

final de sentença e quebrar o texto em sentenças. Para ativar os *Macros* realizam-se os seguintes passos:

- Ir à função “View” do *MS Word (Windows 8)*,
- Clicar em “Macros” (no canto superior direito),
- Clicar em “View Macros”, vai abrir a janela de “Macros”
- Escolher em “Macros in” o nome do documento
- Acionar o comando “Run”.

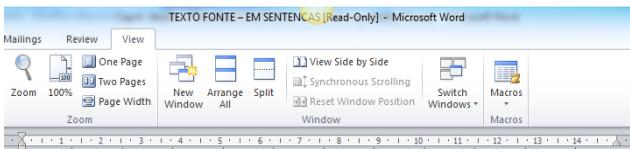
Figura 11: Como ativar os macros do *Template Editor* no *Windows 8*.



- Selecionar todo o texto e acionar o comando gravado **Alt+T**:** com o texto colado, suas linhas e parágrafos formatados e os *Macros* ativados, passa-se as etapas que acionam os *Macros*, colocando-os em atividade. Seleciona-se todo o texto e digita-se **Alt+T** para ativar todos os *Macros*. Caso este comando não os ative ou se algo der errado, verifique se os *Macros* foram ativados. Se foram e ainda

assim o comando continua não funcionando de acordo com o esperado, substitui-se o controle **Alt+T**, que representa a composição de dez controles—de **Alt+0** ao **Alt+9** (descritos no Capítulo 3), por cada um dos controles que os compõem e verifica-se onde o problema ocorreu, não fazendo uso de um dos controles, se necessário. O resultado final é um texto com as sentenças quebradas, pontos destacados (para facilitar o reconhecimento de alguma sentença não quebrada) e pontos de abreviações eliminados:

Figura 12: Resultado do texto processado no *Template Editor*.



Tudo estava tão pendurado para o fim... Derradeiro ainda foi, que eu virei para trás, para repreen-der o cego Borromeu; e que eu estava com dormente dor, nos braços
Sem-ordem daquele cego, estúrdio, agachado lá, cocoral.
Só fez que disse, bronco: – “Quem me dê um de-comer?” Respon-di: ralhei.
Ah, há-de-o, singular ficasse, mesmo ali, mascando fumo grosso e cus-pindo amarelo e preto... Dei num suor.
Vozeiro dele, então, de repente: que principiou a cantar, ele estava cantando um louvado...
Como os braços me testemunhavam um peso... Mesmo estranhei, quando fui notando que o tiro-teio da rua tinha pousado termo; achei que fazia um certo minuto que o fogo teria sopitado.
Cessaram, sim.
Mas gritavam, vuvu vavava de conversa ruim, uns para os outros, de ronda-roda.
Haviam de ter desautorizado toda munição?
Olhando, desentendi.
Atirar eu pudesse?
Acho que quis gritar, e esperei para depoimento, mais tarde.
Mesmo o que vi: aquele mexinflol.
E que quem saía duma porta, para ir se juntar com o bando de todos – armou, segurando frente de si engatilhada uma garrucha de dois canos, pôs a mira – que era o catrumano Teofrácio, como se fosse braço-d’armas!
E vi, chefiando os dele, o Hermógenes!
Chapéu na cabeça era um bandeirão redondo... Homem que se desata...
Entendi.
O senhor me socorre.
Conheci o que estava para ser: que os dele e os meus tinham cruzado grande e doído desafio, conforme para cumprir se arrumavam, uns e outros, nas duas pontas da rua, debaixo de forma; e a frio desembainhavam.
O que vendo, vi Diadorim – movimentos dele.
Querer mil gritar, e não pude, desmim de mim-mesmo, me tonteava, numas ânsias

⁸¹ Os destaques em amarelo visam a facilitar a identificação dos pontos marcadores de final de sentença. Destacá-los leva à identificação rápida das sentenças não quebradas. Nos testes realizados nesta tese foram poucos os casos em que isso acontece, mas se isso demonstrar ser um problema, essas sentenças podem ser quebradas no *Template Corpus Paralelo*, manualmente

- (v) **Copiar todo o texto e colá-lo no *Template Corpus Paralelo*:** abre-se o *Template Corpus Paralelo* e se cola o TF na coluna de texto-fonte (coluna A) da pasta de trabalho chamada “Alinhamento”.
- (vi) **Salvar o documento:** salva-se o documento *Word* indicando qual texto é TF e qual é TA. Repete-se o processo com o Texto-Alvo, colando o texto na coluna B do *Template Corpus Paralelo*.

Esse processo, composto por estes seis passos, não dura mais do que cinco minutos para transformar cada texto em um texto passível de ser alinhado.

5.2. *TEMPLATE CORPUS PARALELO*: ALINHAMENTO, CLASSIFICAÇÃO DOS OBJETOS ANALISADOS, QUANTIFICAÇÃO E EXPOSIÇÃO DOS DADOS

Com o propósito de agilizar o processo de alinhamento, flexibilizar a anotação do corpus, facilitar o processo de revisão dos elementos rotulados e oferecer uma visualização gráfica e automática da quantificação dos dados, de modo a evitar transições e a gerações de inúmeros documentos, este *Template* é organizado em cinco pastas de trabalho, descritas nas seguintes subseções: (5.2.1) pasta de alinhamento, (5.2.2) de rotulação TF, de rotulação TA, (5.2.3) de listas e gráficos e (5.2.4) de dados. Em cada uma dessas subseções é exposta uma figura da pasta, seguida da explicação do seu funcionamento. Todas as pastas apresentam a primeira linha dedicada às instruções de uso e às anotações de pesquisa. Espera-se, com este *Template*, que o manuseio com corpora paralelos de pequena dimensão se torne mais econômico em termos de tempo de compilação e de custo, e mais acessível em no que se refere à familiaridade com linguagem dos programas envolvidos.

no *Template Editor* —se forem poucas ocorrências, ou o pesquisador pode reprogramar esse *Macro* ou criar um novo *Macro* que solucione esse problema.

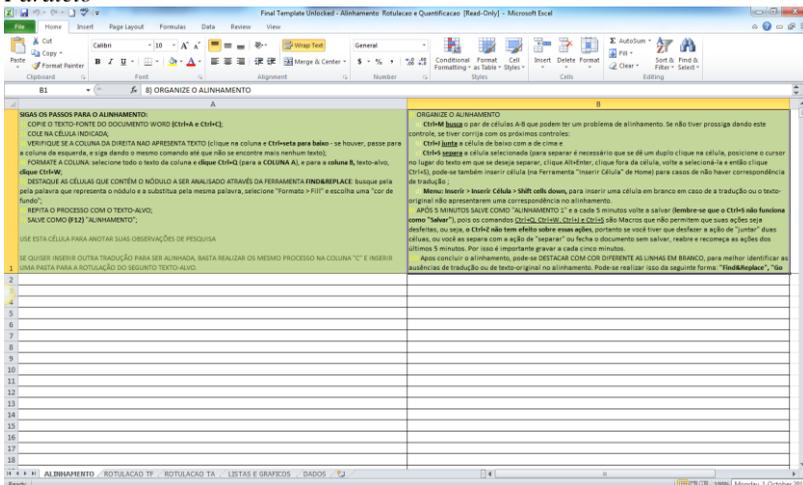
5.2.1. Pasta de Alinhamento

A “Pasta de Alinhamento” é o local onde se realiza o alinhamento do corpus paralelo bilíngue de pequena dimensão. Conforme foi visto no Capítulo 3 o *Template Corpus Paralelo* apresenta *Macros* que desempenham a função de:

- (i) Formatar coluna A (gravado no comando Ctrl+Q);
- (ii) Formatar coluna B (comando Ctrl+W);
- (iii) Buscar por linhas desalinhadas (Ctrl+M);
- (iv) Juntar a célula de baixo com a de cima (comando Ctrl+J);
- (v) Separar uma célula em duas, posicionando uma abaixo da outra (comando Alt+Enter e Ctrl+S).

Esta pasta se apresenta da seguinte *layout*, quando o documento é aberto:

Figura 13: Tela da primeira pasta, pasta de alinhamento, do *Template Corpus Paralelo*



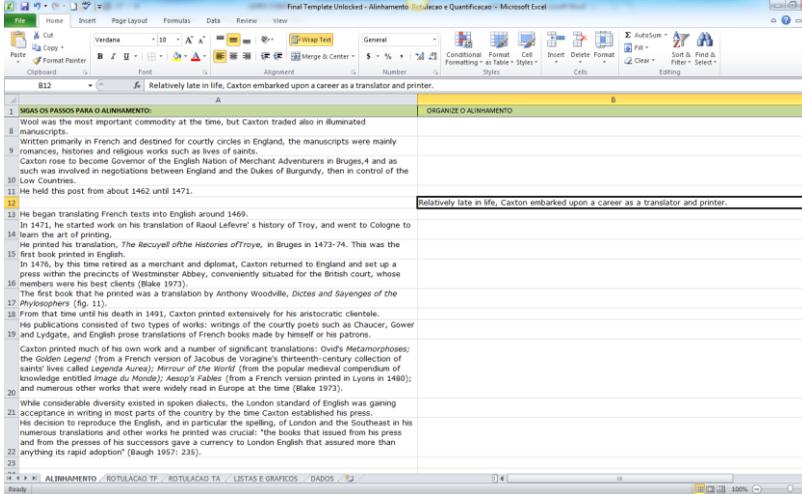
A ‘Pasta de Alinhamento’ tem a coluna A dedicada ao TF e a coluna B ao TA⁸². As etapas para se realizar o alinhamento são as expostas nos seguintes passos e ilustrações:

⁸² Em caso de se desejar alinhar um terceiro texto, aconselha-se a realizar em outra planilha *Template*, pois o *Macro* de ‘Busca por desalinhamento’ não foi programado para buscar linhas da coluna C. Se for o caso de se desejar usar a coluna C, tal *Macro* então teria que sofrer uma modificação para se adaptar ao novo propósito.

- (i) Acionar os Macros do *Template Corpus Paralelo* (conforme exposto na seção 5.1 desta tese);
- (ii) Selecionar e copiar o texto-fonte do documento no *Template Editor*;
- (iii) Colar texto copiado na célula indicada (célula A2 para o texto-fonte e B2 para o texto-alvo);
- (iv) Verificar se a coluna B não apresenta texto (conforme ilustra a Figura a seguir). Para isso, clicar na célula B2 e, em seguida, *Ctrl+seta para baixo*. Esse comando leva o cursor à próxima célula com texto. Se houver texto, passá-lo para a coluna A e seguir realizando o mesmo comando até que nenhum texto seja encontrado. Isso pode ocorrer quando o texto colado apresenta o caractere escondido de *Tab characters*⁸³. Se for utilizado o *Template Editor*, que elimina esse caractere, para editar o texto colado, isso não deveria acontecer. Em todos os casos, aconselha-se que seja conferido neste estágio, para não correr o risco de comprometer a montagem do corpus paralelo e a pesquisa.

⁸³ “Tab character” é um “caractere especial que pode ser inserido em um document de texto. Diferentes programas reagem de modos diferentes a este caractere. A maioria dos processadores de palavra, por exemplo, movem o cursor ou o ponto de inserção para o próximo ‘tab stop’, e a maioria das impressoras também movem o ponto de impressão ao próximo ‘tab stop’. Alguns programas, entretanto, simplesmente ignoram esse caractere”. (Fonte: http://www.webopedia.com/TERM/T/tab_character.html, consultada no dia 06/09/13).

Figura 14: Tela de como passar textos da coluna B para a A (pasta 1 T.C.P)



- (v) Formatar a coluna: selecionar todo o texto da coluna A e acionar o comando de formatação da coluna A (**Ctrl+Q**) e da coluna B (**Ctrl+W**). As Figuras a seguir ilustram como um texto-fonte não formatado se apresenta antes e depois de acionar o comando de formatação da coluna A:

Figura 15: Tela – ativando a formatação da coluna A (texto recém-colado)

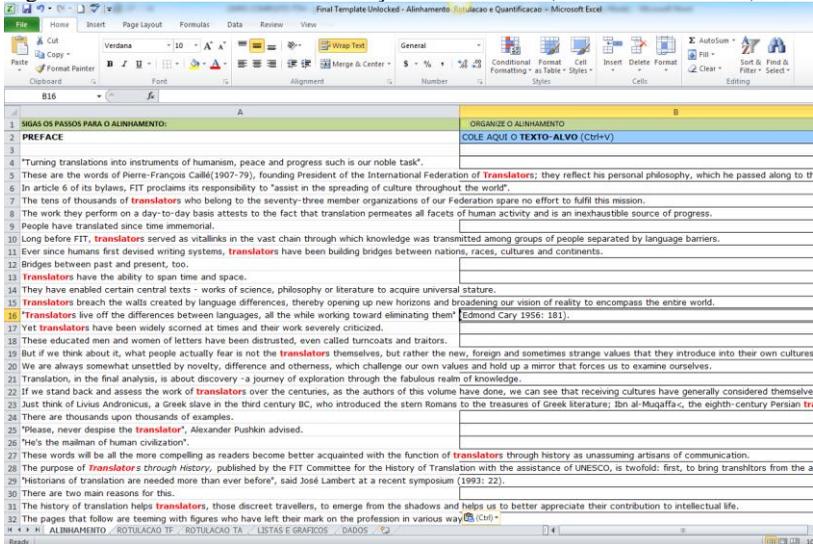
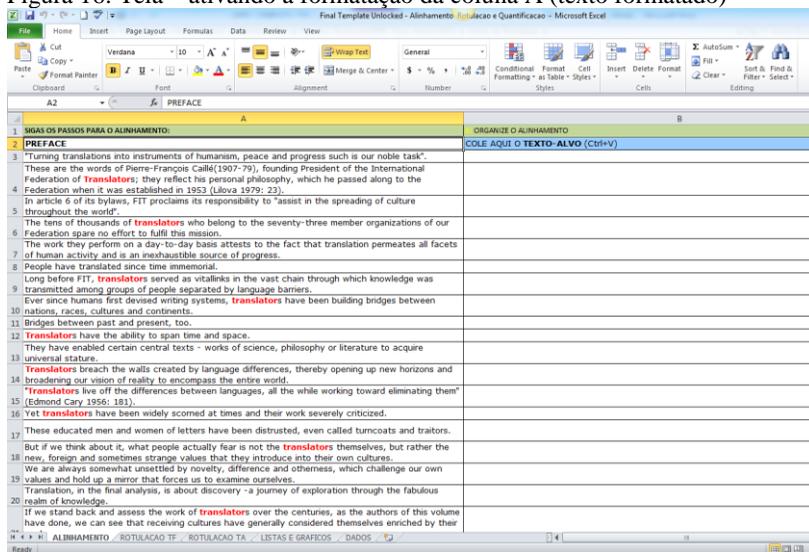


Figura 16: Tela – ativando a formatação da coluna A (texto formatado)



- (vi) Destacar, se desejar, as células que contém o nódulo a ser analisado através da ferramenta *Find & Replace*: busque pela palavra que representa o nódulo e a substitua pela mesma palavra, selecione *Formato > Fill* e escolha uma *cor de fundo*, da seguinte forma (as Figuras a seguir ilustram os comandos do Excel acionados para realizar colorir as palavras nódulos e as células que apresentam tais nódulos):

Figura 17: Tela de formatação da coluna A para alinhamento (parte 1)

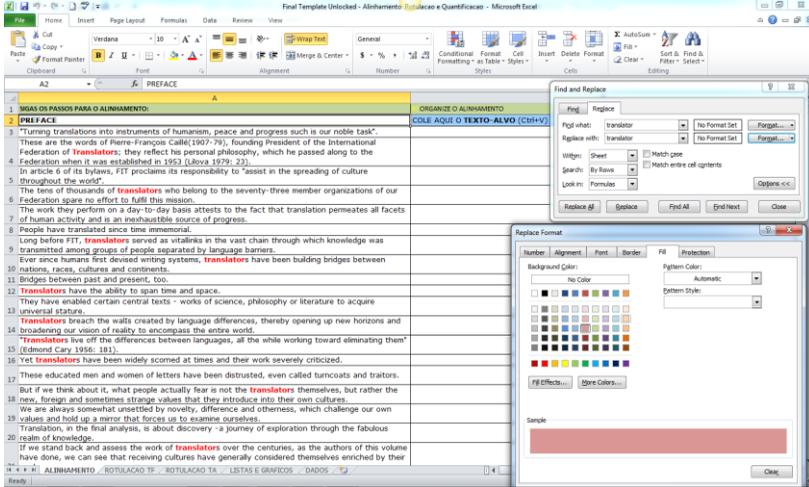
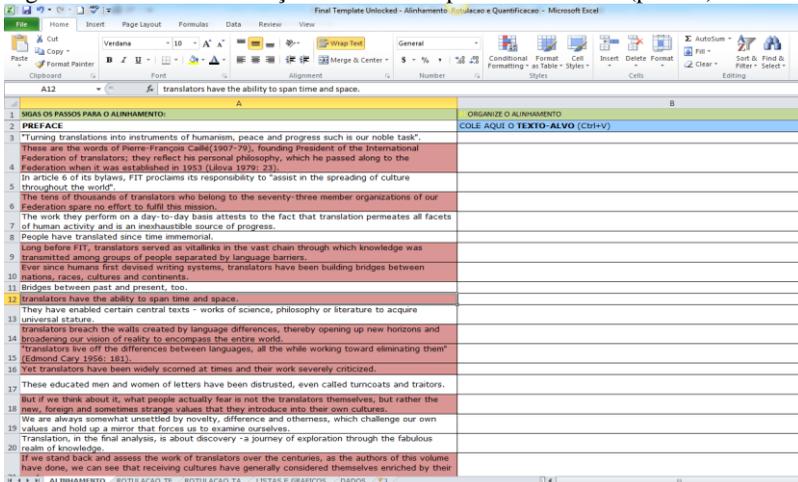


Figura 18: Tela de formatação da coluna A para alinhamento (parte 2)

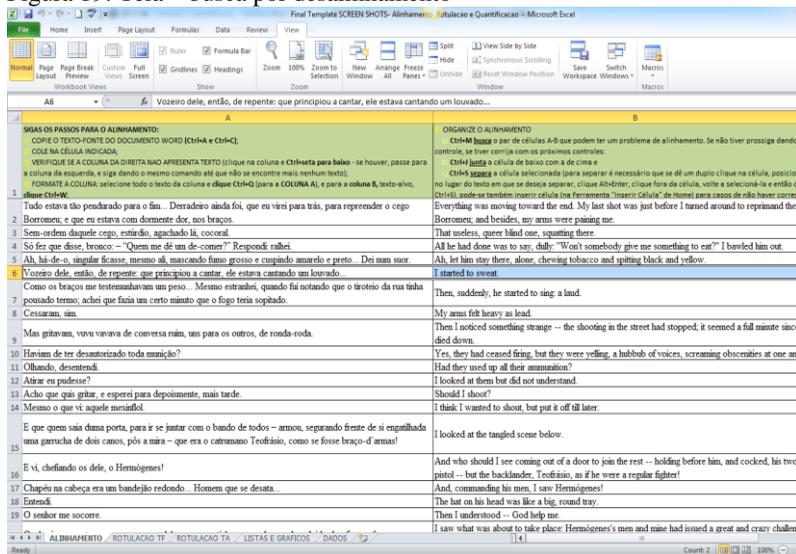


- (vii) Repetir o processo com o texto-alvo;
- (viii) Salvar como "Alinhamento";
- (ix) Organizar o alinhamento, a partir das seguintes etapas:

a) **Buscar por linhas desalinhadas (Ctrl+M).** Acionando este comando, o cursor é levado a um possível par TF/TA desalinhados. Conferir se ocorre mesmo desalinhamento. Se não ocorrer, prosseguir

acionando o comando de Busca; se ocorrer, corrigir o desalinhamento utilizando os próximos comandos. A Figura a seguir ilustra um caso de desalinhamento encontrado através desse comando:

Figura 19: Tela – busca por desalinhamento



b) **Juntar células em caso de desalinhamento (Ctrl+J).** As Figuras a seguir ilustram o antes e o depois da junção das células. Demonstra com as células destacadas em vermelho escuro como duas células podem ser combinadas:

Figura 20: Tela – ativando a função de *Juntar* do T. C.Paralelo (parte 1)

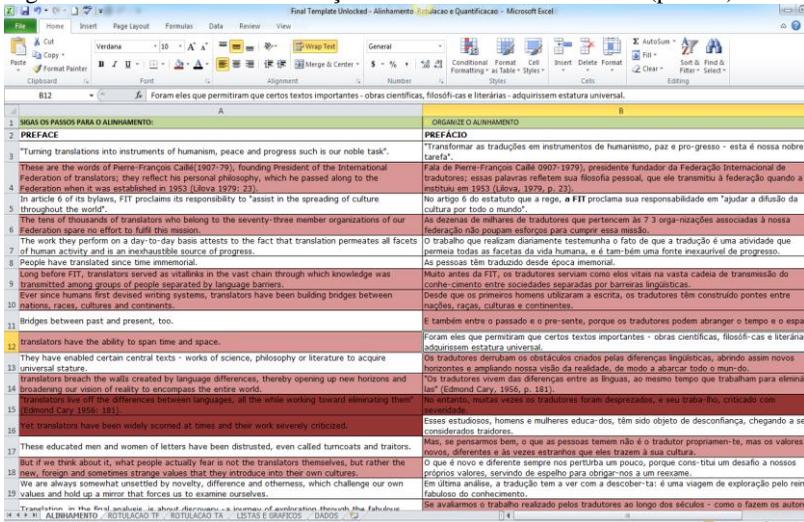
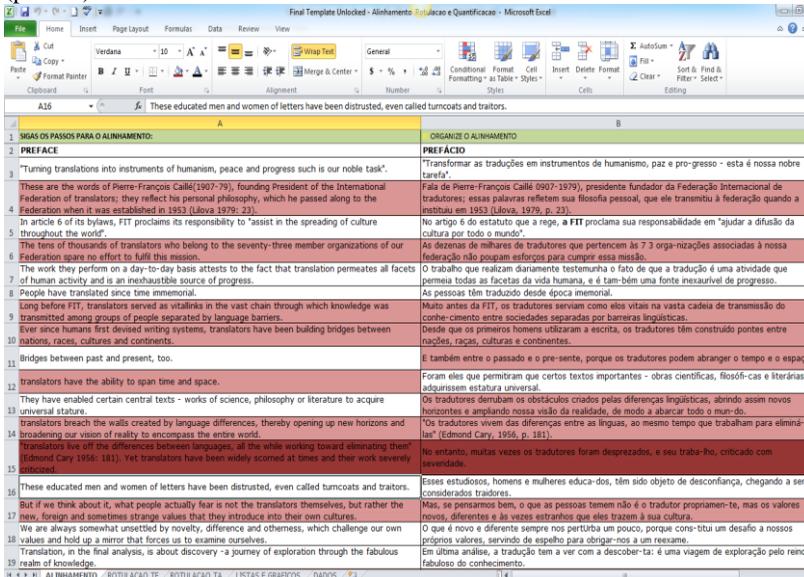


Figura 21: Tela – ativando a função de *Juntar* do Template Corpus Paralelo (parte 2)



- c) **Separar células desalinhadas (Ctrl+S)**. Separar células é um processo mais indireto. É necessário passar por três etapas, que são:
- (c.1) posicionar cursor no lugar do texto em que se deseja separar (ilustrado na Figura a seguir — o fragmento que se deseja separar é destacado em negro):

Figura 22: Tela – ativando a função de *Separar* do Template Corpus Paralelo (parte 1)

	A	B
	A36	This young discipline cannot claim a future if it is unable to build upon earlier experience and seek fresh ideas based on models from the past Constructing a history of translation
1	SIGAS OS PASSOS PARA O ALINHAMENTO:	ORGANIZE O ALINHAMENTO
2	Translation cannot be dissociated from the notion of progress; some even maintain that a society can be measured by the translations it accepts. This points to the importance of the work done by translators.	A tradução não pode ser dissociada da ideia do progresso; há mesmo quem sustente que uma sociedade pode ser avaliada pela qualidade das traduções que aceita, o que indica a importância do trabalho dos tradutores.
3	"The construction of a history of translation is the first task of a modern theory of translation" (Berman 1992: 1).	"A elaboração de uma história da tradução é a primeira tarefa de uma teoria moderna da tradução" (Berman, 1992, p. 1).
33	The study of our profession's antecedents will help to legitimate translation as an independent discipline, capable of defining itself, of sustaining a discourse sui generis.	O estudo dos antecedentes da nossa profissão nos ajudará a legitimar a tradução como uma disciplina independente, capaz de se definir, de sustentar um discurso sui generis.
35	It has already been given a name: "translation studies" or "traductologie" in French.	Já, ela já recebeu um nome: estudos de tradução ou tradutologia.
36	This young discipline cannot claim a future if it is unable to build upon earlier experience and seek fresh ideas based on models from the past Constructing a history of translation means bringing to light the complex network of cultural exchanges between people, cultures and civilizations down through the ages.	Essa nova disciplina não terá futuro se não for construída com base na experiência do passado, se não buscar ideias frescas em modelos anteriores.
37	It means drawing a portrait of these import-export workers and attempting to unravel their deep-rooted reasons for translating one particular work instead of another.	Elaborar uma história da tradução significa trazer à luz a trama complexa de trocas culturais entre os povos, as culturas e civilizações, através dos tempos.
38	It means finding out why their sponsors (kings, aristocrats, patrons, high-ranking clergy, etc.) asked them to translate a given work.	Significa retratar esses trabalhadores da importação e exportação e tentar descobrir as razões profundas que os levaram a traduzir uma obra em particular, e não outra.
39	It means taking into account what the translators themselves have written about their work, its difficulties and constraints.	Significa descobrir por que os seus patronos (monarcas, aristocratas, autoridades clérigas, etc.) lhes encomendaram a tradução de determinadas obras.
40	In short, Lieven D'huist has observed, history "is virtually the only means by which the discipline of translation studies can achieve some measure of coherence - by showing how divergent traditions of thought and activity are in fact similar or interconnected, by linking the past to the present" (1994: 13).	Significa levar em conta o que os próprios tradutores escreveram sobre o seu trabalho, suas dificuldades e limites.
41	I hope that today's translators will recognize themselves in this book devoted to their predecessors.	Em suma, como observou Lieven D'huist, a história "é virtualmente o único meio pelo qual a disciplina dos estudos de tradução pode atingir um certo grau de coerência, mostrando como tradições divergentes de pensamento e atividade são na verdade similares ou inter-conexas, associando o passado ao presente" (1994, p. 13).
42	And I include the many administrative and technical translators of the latter half of the twentieth century who, while they may not participate in cultural advances to the same extent as the individuals from the past presented here, nonetheless fulfil an essential function in modern society.	Espero que os tradutores contemporâneos se reconheçam neste livro, de-dicado aos seus predecessores.
43	Readers will appreciate the immense task involved in mapping the vast but uncharted territory of the universal history of translation.	E inclui aí os numerosos tradutores técnicos e administrativos da segunda metade do século XX que, embora possam não participar do progresso cultural na mesma medida em que o fizeram as figuras do passado aqui apresentadas, desempenham uma função essencial na sociedade moderna.
44	This work, itself only a beginning, could never have been carried out without an international team and reflecting the make-up of FIT itself.	Os leitores levarão em conta a menesidade da tarefa representada pela exploração do amplo e mal cartografado território da história universal da tradução.
45	We would like to express our heartfelt gratitude to all those who had a hand in the project, as	Este trabalho, que representa apenas um ponto de partida, não poderia ter sido realizado sem os

- (c.2) acionar o comando do Excel de quebrar sentença dentro da própria célula (**Alt+Enter**) e clicar na célula abaixo (Figura a seguir ilustra o texto da célula A36 após acionado o comando Alt+Enter):

Figura 23: Tela – ativando a função de *Juntar* do T. Corpus Paralelo (parte 2)

SÍDAS OS PASSOS PARA O ALINHAMENTO:	ORGANIZE O ALINHAMENTO:
Translation cannot be dissociated from the notion of progress; some even maintain that a society can be measured by the translations it accepts. This points to the importance of the work done by translators.	A tradução não pode ser dissociada da ideia do progresso; há mesmo quem sustente que uma sociedade pode ser avaliada pela qualidade das traduções que aceita, o que indica a importância do trabalho dos tradutores.
"The construction of a history of translation is the first task of a modern theory of translation" (Berman 1992: 1).	"A elaboração de uma história da tradução é a primeira tarefa de uma teoria moderna da tradução" (Berman, 1992, p. 1).
The study of our profession's antecedents will help to legitimize translation as an independent discipline, capable of defining itself, of sustaining a discourse sui generis.	O estudo dos antecedentes da nossa profissão nos ajudará a legitimar a tradução como uma disciplina independente, capaz de se definir, de sustentar um discurso sui generis.
It has already been given a name: "translation studies" or "traductologie" in French.	Ainda, ela já recebeu um nome: estudos de tradução ou traductologia.
This young discipline cannot claim a future if it is unable to build upon earlier experience and seek fresh ideas based on models from the past.	Essa nova disciplina não terá futuro se não for construída com base na experiência do passado, se não buscar ideias frescas em modelos anteriores.
Constructing a history of translation means bringing to light the complex network of cultural exchanges between people, cultures and civilizations down through the ages.	Elaborar uma história da tradução significa trazer à luz a trama complexa de trocas culturais entre os povos, as culturas e civilizações, através dos tempos.
It means drawing a portrait of these import-export workers and attempting to unravel their deep-rooted reasons for translating one particular work instead of another.	Significa retratar esses trabalhadores da importação e exportação e tentar descobrir as razões profundas que os levaram a traduzir uma obra em particular, e não outra.
It means finding out why their sponsors (kings, aristocrats, patrons, high-ranking clergy, etc.) asked them to translate a given work.	Significa descobrir por que os seus patronos (monarcas, aristocratas, autoridades clerical, etc.) lhes encomendaram a tradução de determinadas obras.
In short, Lieven D'huilst has observed, history "is virtually the only means by which the discipline of translation studies can achieve some measure of coherence - by showing how divergent traditions of thought and activity are in fact similar or interconnected, by linking the past to the present" (1994: 13).	Significa levar em conta o que os próprios tradutores escreveram sobre o seu trabalho, suas dificuldades e limites.
I hope that today's translators will recognize themselves in this book devoted to their predecessors.	Em suma, como observou Lieven D'huilst, a história "é virtualmente o único meio pelo qual a disciplina dos estudos de tradução pode atingir um certo grau de coerência, mostrando como tradições divergentes de pensamento e atividade são na verdade similares ou inter-conexas, associando o passado ao presente" (1994, p. 13).
And I include the many administrative and technical translators of the latter half of the twentieth century who, while they may not participate in cultural advances to the same extent as the individuals from the past presented here, nonetheless fulfil an essential function in modern society.	Espero que os tradutores contemporâneos se reconheçam neste livro, de-dicado aos seus predecessores.
Readers will appreciate the immense task involved in mapping the vast but uncharted territory of the universal history of translation.	E incluo aí os numerosos tradutores técnicos e administrativos da segunda metade do século XX que, embora possam não participar do progresso cultural na mesma medida em que o foram as figuras do passado aqui apresentadas, desempenham uma função essencial na sociedade moderna.
This work, itself only a beginning, could never have been carried out without an international team reflecting the make-up of FIT itself.	Os leitores levarão em conta a mensuração da tarefa representada pela exploração do amplo e mal cartografado território da história universal da tradução.
We would like to express our heartfelt gratitude to all those who had a hand in the project, as	Este trabalho, que representa apenas um ponto de partida, não poderia ter sido realizado sem os

(c.3) voltar a selecionar a célula que se deseja separar e acionar o comando criado neste *Template* para separar células (Ctrl+S). A Figura a seguir ilustra a célula A36 separada em duas, destacadas em amarelo:

Figura 24: Tela – ativando a função de *Juntar* do Template Corpus Paralelo (parte final)

A	B
1 SIGAS OS PASSOS PARA O ALINHAMENTO:	ORGANIZE O ALINHAMENTO
2 Translation cannot be dissociated from the notion of progress; some even maintain that a society can be measured by the translations it accepts. This points to the importance of the work done by translators.	A tradução não pode ser dissociada da ideia do progresso; há mesmo quem sustente que uma sociedade pode ser avaliada pela qualidade das traduções que aceita, o que indica a importância do trabalho dos tradutores.
3 "The construction of a history of translation is the first task of a modern theory of translation" (Berman 1992: 1).	"A elaboração de uma história da tradução é a primeira tarefa de uma teoria moderna da tradução" (Berman, 1992, p. 1).
4 The study of our professor's antecedents will help to legitimize translation as an independent discipline, capable of defining itself, of sustaining a discourse sui generis.	O estudo dos antecedentes da nossa profissão nos ajudará a legitimar a tradução como uma disciplina independente, capaz de se definir, de sustentar um discurso sui generis.
5 It has already been given a name: "translation studies" or "traductologia" in French.	Aliás, ela já recebeu um nome: estudos de tradução ou traductologia.
6 This young discipline cannot claim a future if it is unable to build upon earlier experience and seek fresh ideas based on models from the past.	Essa nova disciplina não terá futuro se não for construída com base na experiência do passado; se não buscar ideias frescas em modelos anteriores.
7 Constructing a history of translation means bringing to light the complex network of cultural exchanges between people, cultures and civilizations down through the ages.	Elaborar uma história da tradução significa trazer à luz a trama complexa de trocas culturais entre os povos, as culturas e civilizações, através dos tempos.
8 It means drawing a portrait of these import-export workers and attempting to unravel their deep-rooted reasons for translating one particular work instead of another.	Significa retratar esses trabalhadores da importação e exportação e tentar descobrir as razões profundas que os levaram a traduzir uma obra em particular, e não outra.
9 It means finding out why their sponsors (kings, aristocrats, patrons, high-ranking clergy, etc.) asked them to translate a given work.	Significa descobrir por que os seus patronos (monarcas, aristocratas, autoridades clero-cais, etc.) lhes encomendaram a tradução de determinadas obras.
10 It means taking into account what the translators themselves have written about their work, its difficulties and constraints.	Significa levar em conta o que os próprios tradutores escreveram sobre o seu trabalho, suas dificuldades e limitações.
11 In short, Lieven D'huut has observed, history is virtually the only means by which the discipline of translation studies can achieve some measure of coherence - by showing how divergent traditions of thought and activity are in fact similar or interconnected, by linking the past to the present" (1994: 41, 13).	Em suma, como observou Lieven D'huut, a história "é virtualmente o único meio pelo qual a disciplina dos estudos de tradução pode atingir um certo grau de coerência, mostrando como tradições divergentes de pensamento e atividade são na verdade similares ou inter-conexas, associando o passado ao presente" (1994, p. 13).
12 I hope that today's translators will recognize themselves in this book devoted to their predecessors.	Espero que os tradutores contemporâneos se reconheçam neste livro, dedicado aos seus predecessores.
13 And I include the many administrative and technical translators of the latter half of the twentieth century who, while they may not participate in cultural advances to the same extent as the individuals from the past presented here, nonetheless fulfil an essential function in modern society.	E incluo aí os numerosos tradutores técnicos e administrativos da segunda metade do século XX que, embora possam não participar do progresso cultural na mesma medida em que o fizeram as figuras do passado, aqui apresentadas, desempenham uma função essencial na sociedade moderna.
14 Readers will appreciate the immense task involved in mapping the vast but uncharted territory of the universal history of translation.	Os leitores levarão em conta a imensidão da tarefa representada pela exploração do amplo e mal cartografado território da história universal da tradução.
15 This work, itself only a beginning, could never have been carried out without an international team reflecting the make-up of FIT itself.	Este trabalho, que representa apenas um ponto de partida, não poderia ter sido realizado sem os esforços de um grupo internacional que reflete a constituição da própria FIT.
16 We would like to express our heartfelt gratitude to all those who had a hand in the project, as contributing authors, editors, translators or proofreaders, helping to produce the original work from a wealth of documents.	Gostariamos de expressar nossa sincera gratidão a todos os que participaram do projeto como autores, editores, tradutores ou revisores, ajudando a produzir este trabalho original a partir de uma riqueza de documentos.
17 Some fifty individuals from twenty different countries helped to mount this impressive portrait gallery of	Cerca de cinquenta pessoas, de vinte países diferentes, ajudaram a montar esta impressionante

d) Inserir célula da seguinte forma: na Função "Inserir Célula" de *Home*: Menu: *Inserir* > *Inserir Célula* > *Shift cells down*. Essa série de comandos é ilustrada nas seguintes Figuras:

Figura 25: Tela – inserindo linha em branco (parte 1)

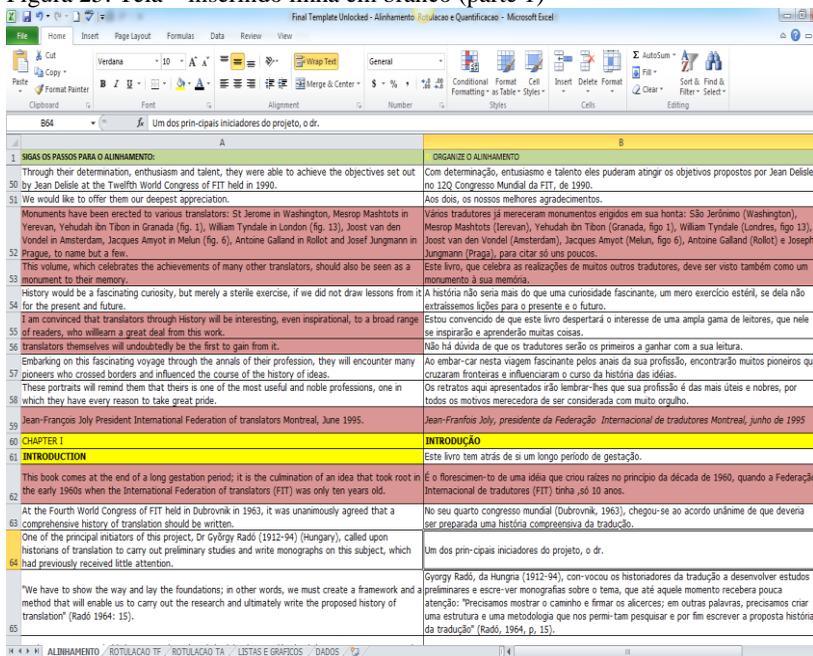


Figura 26: Tela - inserindo linha em branco (parte 2)

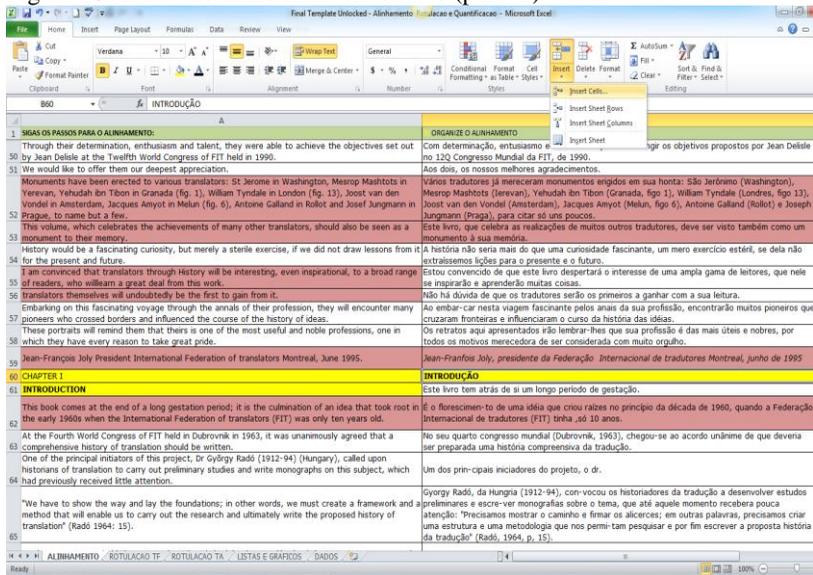


Figura 27: Tela - inserindo linha em branco (parte 3)

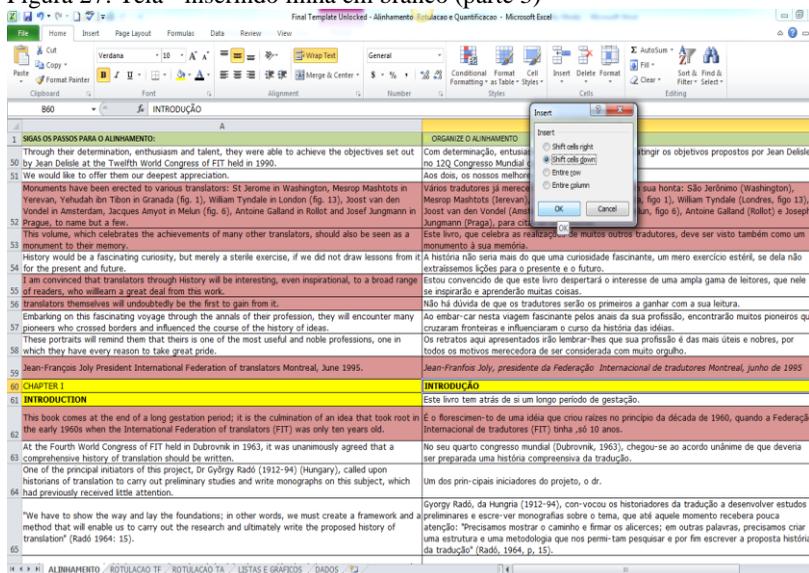
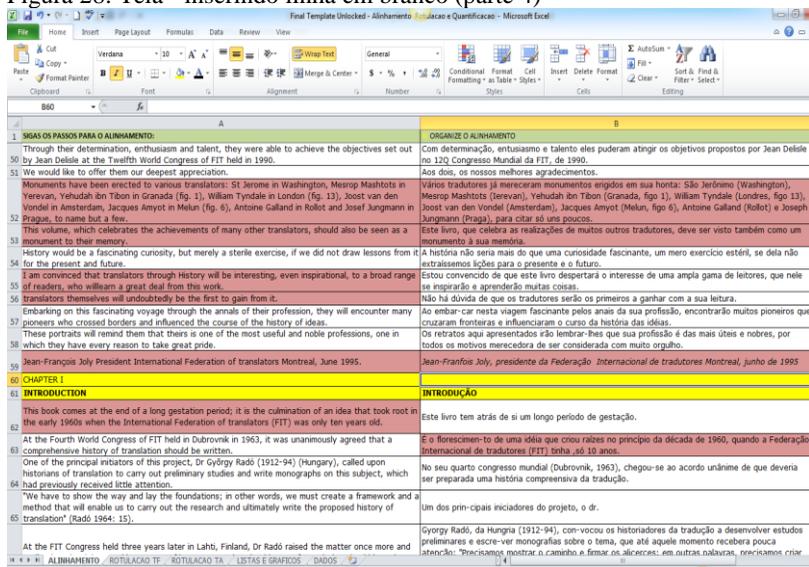


Figura 28: Tela - inserindo linha em branco (parte 4)



(x) Salvar o alinhamento frequentemente, visto que os comandos criados (Ctrl+Q, Ctrl+W, Ctrl+J e Ctrl+S) são Macros que

não permitem que suas ações sejam desfeitas, isto é, o Ctrl+Z não tem efeito sobre essas ações.

O processo de alinhamento é concluído quando não há mais ocorrências de desalinhamentos. O pesquisador pode optar em formatar o *layout* do alinhamento, por exemplo, destacando com cor diferente as linhas em branco, para melhor identificar as ausências de tradução ou de texto-original no alinhamento. Esses processos de formatação em geral encontram-se disponíveis nos comandos do próprio *Excel*. Para destacar com uma cor as linhas em branco, por exemplo, acionam-se uma série de comandos que são: *Find & Replace > Go to Special > Blank*, selecione uma cor para preencher as células em branco. O resultado é um corpus bilíngue alinhado em paralelo, com as ausências de tradução/texto original destacados, conforme ilustrado na Figura a seguir:

Figura 29: Tela Template Corpus Paralelo – formatação final, colorindo linhas em branco

	A	B
	Como vinham de lá e de lá, em contra-rancho, a tomar armas, as cartucheiras de tiracolo.	
1	SIGAS OS PASSOS PARA O ALINHAMENTO: COPIE O TEXTO-FONTE DO DOCUMENTO WORD (Ctrl+A e Ctrl+C). COLE NA CÉLULA INDICADA. VERIFIQUE SE A COLUNA DA DIREITA NÃO APRESENTA TEXTO (clique na coluna e Ctrl+seta para baixo - se houver, passe para a coluna da esquerda, e siga dando o mesmo comando até que não se encontre mais nenhum texto). FORMATAR A COLUNA: selecione todo o texto da coluna e clique Ctrl+Q (para a COLUNA A), e para a coluna B, texto-alto.	ORGANIZE O ALINHAMENTO Ctrl+M basta o par de células A-B que podem ter um problema de alinhamento. Se não tiver prossegua dando este comando, se não corrigir com os próximos comandos: Ctrl+J junta a célula de baixo com a de cima e Ctrl+S separa a célula selecionada (para separar é necessário que se dê um duplo clique na célula, posicione o cursor no lugar do texto em que se deseja separar, clique Alt+Enter, clique fora da célula, volte a selecionar a célula e dê um clique Ctrl+Q), onde se houver inserir célula na ferramenta "Inserir Célula" de nome para caso de não haver correspondência.
8	Cessaram, sim. Mas gritavam, vivia vivava de conversa riam, uns para os outros, de ronda-ronda.	Yes, they had ceased firing, but they were yelling, a hubbub of voices, screaming obscenities at one another.
9	Haviam de ter desautorizado toda minação?	Had they used up all their ammunition?
10	Ofandando desentendi.	I looked at them but did not understand.
11	Aftrar eu pudesse?	Should I shoot?
12	Acho que quis gritar, e esperei para depoismente, mais tarde.	I think I wanted to shout, but put it off all later.
13	Mesmo o que vi aparte metralhad.	I looked at the tangled scene below.
14	E que quem saia duna porta, para si se juntar com o bando de todos - amou, segurando frente de si enghalhada uma garrafa de dois canos, pôs a mira - que era o catramano Teofrasto, como se fosse braco-d'armas!	And who should I see coming out of a door to join the rest - holding before him, and cocked, his two-barreled pistol -- but the backlander, Teofrasto, as if he were a regular fighter!
15	E vi, chafando os dentes, o Hermógenes!	And, commanding his men, I saw Hermógenes!
16	Chapéu na cabeça era um bandejeilo redondo. Homem que se desata.	The hat on his head was like a big, round tray.
17	Entendi. O senhor me socorre.	Then I understood -- God help me.
18	Codeci o que estava para ser: que os dele e os meus tinham cruzado grande e doído desafio, conforme para cumprir se armanavam, uns e outros, nas duas pontas da rua, debateo de forma, e a fio desembainhavam.	I saw what was about to take place: Hermógenes's men and mine had issued a great and crazy challenge to each other, and were preparing to carry it out: they were gathering at opposite ends of the street -- and drawing their knives.
19	O que vendo, vi Diadorim - movimentos dele.	Looking I saw Diadorim and watched his movements.
20	Querer não gritar, e não pude, demais de mim-mesmo, me tonteava, namas insias.	I strove to shout but couldn't, I was beside myself, my head was spinning.
21	E inha o inferno daquela rua, para escutar compredo. Tiraram minha voz.	That street was like a long pit in hell I lost my voice.
22	Como vinham de lá e de lá, em contra-rancho, a tomar armas, as cartucheiras de tiracolo.	
23	Aftrar eu pude?	
24	A brace torceu e lesou meus braços, estorvados. Pela espinha doíam, em suar em fo veritigioso.	A cramp twisted and paralyzed my arms. A thread of sweat ran down my spine.
25	Quem era que desbracava e me peava, suplindo minhas forças? - "Tua honra. Minha honra de homem de valores!" - "eu me, em mim, geni alma que perdeu o corpo.	Who was disarming and stacking me, robbing me of my strength? "My honor -- my honor as a brave man!" I said, groaning, to myself a soul which had lost its body.
26	O fuc caiu de minhas mãos, que nem pude separar com o queito e com os peitos.	The gun fell from my hands, I could not even hold it against my chest.

O resultado final é dois textos alinhados lado-a-lado com os segmentos em cujo os títulos se encontram destacados para uma busca visual mais rápida e uma análise do corpus paralelo mais acurada.

5.2.2. Pasta de Rotulação TF e a Pasta de Rotulação TA

As Pastas de Rotulação TF e a de Rotulação TA apresentam mesmo *layout*, fórmulas e *macros*. A diferença é que uma é dedicada ao texto-fonte e outra ao texto-alvo, ambos colados na coluna A de cada pasta. Nestas pastas é realizada a classificação dos objetos analisados. Conforme ilustrado na Figura a seguir:

Figura 30: Tela T. Corpus Paralelo– pasta “Rotulação TF”

PARA A ROTULAÇÃO SIGA OS PASSOS: Copie a coluna A da pasta de "Alinhamento" e cole na célula indicada. Escreva o nome das categorias para sua análise nas células "CATEG"; adicione mais colunas se necessário ("insert coluna") ou utilize "Clonar coluna"; na coluna A, as células destacadas contém o título de sua pesquisa, fuque essas células para a categorização; marque células (EM) (a coluna A se necessário, por cada linha deve equivar a uma categoria; antes a categoria marcando "1" na célula, só se pode marcar uma linha por vez); para que funcione as formulas da coluna "Categorias Correspondentes"; se desejar, escreva na coluna Q ao que se refere exatamente a categoria analisada (se a categoria é, por exemplo, Processo Material, na coluna Q você escreve o verbo que está sendo rotulado, por exemplo, "fazer"); salve como "Anotação"; Para visualizar o Gráfico, selecione todas as categorias, mantenha a Tecla Ctrl pressionada e selecione as células de número, correspondente a Tecla "Total" com ambas linhas selecionadas, clique em "Inserir>Gráfico" e selecione o Gráfico que quiser. Acoluna A e as linhas 1 e 2 foram FREEZE, se desejar UNFREEZE, vá em "View>Freeze Panels>Unfreeze Panel". PARA A QUANTIFICAÇÃO SE OS GRÁFICOS CORRESPONDEM AS DE SUAS CATEGORIAS (PARA A TABELA DO TEMPLATE SE CONSIDERAM APENAS AS LINHAS R2,N2, REVISAR CASO VOCÊ INSIRA MAIS OU MENOS CATEGORIAS) E A LINHA DO TOTAL (NESTE TEMPLATE O TOTAL SE ENCONTRA NA LINHA 104,699).

COLA AQUI O TEXTO-FONTE DO "ALINHAMENTO"	CAT01	CAT02	CAT03	CAT04	CAT05	CAT06	CAT07	CAT08	CAT09	CAT10	CAT11	CAT12	CAT13	CAT14	CAT15	INSERIMENTO	CATEGORIA	DOWNT
			1													verbo1	CATEG1	
				1												verbo2	CATEG2	
					1											verbo3	CATEG3	
						1										verbo4	CATEG4	
							1									verbo5	CATEG5	
								1								verbo6	CATEG6	
									1							verbo7	CATEG7	
										1						verbo8	CATEG8	
											1					verbo9	CATEG9	
												1				verbo10	CATEG10	
													1			verbo11	CATEG11	
														1		verbo12	CATEG12	
															1	verbo13	CATEG13	
																verbo14	CATEG14	
															1	verbo15	CATEG15	
																verbo16	0	
																0	0	
																0	0	
																0	0	
																0	0	
																0	0	
																0	0	
																0	0	

- A **coluna A** é dedicada ao texto (fonte e alvo);
- A primeira linha das **colunas B a P** são dedicadas à nomenclatura dos elementos de classificação (no *Template Corpus Paralelo* são chamados de categoria 1, categoria 2 até categoria 15, esses nomes são provisórios e devem ser substituídos pelas categorias consideradas na pesquisa);
- A **coluna Q** é dedicada à transcrição do elemento analisado. Enquanto na coluna A contém elemento analisado se encontra inserido em seu contexto, dentro da oração ou complexo oracional, na coluna Q esse elemento pode ser digitado pelo pesquisador, com o intuito de destacá-lo para futura revisão e listagem;
- A **coluna R** contém uma fórmula que copia o nome da categoria de B2 a P2 que é marcada. Então se uma determinada categoria, por exemplo a categoria escrita em I2, é marcada na linha I36 com o número 1, a célula R36 receberá o mesmo nome da

categoria em I2. As colunas Q e R lado-a-lado permitem que se visualize o objeto analisado ao lado da classificação dada a este objeto, permitindo uma revisão rápida da análise.

- A **coluna S** é dedicada a anotações do pesquisador em relação à linha analisada.

Para que as fórmulas não sejam alteradas, recomenda-se que nenhuma coluna seja deletada. Se for necessário excluir alguma coluna, aconselha-se que se escolha a opção de “esconder” (*hide*) coluna.

As etapas para se realizar a rotulação são as seguintes:

- Apagar o texto de exemplo da célula B2 à Q18. **Não apagar nenhuma célula da coluna R**, pois esta apresenta fórmulas.
- Copiar a coluna A da Pasta de "Alinhamento" e colar na célula A2 da pasta de “Rotulação TF”;
- Escrever o nome das categorias para sua análise nas células "CATEG" (B2 a P2). **Não inserir ou deletar colunas antes da coluna Q**;
- Separar células (Ctrl+S) da coluna A se necessário, pois **cada linha deve equivaler a uma categoria**;
- Anotar a categoria marcando "1" na célula da coluna da categoria correspondente à linha em que se encontra o objeto a ser classificado. Não marcar duas categorias para um mesmo elemento, isso leva a não anotação automática da categoria na coluna R;
- Se desejar, escrever na coluna Q ao que se refere exatamente a categoria analisada (se a categoria é, por exemplo, Processo Material, na coluna Q você escreve o verbo que está sendo rotulado, por exemplo, "fazer");
- Salvar o documento.

A Figura abaixo ilustra uma pasta de “Rotulação TF” anotada:

Figura 31: Tela T. Corpus Paralelo– Classificação dos elementos analisados

Final Template Unlocked - Alinhamento - Rotulacao e Quantificacao - Microsoft Excel

PARA A ROTULACAO SIGA OS PASSOS: Copie a coluna A da pasta de "Alinhamento" e cole na célula indicada. Escreva o nome das categorias para sua análise nas células "CATEG", adicione mais colunas se necessário ("insert column") ou delete as colunas que não deseja ("delete column").

	MATERIAL	MENTAL	RELACIONAL	VERBAL	BEHAV	EXIST	ELEMENTO	CATEGORIA	Correspondente	Duvidas										
3																				
4	PREFACE	ACTOR	GOAL	CLIENT	RECIPIENT	SENS	PHENO	IDENTR	IDENTD	SAVER	VERBI	ADDRESS	BEHAV	EXIST						
5	Turning translations into instruments of humanism, peace and progress such is our noble task.																			
6	These are the words of Pierre-François Callet (1907-79), founding President of the International Federation of translators; they reflect his personal philosophy, which he passed along to the Federation when it was established in 1953 (Léves 1976: 22).																			
7	In article 6 of its bylaws, FIT proclaims its responsibility to assist in the spreading of culture throughout the world.																			
8	The tens of thousands of translators who belong to the seventy-three member organizations of our Federation spare no effort to fulfill this mission.																			
9	The work they perform on a day-to-day basis attests to the fact that translation permeates all facets of human activity and is an inexhaustible source of progress.																			
10	People have translated since time immemorial.																			
11	Long before FIT, translators served as trailblazers in the vast chain through which knowledge was transmitted among groups of people separated by language barriers.																			
12	Ever since humans first devised writing systems, translators have been building bridges between nations, races, cultures and continents.																			
13	Tridgen between east and present, too.																			
14	Translators have the ability																			
15	to span time and space.																			
16	They have enabled certain central texts - works of science, philosophy or literature to acquire universal status.																			
17	Translators breach the walls created by language differences																			
18	thereby opening up new horizons and																			
19	broadening our vision of reality.																			
20	To encompass the entire world.																			
21	Translators live off the differences between languages,																			
22	and the while working																			
23	toward eliminating them" (Edmond Cary 1956: 181).																			
24																				

Final Template SCREEN SHOTS: Alinhamento - Rotulacao e Quantificacao - Microsoft Excel

B104695 =SUM(B2:B104694)

A linha 104695 é dedicada ao cálculo automático do total, conforme ilustra a Figura a seguir:

Figura 32: Tela Template Corpus Paralelo – gerando gráficos (parte 1)

Final Template SCREEN SHOTS: Alinhamento - Rotulacao e Quantificacao - Microsoft Excel

B104695 =SUM(B2:B104694)

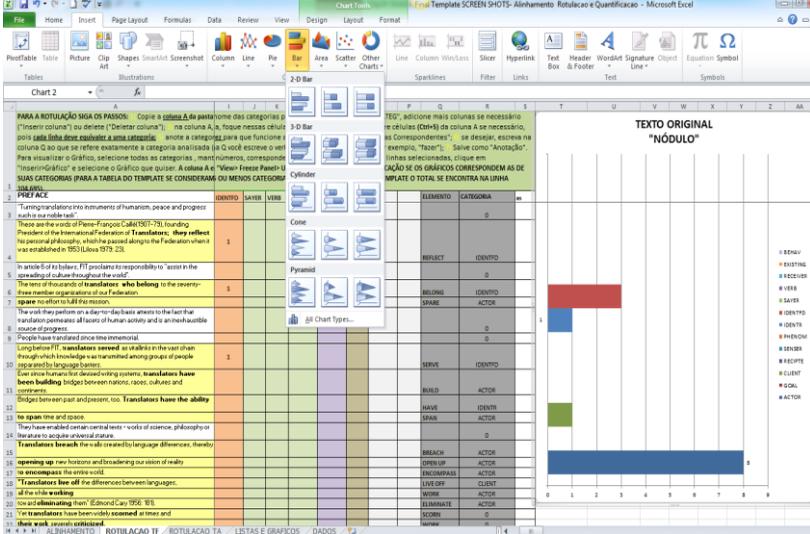
PARA A ROTULACAO SIGA OS PASSOS: Copie a coluna A da pasta de "Alinhamento" e cole na célula indicada. Escreva o nome das categorias para sua análise nas células "CATEG", adicione mais colunas se necessário ("insert column") ou delete as colunas que não deseja ("delete column"). Para cada linha deve escolher a uma categoria; anote a categoria marcando "1" na célula, se se pode marcar uma linha por vez para que funcione as formulas de coluna "Categorias Correspondentes", se desejar, escreva na coluna Q o que se refere exatamente a categoria analisada (pe a categoria é, por exemplo, Processo Material), na coluna Q você escreve o verbo que está sendo rotulado, por exemplo, "fazer". Salve como "Anotação". Para visualizar o Gráfico, selecione todas as categorias, mantenha a tecla Ctrl pressionada e selecione as células de número, correspondente à linha "Total", com ambas linhas selecionadas, clique em "Inserir>Gráfico" e selecione o Gráfico que quiser. A coluna A e as linhas 1 e 2 foram FREEZED, se desejar UNFREEZE, vá em "View>Freeze Panel>Unfreeze Panel". PARA A QUANTIFICACAO SE OS GRAFICOS CORRESPONDEM AS SUAS CATEGORIAS PARA A TABELA DO TEMPLATE SE CONSIDERAM APENAS AS LINHAS B2:B23. REVISAR CASO VOCÊ INSIRA MAIS OU MENOS CATEGORIAS E A LINHA DO TOTAL (NESTE TEMPLATE O TOTAL SE ENCONTRA NA LINHA 104695).

	MATERIAL	MENTAL	RELACIONAL	VERBAL	BEHAV	EXIST	ELEMENTO	CATEGORIA	Correspondente	Duvidas										
2	PREFACE	ACTOR	GOAL	CLIENT	RECIPIENT	SENS	PHENO	IDENTR	IDENTD	SAVER	VERBI	RECEIVER	EXISTING	BEHAV						
304683																				
304684																				
304685																				
304686																				
304687																				
304688																				
304689																				
304701																				
304702																				
304703																				
304704																				
304705																				
304706																				
304707																				
304708																				
304709																				
304710																				
304711																				
304712																				
304695	TOTAL																			
304696																				
304697																				
304698																				
304699																				
304700																				
304701																				
304702																				
304703																				
304704																				
304705																				
304706																				
304707																				
304708																				
304709																				
304710																				
304711																				
304712																				

Final Template SCREEN SHOTS: Alinhamento - Rotulacao e Quantificacao - Microsoft Excel

Um gráfico que proporcione a quantificação de cada categoria pode ser gerado ao se selecionar as células B104695 a P104695 e as células B2 a P2. A Figura a seguir representa a geração de um gráfico a partir desta tabela:

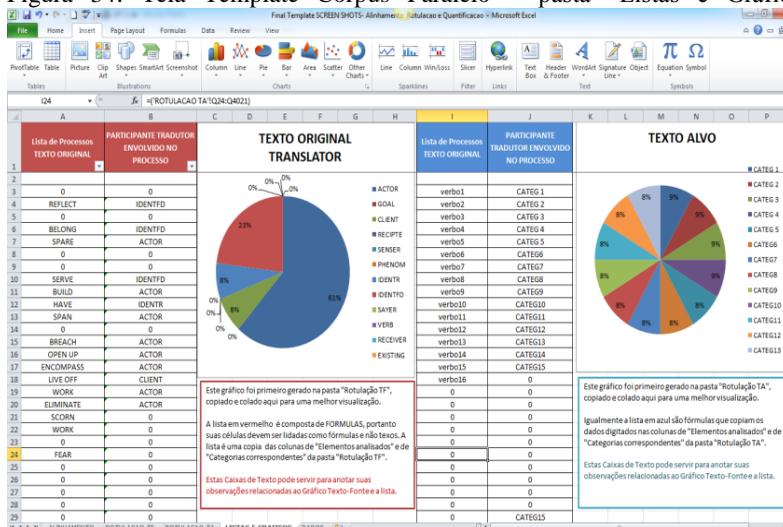
Figura 33: Tela Template Corpus Paralelo – gerando gráficos (parte 2)



5.2.3. Pasta de Listas e Gráficos

Na “Pasta de Listas e Gráficos” são expostos os gráficos de quantificação dos elementos classificados e as listas de elementos e suas categorias. Esta pasta não apresenta Macros, apenas fórmulas que copiam dados da pasta de “Rotulação TF” e de “Rotulação TA” e os organiza em listas e os quantifica em gráficos. Esta pasta apresenta o seguinte *layout*:

Figura 34: Tela Template Corpus Paralelo – pasta “Listas e Gráficos”



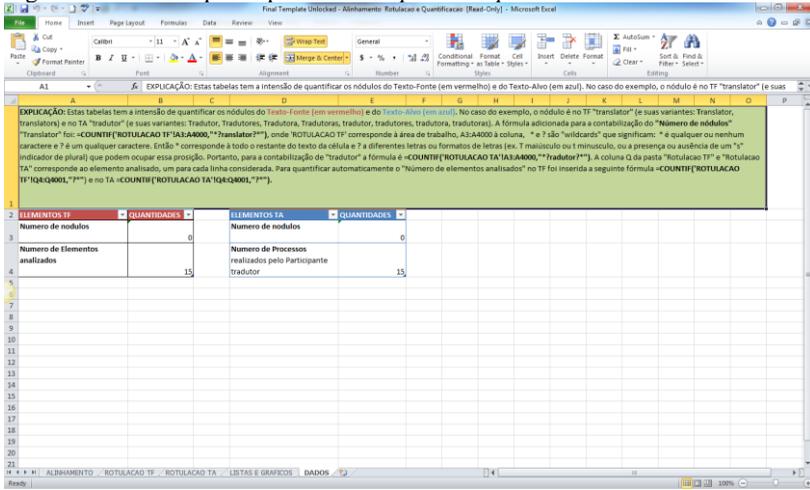
Esta pasta expõe listas de gráficos do TF e do TA lado-a-lado, para permitir uma comparação direta. A lista em detalhes vermelhos e o gráfico ao lado correspondem ao texto-fonte e a em detalhes azuis e o gráfico ao lado correspondem ao texto-alvo. As caixas abaixo do gráfico explicam a pasta e com a tal explicação apagada podem servir como um espaço para o pesquisador anotar suas observações. As listas e os gráficos são gerados automaticamente, a medida que os dados são anotados nas pastas de “Rotulação TF” e de “Rotulação TA”. No caso da Figura acima apenas o TF foi classificado, o TA encontra-se ainda no padrão original do *Template Corpus Paralelo*.

5.2.4. Pasta de Dados

A pasta de “Dados”, como mencionado no Capítulo 3, é criada para demonstrar as possibilidades de quantificação de certos termos que se busca no texto-fonte e no texto-alvo. Como a compilação de corpus realizada no Estudo Piloto é comparado com a compilação do corpus em Fleuri (2006), considerou-se a possibilidade de quantificar os nódulos “tradutor” e “translator” anotados/analísados e os mesmos nódulos presentes em todo o TF e TA. Para isso, são desenvolvidas fórmulas relacionadas ao TF e ao TA da pasta de “Alinhamento” e aos elementos

anotados das pastas de “Rotulação TF” e de “Rotulação TA”. Esta pasta apresenta o seguinte *layout*:

Figura 35: Tela da quinta pasta de *Template Corpus Paralelo*



Há duas listas de duas linhas e duas colunas. A lista em detalhes vermelho corresponde ao TF e a em detalhes azuis corresponde ao TA. Cada lista exhibe o número de nódulos presentes em cada texto, assim como o número de elementos analisados. Isso indica ao pesquisador a quantidade de nódulos “tradutor” em relação ao número de Processos que tal nódulo realiza como Participante.

Para a elaboração do *Template Corpus Paralelo* foram observadas as necessidades mais emergentes nas pesquisas de mestrado em ESFTBC. Observou-se que 75% das 20 pesquisas lidaram com corpus paralelo bilíngue alinhado de pequena dimensão (ver capítulo de análise desta tese). Praticamente metade das pesquisas optaram por um alinhamento manual em tabelas e outra metade por um alinhamento semiautomático, com o auxílio do utilitário *Viewer and Aligner* do *WST*. Considerando que 100% dos pesquisadores mencionaram ou indicaram de alguma forma ter trabalhado ou acessado durante a produção da dissertação o *MS Office*, conclui-se que tal programa é familiar e acessível a todos. Como o objetivo desta tese é oferecer um método acessível, econômico e prático à pesquisa em ESFTBC, pensou-se em desenvolver um *Template* de alinhamento utilizando tal Suíte de Aplicativos para Escritório.

Concluo que com esta proposta de compilação de corpus paralelo bilíngue de pequena dimensão, o *MS Excel* é ideal para lidar com tabelas e gráficos, ambos recursos escolhidos pelos pesquisadores que organizam seu corpus alinhado manualmente e que expõem a quantificação dos dados em gráficos. Isso porque suas tabelas podem ser formatadas automaticamente, oferecendo uma praticidade de organização, além de uma interface visualmente organizada de alinhamento e de listagem. Seu sistema de fórmulas pode oferecer uma quantificação imediata dos objetos analisados e classificados, podendo ser ilustrada simultaneamente em gráficos. Entretanto, por ser um programa desenhado para lidar principalmente com números⁸⁴, lidar com textos demanda criatividade e planejamento do elaborador/adaptador do *Template*, pois certos comandos devem ser desenvolvidos. É o caso das funções de “juntar” textos de células diferentes ou de “separá-los”, desenvolvidos em *Macros* na pasta “Alinhamento”. Algumas funções criadas, como a de “buscar” células desalinhadas nessa pasta, demandam combinações de fórmulas matemáticas com letras; uma mistura de funções numéricas com funções textuais. Em todos os casos, o *MS Excel* demonstra-se ser capaz de recriar os comandos presentes no *Viewer and Aligner* do *WST* e de quantificar os objetos analisados, respondendo às necessidades metodológicas de compilação de corpus paralelo bilíngue, demonstrada na maioria das pesquisas em ESFTBC desenvolvidos no Brasil de 2003 a 2010.

6 ESTUDO PILOTO: TESTE DOS TEMPLATES

Neste capítulo, tomo os processos de compilação da minha pesquisa de mestrado (Fleuri, 2006, p. 65-87), realizados com o auxílio do *WordSmith Tools* (Scott, 1999), e os comparo com os processos de compilação propostos nesta tese, que utilizam o *Template Editor* e o *Template Corpus Paralelo*. Os processos de compilação observados em Fleuri (2006) e testados no Estudo Piloto são descritos nas seguintes subseções: (6.1) Preparação dos textos para o alinhamento; (6.2) Alinhamento; (6.3) Anotação do Corpus / Classificação dos objetos analisados; (6.4) Revisão desta rotulação e (6.5) Quantificação dos dados e o modo de exposição.

Neste Estudo Piloto não é considerado o processo de escaneamento e de correção dos erros de escaneamento. Como a proposta desta pesquisa é: ‘oferecer um método econômico que **agilize** e organize

⁸⁴ “Fonte: http://en.wikipedia.org/wiki/Microsoft_Excel, em 06/09/13

os processos de compilação, que **diminua o número das etapas** em cada processo, **de transições** entre programas e **da geração de diferentes documentos** e que **flexibilize o modo de anotação** do corpus⁷; comparo, a partir da observação que realizo em Fleuri (2006) e dos testes nos *Templates*, as etapas de compilação em relação ao (a) tempo gasto, ao (b) número de etapas, ao (c) número de transições entre programas utilizados no processo de compilação e (d) o número de documentos gerados ao final do processo. Ainda que o tempo gasto na pesquisa de mestrado não tenha sido medido, a descrição detalhada dos processos de compilação no Capítulo de Metodologia da dissertação de Fleuri (2006) e minha experiência como autora desta dissertação oferecem-me uma estimativa de tempo em termos de “minutos”, “horas”, “dias” e “semanas”. Já no Estudo Piloto nos *Templates*, o tempo é medido nesses mesmos termos.

Cada processo de compilação corresponde a uma seção deste capítulo. O corpus em Fleuri (2006) é composto por dois textos: o TF com 148 páginas (Fonte Verdana, tamanho 10, espaço simples) e o TA com 146 páginas (idem), que foram escaneados e tiveram os erros de leitura de escâner corrigidos. Este mesmo corpus é utilizado no Estudo Piloto. Cada subseção descreve primeiramente o processo de compilação e suas etapas em Fleuri (2006) —de acordo com o que está descrito na própria dissertação, em seguida o processo de compilação e suas etapas no Estudo Piloto realizado com *Templates* propostos nesta tese. Ao final comparo entre Fleuri (2006) e o Estudo Piloto o (i) tempo gasto, (ii) número de etapas em cada processo de compilação; (iii) o número de transições entre programas utilizados em cada processo e o (iv) número de documentos gerados ao final de cada processo de compilação.

6.1. PREPARAÇÃO PARA O ALINHAMENTO

Em Fleuri (2006, p. 74) os textos (TF e TA), após terem sido escaneados e os erros de escaneamento corrigidos, foram preparados para o alinhamento automático no *Viewer and Aligner* do *WordSmith Tools* (*WST*). Tal preparação se deu parcialmente no próprio documento *MS Word* e parcialmente no *WST*, passando pelas seguintes etapas:

- (i) Ler cada texto e identificar as abreviações mais recorrentes (i.e. Rev., Cap., U.S.A., fig., Sr., 1. Dr. entre outros);
- (ii) Substituir através do *Find & Replace* tais abreviações, substituindo a abreviação com ponto pela mesma abreviação sem ponto;

- (iii) Quebrar os textos em sentenças, utilizando o comando do *Find & Replace*, substituindo ponto final por “Quebra de linha manual”;
- (iv) Salvar cada texto em *Plain Text*;
- (v) Alinhar ambos os textos no *Viewer and Aligner* do *WordSmith Tools*;
- (vi) Identificar as ocorrências de desalinhamento por presença de abreviações;
- (vii) Voltar ao documento *Word* e substituir tais abreviações através do *Find & Replace*;
- (viii) Salvar em *Plain Text*;
- (ix) Realizar novamente o alinhamento;
- (x) Repetir as etapas “vi” a “ix” se é encontrado outro desalinhamento por presença de abreviação.

O tempo de preparação para o alinhamento, o número de etapas e de programas e os documentos gerados são resumidos no Quadro abaixo:

Quadro 44: Tempo, o número de etapas e os documentos gerados em Fleuri (2006) —Preparação para alinhamento.

Tempo	Número de etapas	Número de programas	Documentos gerados
Horas	10	2 <i>WST</i> <i>MS Word</i>	+2 Word <i>Plain Text</i>

O processo de preparação para o alinhamento em Fleuri (2006) realizou mais de 10 etapas, transitando entre dois programas, considerando que houve algumas repetições das etapas (vi) a (ix). A realização dessas etapas levou horas de trabalho contínuo para cada texto e gerou mais de dois documentos.

No estudo piloto esta preparação se deu no *Template Editor*. Cada texto foi processado separadamente, seguindo as etapas expostas no Capítulo 5 (com exceção da etapa ii), que são:

- (i) Colar o texto no *Template*;
- (ii) Formatar o texto se necessário: não foi necessário.
- (iii) Ativar os Macros criados;
- (iv) Colar o texto no *Template*, selecioná-lo todo e se acionar o comando gravado Alt+T;
- (v) Copiar todo o texto e colá-lo no *Template Corpus Paralelo*;
- (vi) Salvar o documento como TF ou TA:

Quadro 45: Tempo, o número de etapas e os documentos gerados no Estudo Piloto—Preparação para alinhamento.

Tempo	Número de etapas	Número de programas	Documentos gerados
Minutos	5	1 <i>MS Word</i>	2 <i>MS Word</i>

O processo de preparação para o alinhamento no Estudo Piloto com o *Template Editor* realizou 5 etapas automáticas, desenvolvidas em apenas um programa. A realização dessas etapas levou alguns minutos para ser completada e gerou apenas um documento.

Comparamos ambos procedimentos, como observado no Quadro abaixo:

Quadro 46: Comparação Fleuri (2006) / Estudo Piloto no *Template Editor* do processo de preparação para alinhamento.

	Fleuri (2006)	Estudo Piloto <i>Template Editor</i>
Tempo	Horas	Minutos
Número de etapas	+10	5
Número de programas	2	1
Documentos gerados	+2	2

Observo que o *Template Editor* agilizou muito o processo de preparação para o alinhamento em relação ao processo adotado em Fleuri (2006). Diminuiu para mais que a metade o número de etapas de preparação e gerou apenas um documento para cada texto (TF e TA), enquanto a dissertação Fleuri (2006) gerou mais que dois documentos.

6.2. EXECUÇÃO DO ALINHAMENTO

O processo de alinhamento do corpus em Fleuri (2006, p. 74-76) foi realizado em todo o texto, e não apenas com os complexos oracionais que seria analisados. As unidades de alinhamento desses complexos oracionais foram isoladas, apagando-se todas as unidades de alinhamento que não seriam analisadas. Esse processo de alinhamento em Fleuri (2006) se deu no *WordSmith Tools* e *MS Word*, sendo frequentemente gravado em *Plain Text*. Tal alinhamento passou pelas seguintes etapas:

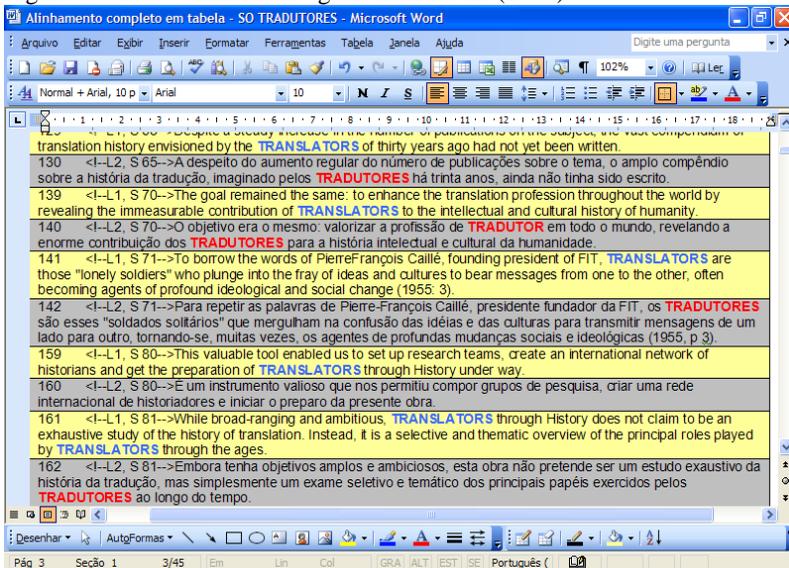
- (i) Corrigir o desalinhamento no próprio *WordSmith Tools* acionando os comandos de “juntar” (*Join*) ou “separar” (*Split*) sentenças (nos casos de desalinhamento causados por tradução de uma frase em duas, duas em uma ou de falta de tradução ou de original);
- (ii) Salvar o alinhamento final em *Plain Text*;
- (iii) Copiar e colar este alinhamento em um Documento *Word*;
- (iv) Converter o texto em tabela;
- (v) Editar e formatar tabela, atribuindo cores diferentes ao TF e ao TA. Essa coloração teve a função de agilizar e garantir o sucesso da etapa de eliminação do par de sentenças que não apresentavam o nóculo;
- (vi) Salvar alinhamento final do texto todo em Documento *Word*;

(vii) Eliminar unidades de alinhamento que não são considerados na análise. É uma etapa que demanda muita atenção, para que não fossem eliminadas unidades analisáveis;

(viii) Salvar documento.

O resultado final é ilustrado na Figura a seguir:

Figura 36: Tela – alinhamento gerado em Fleuri (2006)



O TF e o TA são organizados um abaixo do outro e as unidades de alinhamento são o par de linhas amarelas (TF)/cinza (TA). O nódulo do TF é destacado em azul e o do TA em vermelho. O tempo de preparação para o alinhamento, o número de etapas e de transições entre programas e os documentos gerados são resumidos no Quadro abaixo:

Quadro 47: Tempo, o número de etapas e os documentos gerados em Fleuri (2006)—Alinhamento.

Tempo	Número de etapas	Número de programas	Documentos gerados
Dias	8	2 WST MS Word	+2 Word Plain Text

O processo de alinhamento em Fleuri (2006) se realizou em dois programas em 8 etapas que levaram dias de trabalho cuidadoso, gerando mais de dois documentos.

A realização do processo de alinhamento no estudo piloto, no mesmo corpus, é realizado no *Template Corpus Paralelo*, na pasta “Alinhamento”, seguindo as etapas de:

- (i) Copiar o texto-fonte e o texto-alvo do documento no *Template Editor*;
- (ii) Abrir o *Template Corpus Paralelo* (pasta “Alinhamento”) e colar o TF na célula A2 e o TA na célula B2;
- (iii) Assim que colar o TF, antes de colar o TA, verificar se a coluna B não apresenta texto (conforme explicado no Capítulo 5) e após colar o TA, fazer o mesmo com a coluna C;
- (iv) Formatar as colunas A e B através do comando Ctrl+Q para a coluna A Ctrl+W para a coluna B;
- (v) Destacar, se desejar, as células que contém o nódulo a ser analisado através da ferramenta *Find & Replace*;
- (vi) Organizar o alinhamento, a partir dos seguintes comandos: “buscar” por linhas desalinhadas (Ctrl+M), “juntar” células em caso de desalinhamento (Ctrl+J) e “separar” células desalinhadas (Ctrl+S);
- (vii) Inserir célula manualmente, se necessário;
- (viii) Salvar o alinhamento frequentemente,

O resultado final é ilustrado na Figura a seguir:

Figura 37: Tela Template Corpus Paralelo – alinhamento gerado no estudo piloto

	A	B
1	DIGAS OS PASSOS PARA O ALINHAMENTO:	ORGANIZE O ALINHAMENTO
2	PREFACE	PREFÁCIO
3	"Turning translations into instruments of humanism, peace and progress such is our noble task".	"Transformar as traduções em instrumentos de humanismo, paz e progresso - esta é nossa nobre tarefa".
4	These are the words of Pierre-François Calle (1907-79), founding President of the International Federation of translators; they reflect his personal philosophy, which he passed along to the Federation when it was established in 1953 (Léves, 1979, p. 23).	Fala de Pierre-François Calle (1907-1979), presidente fundador da Federação Internacional de tradutores; essas palavras refletem sua filosofia pessoal, que ele transmitiu à federação quando a instituiu em 1953 (Léves, 1979, p. 23).
5	In article 6 of its bylaws, FIT proclaims its responsibility to "assist in the spreading of culture throughout the world".	No artigo 6 do estatuto que a rege, a FIT proclama sua responsabilidade em "ajudar a difusão da cultura por todo o mundo".
6	The tens of thousands of translators who belong to the seventy-three member organizations of our Federation spare no effort to fulfil this mission.	As dezenas de milhares de tradutores que pertencem às 73 organizações associadas à nossa federação não pougam esforços para cumprir essa missão.
7	The work they perform on a day-to-day basis attests to the fact that translation permeates all facets of human activity and is an inexhaustible source of progress.	O trabalho que realizam diariamente testemunha o fato de que a tradução é uma atividade que permeia todas as facetas da vida humana, e é também uma fonte inesgotável de progresso.
8	People have translated since time immemorial.	As pessoas têm traduzido desde época imemorial.
9	Long before FIT, translators served as vitalinks in the vast chain through which knowledge was transmitted among groups of people separated by language barriers.	Muito antes da FIT, os tradutores serviam como elos vitais na vasta cadeia de transmissão do conhecimento entre sociedades separadas por barreiras linguísticas.
10	Ever since humans first devised writing systems, translators have been building bridges between nations, races, cultures and continents.	Desde que os primeiros homens utilizaram a escrita, os tradutores têm construído pontes entre nações, raças, culturas e continentes.

Como ilustra a Figura acima, o TF e TA são alinhados lado-a-lado facilitando o reconhecimento das unidades de alinhamento. As células em vermelho correspondem às células que apresentam o nódulo analisado. O tempo de preparação para o alinhamento, o número de etapas e de transições entre programas e de documentos gerados são resumidos no Quadro abaixo:

Quadro 48: Tempo, o número de etapas e os documentos gerados no Estudo Piloto—Alinhamento.

Tempo	Número de etapas	Número de programas	Documentos gerados
Horas (6h)	8	1 <i>MS Excel</i>	1 Excel

A realização deste processo no estudo piloto, no mesmo corpus, levou exatamente seis horas para ser concluído. Houve 8 etapas realizadas em um só programa, o *MS Excel*, gerando portanto apenas um documento.

A comparação de ambos procedimentos podem ser observados no Quadro a seguir:

Quadro 49: Comparação Fleuri (2006) / Estudo Piloto no *Template Corpus Paralelo* do processo de alinhamento.

	Fleuri (2006)	Estudo Piloto <i>Template Corpus Paralelo</i>
Tempo	Dias	Horas (6h)
Número de etapas	8	8
Número de programas	2	1
Documentos gerados	+2	1

Observo que a pasta “Alinhamento” no *Template Corpus Paralelo* agilizou muito o processo de alinhamento em relação ao processo adotado em Fleuri (2006), dias de trabalho foram resumidos em 6 horas apenas. O número de etapas para o alinhamento é o mesmo, o que mudou foi que as etapas no *Template* são automáticas enquanto em Fleuri (2006) foram manuais. Como tudo é realizado apenas no *Template Corpus Paralelo*, o Estudo Piloto gerou apenas um documento, enquanto em Fleuri (2006), por ter sido realizado no *MS Word* e no *WST* gerou mais de dois. O alinhamento no Estudo Piloto, em relação ao alinhamento em Fleuri (2006), não passou pela etapa de levar o alinhamento do *WST* ao *Word*, de transformar em tabela e de eliminar linhas que não apresentavam o nódulo.

6.3. ANOTAÇÃO DO CORPUS / CLASSIFICAÇÃO DOS OBJETOS DE ANÁLISE

Em Fleuri (2006) é realizada uma anotação do corpus, pois etiquetas são adicionadas no corpo do texto rotulando cada objeto analisado (cf. Leech, 2004). A classificação dos objetos de análise é realizada no Estudo Piloto, pois o corpus não é anotado diretamente, os objetos de análise, ao invés, são classificados em tabelas. O processo de anotação de corpus em Fleuri (2006) se deu *MS Word*, sendo ao final gravado em *Plain Text* para ser quantificado no *WST*. Foi adotado o CROSF14 (Feitosa, 2005) para anotar os objetos de análise, sendo inserido no documento final de alinhamento em *Word*. Esse processo passou pelas seguintes etapas:

- (i) Estudar o CROSF14, lendo a dissertação de Feitosa (2006);
- (ii) Inserir as primeiras etiquetas (<0010) antes do objeto analisado, identificando onde há objetos de análise (no caso da pesquisa de Fleuri, 2006, eram os Processos e Participantes);
- (iii) Criar uma tabela resumindo o significado dos três últimos algarismos do CROSF, para auxiliar na etapa de classificação dos objetos analisados;
- (iv) Completar essas etiquetas com os últimos três algarismos do CROSF (<0010yyy>) de classificação dos objetos de análise, marcando com pontos de interrogação (<0010???) as etiquetas cuja classificação necessita estudo;
- (v) Separar as unidades de alinhamento em Texto-fonte e em Texto-alvo, gerando dois documentos, um para o TF e outro para o TA;
- (vi) Converter as tabelas em textos;
- (vii) Salvar cada texto em *Plain Text*;
- (viii) Buscar as etiquetas com símbolos de interrogação no *Concord* no *WST*;
- (ix) Estudar e re-etiquetar os rótulos com símbolos de interrogação, nos TF e TA em Documento *Word*;
- (x) Salvar cada texto em *Plain Text* para ser utilizado no processo de quantificação.

O tempo de anotação do corpus, o número de etapas e de programas e os documentos gerados são resumidos no Quadro a seguir:

Quadro 50: Tempo, o número de etapas e os documentos gerados em Fleuri (2006)—Alinhamento.

Tempo	Número de etapas	Número de programas	Documentos gerados
Semanas	10	2 <i>WST</i> <i>MS Word</i>	+3 Word <i>Plain Text</i>

Este processo levou semanas de trabalho contínuo, pois foram mais de mil elementos rotulados. As etapas que mais levam tempo são (i), (ii), (iii) e (viii). Passou-se por 10 etapas para que todo o corpus (TF e TA) fosse anotado, gerando mais de três documentos: o corpus alinhado e anotado, o TF anotado, o TA anotado e os respectivos textos salvos em *Plain Text* para serem processados no *WST*.

O Estudo Piloto desta tese realiza a classificação dos objetos de análise no próprio *Template Corpus Paralelo*, na pasta “Rotulação TF” e na “Rotulação TA”. O processo de classificação é composto pelas seguintes etapas:

- (i) Apagar o texto de exemplo da célula B2 à Q18 e copiar o TF e o TA em sua respectiva pasta, na coluna A;
- (ii) Escrever o nome das categorias nas células B2 a P2 de cada pasta;
- (iii) Colorir as células que apresentam os nódulos analisados;
- (iv) Quebrar, na coluna A, os complexos oracionais em unidades de análise (cada linha deve corresponder a uma categoria);
- (v) Identificar a categoria a qual o elemento da linha pertence e marcar com o número “um” a célula correspondente;
- (vi) Caso uma linha apresente incerteza quanto à classificação, tingir toda a linha com alguma cor extravagante que permita a rápida identificação e o retorno posterior;
- (vii) Estudar e retornar a essas linhas tingidas para classificá-las, apagando o tingimento uma vez que a classificação é concluída.
- (viii) Escrever o objeto analisado na coluna Q;
- (ix) Salvar o documento.

O tempo de classificação dos objetos analisados, o número de etapas e de transições entre programas e os documentos gerados são resumidos no Quadro a seguir:

Quadro 51: Tempo, o número de etapas e os documentos gerados no Estudo Piloto—Classificação dos objetos analisados.

Tempo	Número de etapas	Número de programas	Documentos gerados
Dias	9	0	0

Este processo levou dias de trabalho. As etapas (i) e (ix) são completamente automáticas e as etapas (iv), (v), (vii) e (viii) são as mais lentas, pois requerem estudo e análise do pesquisador. Passou-se por 9 etapas para que todo o corpus (TF e TA) fosse classificado. Como tudo foi realizado na mesma planilha do *MS Excel*, não foi gerado nenhum documento novo.

A comparação de ambos procedimentos podem ser observados no Quadro a seguir:

Quadro 52: Comparação Fleuri (2006) / Estudo Piloto no *Template Corpus Paralelo* do processo de anotação de corpus e classificação dos objetos analisados.

	Fleuri (2006)	Estudo Piloto <i>Template Corpus Paralelo</i>
Tempo	Semanas	Dias
Número de etapas	10	9
Número de programas	2	0
Documentos gerados	+3	0

Observo que a organização da pasta “Rotulação TF” e “Rotulação TA” no *Template Corpus Paralelo* agilizou muito o processo de classificação dos objetos analisados em relação ao processo anotação de corpus adotado em Fleuri (2006), semanas de trabalho foram resumidos em dias apenas. O número de etapas no processo de anotação do corpus em Fleuri (2006) é um pouco maior que de classificação dos objetos analisados no Estudo Piloto. As etapas no *Template Corpus Paralelo* são mais automáticas que em Fleuri (2006). A adoção de um código de rotulação demandou: (i) um estudo do código, (ii) a inserção de um etiqueta de identificação dos objetos analisados, (iii) a elaboração de uma tabela resumindo o significado do CROSF para acessar durante a classificação dos objetos analisados e (iv) a inserção dos 3 últimos algoritmos do CROSF. No Estudo Piloto no *Template Corpus Paralelo*, a classificação dos objetos analisados excluiu as etapas (i) e (iii) e substituiu as etapas (ii) e (iv) por etapas mais automatizadas de quebrar

células (um comando: Ctrl+S) e de identificar a coluna de categorias a qual pertence objeto a ser analisado (inserir o número 1 na célula correspondente). Como tudo é realizado apenas no mesmo *Template Corpus Paralelo*, o Estudo Piloto não gerou um novo documento, enquanto em Fleuri (2006) mais de três documentos foram gerados.

6.4. REVISÃO DAS ETIQUETAS

Em Fleuri (2006) o processo de revisão das etiquetas inseridas no corpus durante o processo de anotação ocorreu no *MS Word*. Para visualizar todos os objetos de análise do TF e do TA organizei-os no *MS Word* em tabelas conforme ilustrado na Figura a seguir:

Figura 38: Tela – tabela para revisão das classificações em Fleuri (2006)

PROCEOS	VERBOS	SENTENÇAS	PROCESSO	VERBO SERVINDO DE PROCESSO	SENTENÇAS	No.
<0010310>	cumprir essa missão	4	<0010310>	fulfil	4	1.
<0010330>	pertencem	4	<0010310>	spare	4	2.
<0010310>	poupan esforços	4	<0010330>	belong to	4	3.
<0010330>	serviam como	7	<0010330>	served as	7	4.
<0010330>	abraner	8	<0010310>	building	8	5.
<0010310>	construido	8	<0010330>	have the ability	8	6.
<0010330>	abarcar	10	<0010330>	to span	8	7.
<0010310>	abrindo	10	<00103??>	to encompass	10	8.
<0010310>	ampliando	10	<0010310>	breach	10	9.
<0010310>	derrubam os obstáculos	10	<0010310>	broadening	10	10.
<0010310>	elimina-las	11	<0010310>	opening up	10	11.
<0010310>	trabalham	11	<0010310>	eliminating	11	12.
<0010310>	Vivem	11	<0010310>	live off	11	13.
<0010320>	desprezados	12	<0010310>	working toward	11	14.

A tabela ilustrada na Figura acima tem uma coluna para as etiquetas (i.e. <0010310>), outra para os objetos analisados (i.e. verbo “cumprir”) e outra para o número das sentenças correspondente. Nessa Figura as informações do TF estão em paralelo com as informações do TA.

O processo de elaboração desta tabela passou pelas seguintes etapas:

- (i) Buscar as ocorrências das etiquetas de um dos textos no *Concord* do *WST*, organizando-as por ordem de aparecimento no texto;
- (ii) Copiar e colar as ocorrências do *Concord* em documento *Word*;
- (iii) Converter o texto em tabela;
- (iv) Apagar o texto que não corresponde à etiqueta e ao objeto de análise. Esse trabalho foi realizado em cada linha da tabela, um a um, manualmente.
- (v) Adicionar duas colunas à tabela;
- (vi) Selecionar o objeto analisado e copiá-lo à seguinte coluna, na mesma linha;
- (vii) Retornar à do *Concord*, aumentar (comando *Grow* do *WST*) as ocorrências, procurar o número da sentença em que se encontrava cada ocorrência;
- (viii) Copiar e colar o número de sentença na tabela do *Word*;
- (ix) Colorir cada etiqueta e seus correspondentes de uma cor diferente, facilitando a visualização e identificação das etiquetas;
- (x) Inserir uma quarta coluna, enumerando as ocorrências;
- (xi) Salvar documento;
- (xii) Repetir os processos (i) a (xi) com o TA;
- (xiii) Abrir e salvar um terceiro documento *Word* com ambas as tabelas (do TF e do TA) posicionadas lado-a-lado (conforme a última Figura).

O tempo de elaboração dessas tabelas, o número de etapas e de transições entre programas e os documentos gerados são resumidos no Quadro a seguir:

Quadro 53: Tempo, o número de etapas e os documentos gerados em Fleuri (2006)—Revisão de Etiquetas.

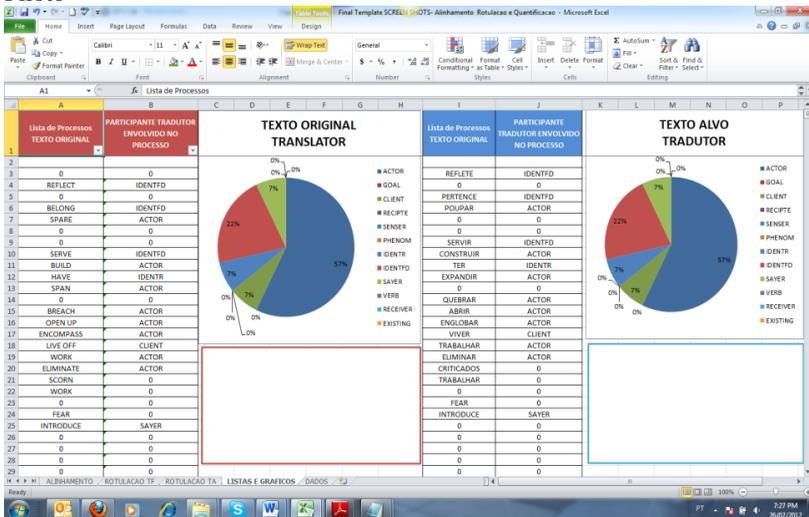
Tempo	Número de etapas	Número de programas	Documentos gerados
Dias	25	2 <i>WST</i> <i>MS Word</i>	3 Word

Em Fleuri (2006) este processo levou dias de trabalho manual de formatação e de passar os fragmentos de textos à tabela. Passou-se por 25 etapas, considerando que as etapas de (i) a (xii) são realizadas duas

vezes, uma para o TF e outra para o TA. Essas etapas demandam transições entre o programa *MS Word* e o *WST*. Foram gerados três documentos no *MS Word*: a tabela do TF, a tabela do TA e a tabela com ambos os textos.

O Estudo Piloto organiza automaticamente os dados em tabelas/listas, na pasta “Listas e Gráficos” ao mesmo tempo que quantifica os dados em gráficos e na linha 104625 das pastas de “Rotulação”. Como os objetos analisados são escritos no momento de sua classificação, nas pastas de “Rotulação TF” e “Rotulação TA” e a nomenclatura da classificação é passada à lista automaticamente através de um fórmula, essa etapa não representa trabalho extra ao pesquisador, nem tempo extra (já que o tempo é considerado na fase de classificação dos objetos analisados). A geração completamente automática da tabela (e dos gráficos de quantificação), na pasta “Lista e Gráficos”, é ilustrada na Figura a seguir:

Figura 39: Tela Template Corpus Paralelo - tabelas e gráficos gerados no Estudo Piloto



A lista/tabela em vermelho corresponde aos dados do TF e em azul ao TA (conferir Capítulo 5 desta tese). O tempo de elaboração dessas tabelas, o número de etapas e de transições entre programas e os documentos gerados são resumidos no Quadro a seguir:

Quadro 54: Tempo, o número de etapas e os documentos gerados no Estudo Piloto— Revisão de Etiquetas.

Tempo	Número de etapas	Número de programas	Documentos gerados
0	0	0	0

O sistema de fórmulas do *MS Excel* permite que o processo de quantificação no *Template Corpus Paralelo* seja completamente automático. O único tempo que pode ser gasto neste processo é na aplicação de um filtro nas tabelas, escolhendo esconder as linhas que não tiveram elementos analisados; esse processo não leva mais do que dois minutos, usando os recursos do próprio *Excel*.

A comparação de ambos procedimentos podem ser observados no Quadro a seguir:

Quadro 55: Comparação Fleuri (2006) / Estudo Piloto no *Template Corpus Paralelo* do processo de revisão da anotação do corpus e da classificação dos objetos analisados.

	Fleuri (2006)	Estudo Piloto <i>Template Corpus Paralelo</i>
Tempo	Dias	0
Número de etapas	25	0
Número de programas	2	0
Documentos gerados	3	0

Observo que a organização de uma tabela para a revisão das etiquetas em Fleuri (2006) foi uma etapa que tomou muito tempo de pesquisa, demandou a transição por dois programas no decorrer da execução de 25 etapas, gerando 3 documentos. Enquanto o *Template Corpus Paralelo* não gerou nenhum documento extra e demonstrou-se capaz de reduzir o tempo e o número de transições entre programas e de etapas a zero, automatizando completamente o processo de revisão.

6.5. QUANTIFICAÇÃO DOS DADOS E O MODO DE EXPOSIÇÃO

Em Fleuri (2006) o processo de quantificação dos dados se deu *WST*. O valor de cada etiqueta foi copiado do *Concord* do *WST* para uma tabela e um gráfico no *MS Word* para ser exposto na dissertação. Esse processo passou pelas seguintes etapas:

- (i) Processar no *Concord* (do *WST*), um texto de de cada vez (o TF e o TA anotados e gravados em *Plain Text*), buscando por cada uma das etiquetas. Essa é uma etapa automática em que se busca pelos rótulos e se verifica o número de ocorrências de cada etiqueta;
- (ii) Elaborar em Documento *Word* uma tabela para se anotar a quantidade de cada etiqueta, gerando uma tabela para cada texto;
- (iii) Elaborar de um gráfico com esses dados, um para cada texto;
- (iv) Edição dos gráficos e tabelas para sua apresentação na pesquisa;
- (v) Salvar o documento.

O *Template Corpus Paralelo* foi capaz de gerar automaticamente um gráfico com a quantidade dos objetos analisados, exposta em porcentagem no gráfico e em valores numéricos na linha 104625 das pastas de “Rotulação”.

A comparação do tempo de quantificação dos dados, o número de etapas e de transições entre programas e os documentos gerados são resumidos no Quadro a seguir:

Quadro 56: Comparação Fleuri (2006) / Estudo Piloto no *Template Corpus Paralelo* do processo de quantificação dos dados e de organização de sua exposição.

	Fleuri (2006)	Estudo Piloto <i>Template Corpus Paralelo</i>
Tempo	Horas	0
Número de etapas	5	0
Número de programas	2	0
Documentos gerados	1	0

Observo que a quantificação dos dados e a sua organização em gráficos da exposição em Fleuri (2005) levou horas de elaboração, demandando a transição por dois programas no decorrer da execução de 5 etapas, gerando 1 documento no *MS Word*. O *Template Corpus Paralelo* demonstrou-se capaz de reduzir o tempo e o número de transições entre programas e etapas a zero, automatizando completamente o processo de revisão, não gerando nenhum novo documento.

6.6. REFLEXÕES SOBRE O ESTUDO PILOTO

O Estudo Piloto, em comparação ao método utilizado em Fleuri (2006), demonstrou que o trabalho combinado com os *Templates*:

- (i) Agiliza o processo de preparação para o alinhamento, de alinhamento do corpus, de anotação do corpus, de revisão das etiquetas e de quantificação dos dados e de organização da exposição desta quantificação;
- (ii) Reduz o número de etapas em cada um desses processos;
- (iii) Reduz o número de transições entre programas utilizados em cada processo;
- (iv) Reduz o número de documentos gerados em cada processo.

Os dois Quadros a seguir permitem visualizar a estimativa da agilização dos processos e o total da redução de etapas, transições entre programas e documentos de cada processo. O Gráfico seguinte permite comparar esses elementos entre ambas as pesquisas:

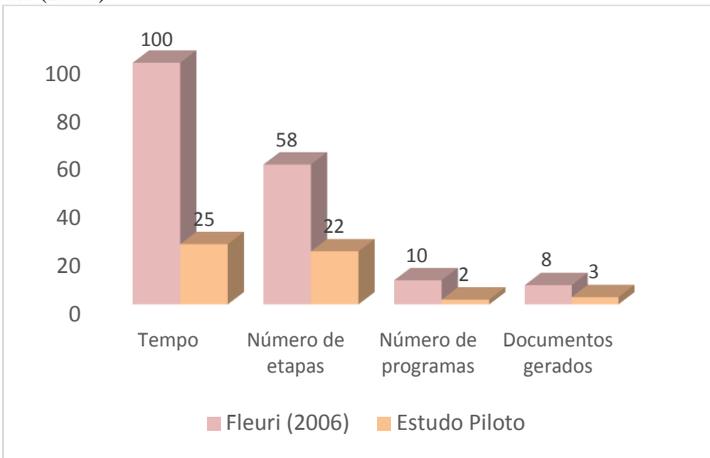
Quadro 57: Total das etapas nos métodos de compilação em Fleuri (2006).

Fleuri (2006)	Tempo	Número de etapas	Número de programas	Documentos gerados
Preparação Alinhamento	Horas	+10	2	+2
Alinhamento	Dias	8	2	+2
Anotação do Corpus	Semanas	10	2	
Revisão das etiquetas	Dias	25	2	3
Quantificação dos dados.	Horas	5	2	1
Total	Mais de 1 mês	+58	10	+8

Quadro 58: Total das etapas nos métodos de compilação no Estudo Piloto.

Estudo Piloto	Tempo	Número de etapas	Número de programas	Documentos gerados
Preparação Alinhamento	Minutos	5	1	2
Alinhamento	Horas	8	1	1
Anotação do Corpus	Dias	9	0	0
Revisão das etiquetas	0	0	0	0
Quantificação dos dados.	0	0	0	0
Total	Mais de 1 semana	22	2	3

Gráfico 12: Comparação do total das etapas metodológicas entre o Estudo Piloto e Fleuri (2006).



Comparando os métodos utilizados em Fleuri (2006) com os testes dos Templates realizado neste Estudo Piloto, observo que no Estudo Piloto o tempo gasto se reduz a aproximadamente $\frac{1}{4}$, o número de etapas executadas a menos da metade, o número de trânsito entre programas a $\frac{1}{5}$ e de documentos gerados a menos da metade.

Isso significa que o trabalho combinado com ambos *Templates* representa uma ágil e organizada alternativa de se compilar corpora

paralelos bilíngues de pequena dimensão. Além disso, por manter aberta sua forma de classificação, oferece um grau de flexibilidade para que o pesquisador escolha a categoria a adotar e sua nomenclatura, sem depender de um código já desenvolvido ou de criar outro.

Outra vantagem destes *Templates* é que eles ainda podem ser adaptados e redesenhados para cumprir com os objetivos da pesquisa. Partir de algo já elaborado é mais fácil do que criar outro do zero. Os resultados dessas adaptações podem gerar um grupo de *Templates*, oferecendo mais opções que supram as diversas necessidades de pesquisas em ESFTBC ou em ETBC.

7 CONSIDERAÇÕES FINAIS

"O que a lagarta chama de fim de mundo, o mestre chama de borboleta."
(Richard Bach, 1977, p.134)

Neste capítulo retomo as reflexões dos capítulos de introdução e referencial teórico. Com base nas conclusões das análises, da proposta metodológica e do estudo piloto, respondo às perguntas de pesquisa e ao final, aponto propostas para futuras pesquisas.

Como explicitado no Capítulo 1, a presente investigação se vincula institucionalmente ao projeto CORDIALL e ao Grupo de Pesquisa TraCor que incluem pesquisas inseridas no contexto das abordagens discursivas e textuais dos Estudos da Tradução, em interface com os *Estudos Sistêmico-Funcionais da Tradução* e os *Estudos da Tradução Baseados em Corpus*. Desde 2003, no Brasil, essa interface de pesquisa vem se tornando uma tradição. Para apontar a interface simultânea chamamo-na nesta tese de *Estudos Sistêmico-Funcionais da Tradução Baseados em Corpus*, evitando localizar estas pesquisas em ‘ESFT com metodologia da LC’ ou ‘ETBC com arcabouço teórico da LSF’. Assis (2012) e Alves e Vasconcellos (no prelo) tratam de mapear respectivamente pesquisas em ESFT e ETBC, mas acabam por produzir uma sobreposição de trabalhos nos mapeamentos propostos, principalmente no que diz respeito àqueles que se contextualizam no que aqui denominamos de ESFTBC.

As pesquisas em ESFTBC são recentes. Até 2010, identificamos 24 publicações de pesquisas acadêmicas na área dos ESFTBC, entre dissertações de mestrado (20) e teses de doutorado (4). A Linguística de Corpus, iniciada na década de 1960, passa a se desenvolver e se consolidar mais na década de 1980 e 1990, com uma proposta de métodos para realizar pesquisa em textos autênticos. A concepção de Estudos Descritivos da Tradução associada à concepção de análise de evidências textuais em textos autênticos (corpus) faz surgir os Estudos da Tradução Baseados em Corpus (Baker, 1995). Os ETBC oferece aos ET uma perspectiva descritiva de estudos em tradução facilitada pelo método proposto pela Linguística de Corpus de lidar com texto em formato eletrônico e processado por *softwares* de análise lexicais. Esta perspectiva associada à forma de interpretar a linguagem textual como um sistema de significados, oferecida pela Linguística Sistêmico-Funcional, possibilitou

desenvolver estudos que analisassem traduções em contextos, através de um aspecto quantitativo e qualitativo, permitindo a associação da análise micro e macro textual. O corpus utilizado pelos ESFTBC no Brasil é em geral de pequena dimensão.

Conforme mencionado no Capítulo 2, esta tese, por tratar da Metodologia de corpus voltada à pesquisas em traduções, localiza-se no contexto teórico-metodológicos dos ETBC, conforme Baker 1993/1995, Olohan, 2004, Bowker, 2001; Alves e Vasconcellos, no prelo; Fernandes, 2006. O objeto de estudo, ou seja, as dissertações analisadas, encontra-se na interface dos ETBC com os ESFT (cf. Vasconcellos, 2009; Assis, 2012; Pagano e Vasconcellos, 2005). A análise, a proposta e o estudo piloto localizam-se nos estudos de corpus voltados para os Estudos da Tradução, que por utilizar a metodologia de corpus, é frequentemente discutido pelos ETBC e pela Linguística de Corpus (cf. Kennedy, 1998; Barnbrook, 1996; Anderman, G. e Rogers, 2008, Mason, 2008; Garretson, 2008; Scott, 2008).

De acordo com Kennedy (1998), o método da LC é associado aos processos de compilar e elaborar corpus; descrever elementos linguísticos/textuais através de corpora computadorizados e desenvolver ferramentas para analisar corpora. Essa última atividade, em que se enquadra a presente pesquisa, é descrita por Mason (2008), Scott (2008) e Garretson (2008). A linguística de corpus, assim como qualquer outra área em estudos de corpus, por ser um campo essencialmente metodológico (Olohan, 2004; Kennedy, 1998), permite que o enfoque do estudo se direcione ao método (como realizei nesta tese), além da usual aplicação da metodologia na descrição e análise textual em si (como ocorre nas vinte dissertações analisadas nesta tese). Estudos sobre métodos e ferramentas de análise linguística são ainda escassos na área da LC e dos ETBC, conforme apontou o encontro de linguistas de corpus em Georgetown University em 1993 (Kennedy, 1998). Tal escassez faz com que a programação de *software* se torne um estudo mais matemático (Mason, 2008) e voltado ou associado à linguística computacional (Kennedy, 1998). Entretanto, Mason (2008), Scott (2008) e Garretson (2008) mostram que examinar o funcionamento de ferramentas computacionais e de suas aplicações aos métodos de corpus não tem que ser tarefa exclusiva do cientista computacional. De fato, Mason (2008) e esta tese mostram que não é necessário que o pesquisador seja um especialista em linguagem de programação e em desenvolvimento de ferramentas para análise textual para realizar estudos nessa área: todo linguista de corpus pode perfeitamente desenvolver esta tarefa, como prova Scott (2008), que assumiu a tarefa de programação inicialmente

como um *hobby*. Meyer (apud Scott, 2008) prevê que no futuro todos os linguistas de corpus serão um pouco programadores e seus *scripts* serão passados adiante de modo que um dia possam apoiar suas pesquisas em programas escritos e adaptados pelos próprios pesquisadores.

Segundo Garretson (2008) e Barnbrook (1996), o projeto de pesquisa em corpus deve considerar a ferramenta e a possibilidade de programar. Segundo Barnbrook (1996), toda pesquisa em corpus requer um projeto e, se possível, um estudo piloto para avaliar se o orçamento, os recursos e o tempo disponíveis são suficientes para se adotar uma metodologia de corpus. Esse projeto define: o desenho do corpus⁸⁵; o modo de compilação deste (Olohan, 2004); a necessidade de anotação e, em caso afirmativo; o código a ser adotado ou criado (cf. Leech, 2005 e Feitosa, 2005); o modo de processamento do corpus (Anderman & Rogers, 2008) e a definição do tipo de programa⁸⁶, e por fim, o modo de expor os dados quantitativos na redação final da pesquisa. A escolha entre um programa pronto ou programa adaptável, como o sugerido neste tese, deve resultar dos objetivos da pesquisa e ser incluída no planejamento que avalia a viabilidade de utilização de um método como o sugerido aqui.

Tendo em vista a escolha do tipo de programa, deve-se considerar que tanto os programas prontos, como o *WordSmith Tools* e o *AntConc*, quanto os programas adaptáveis, como o caso do *MS Word* e do *MS Excel*, apresentam vantagens e desvantagens. Os Aplicativos de Planilha, como o *MS Excel*, são em geral utilizados para desempenhar suas funções básicas de organização de textos e de dados em tabelas e gráficos. O trabalho realizado nestes programas é essencialmente manual e limitado, sem exploração de seus recursos de gravação de *macros*. Entretanto, ao explorar-se a linguagem de programação em *Macros* e Fórmulas, o *MS Excel* demonstra-se muito útil para trabalhar em corpora descartáveis multilíngues de pequena dimensão, pois pode armazenar uma quantidade controlada de texto e ser programado para cumprir propósitos específicos de uma determinada pesquisa. As pesquisas que utilizam o *WST* relatam a praticidade de realizar concordâncias automáticas e de obterem dados estatísticos do corpus em segundos. Entretanto, relatam dificuldades em alinhar textos bilíngues de modo completamente automático. Ambos os programas, para serem explorados em sua totalidade demandam um tempo de aprendizagem. O *WST* e o *AntConc* demandam um estudo e uma

⁸⁵ (cf. Kennedy, 1998; Olohan, 2004; Fernandes, 2006, Borja, 2008 e Varantola, 2002)

⁸⁶ (cf. Bowker, 2004; Garretson, 2008, Mason, 2008 e Sardinha, 2004)

prática de suas funções, enquanto o *MS Excel* demanda estudo de programação de *Macros* e de fórmulas.

O capítulo de metodologia expôs os procedimentos seguidos durante a análise de pesquisas em ESFTBC, o desenvolvimento de uma proposta metodológica e a execução de um Estudo Piloto. Esses procedimentos, desdobrados em várias etapas, foram essencialmente quatro: (i) coleta de pesquisas em ESFTBC; (ii) descrição e análise dessas dissertações; (iii) planejamento e desenvolvimento de uma proposta metodológica de compilação de corpus paralelo bilíngue de pequena dimensão a partir de um Processador de Textos e de um Aplicativo de Planilha e (iv) execução de um Estudo Piloto, testando a proposta.

Os Capítulos 4, 5 e 6 respondem às perguntas de pesquisas levantadas no Capítulo 1 desta tese. Abaixo essas são respondidas:

PP1: Quantas e quais são as dissertações e teses publicadas em ESFTBC no Brasil até 2010?

Através do procedimento de coletas de pesquisas em ESFTBC, que se deu a partir da leitura de dois artigos, Assis (2012) e Alves e Vasconcellos (no prelo) e de uma investigação bibliométrica online em sites acadêmicos, de bibliotecas universitárias e de *curricula vitae* de pesquisadores em ESFTBC, chegou-se a cinco listas de pesquisas que envolvem os campos dos ETBC, ESFT e o cruzamento deste dois campos, chamado de ESFTBC:

Quadro 59: as cinco listas desta pesquisa: duas consultadas, a Lista Assis e a Lista Alves e Vasconcellos, e três geradas, a Lista AAV, a AAVF e a ESFTBC.

Listas geradas	No. trabalhos acadêmicos	Estudos envolvidos
<i>Lista Assis</i>	25	ESFT (com ou sem LC)
<i>Lista Alves e Vasconcellos</i>	47	ETBC (com ou sem LSF)
<i>Lista AAV</i>	30	ESFTBC e ESFT
<i>Lista AAVF</i>	31	ESFTBC e ESFT
<i>Lista ESFTBC</i>	24	ESFTBC

A *Lista Assis*, gerada por Assis (2012), inclui 25 trabalhos acadêmicos em ESFT (com ou sem arcabouço metodológico de corpus). A *Lista Alves e Vasconcellos*, gerada por Alves e Vasconcellos (no prelo), resgata 47 dissertações e teses em ETBC (com ou sem o arcabouço teórico da LSF). O cruzamento entre essas duas listas gerou a *Lista AAV*, que ao

excluir as pesquisas repetidas apontou para um número de 30 investigações em ESFT e ESFTBC. A partir do exame cuidadoso do curriculum de todas as orientadoras destacadas nessas 30 pesquisas, chegou-se à *Lista AAVF* (Assis-Alves e Vasconcellos-Fleuri), que identificou uma pesquisa a mais, não presente na *Lista AAV* ou *Lista Assis*, gerando um número de 31 pesquisas em ESFT e ESFTBC. Dessa lista se separaram as pesquisas em ESFT e as em ESFTBC, chegando-se a *Lista ESFTBC*, contendo 20 dissertações de mestrado e 4 teses de doutorados.

PP2: Qual é o perfil metodológico das dissertações em ESFTBC no que se refere ao perfil de pesquisa e de compilação de corpus, quanto: (i) às Metafunções analisadas, (ii) aos objetos de análise; (iii) à versão original dos textos que compõem o corpus e ao tipo de corpus; (iv) ao método de alinhamento dos corpora paralelos; (v) ao código de anotação e marcação; (vi) aos programas utilizados; (vii) às ferramentas do WST e (viii) ao método de quantificação dos dados?

A presente tese analisou vinte dissertações em ESFTBC publicadas entre 2003 e 2010 no Brasil e descreveu os procedimentos metodológicos aplicados em cada uma delas, identificando seu perfil e necessidades emergentes em relação ao método utilizado. Foram gerados gráficos expondo as características dessas pesquisas em percentuais, relacionados à metafunção explorada, ao objeto de estudo, ao tipo de corpus estudado, à versão dos textos compilados, ao tipo de processo de alinhamento, ao código de rotulação, aos *software* acessados, às ferramentas do WST, à quantidade de ferramentas utilizadas e ao modo de processar a quantificação dos dados. Observou-se que:

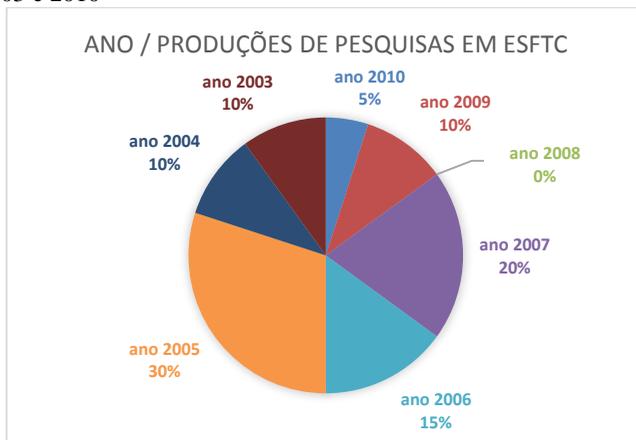
- (i) A maioria estuda a Metafunção Ideacional (65%) e
- (ii) Tem como objeto de análise os Participantes de Processos (29%);
- (iii) Muitos trabalham (nem sempre exclusivamente) sobre um corpus paralelo bilíngue de pequena dimensão (75%) e têm que lidar com a compilação de textos impressos (60%);
- (iv) A grande maioria (85%) opta por utilizar o WST (não exclusivamente);
- (v) Dos trabalhos que alinham o corpus a maioria acessa o utilitário *Viewer and Aligner* do WST (56%);
- (vi) Empatou a proporção majoritária que opta por rotular o corpus com o CROSF e com um código criado (24%);

- (vii) A grande maioria das dissertações acessam o *WST* (85%); desses, grande parte utiliza mais a ferramenta de *Concord* (37%) e optam por utilizar três ferramentas deste programa simultaneamente (40%);
- (viii) Dos trabalhos que realizam análise quantitativa, a maioria opta por contabilizar os dados no *Concord* (40%).

PP3: Como se deu o desenvolvimento diacrônico das pesquisas em ESFTBC no Brasil?

O Gráfico e o Quadro a seguir apresentam a quantidade de produções de pesquisas em ESFTBC afiliadas à UFSC e UFMG, ao longo de sete anos (2003 a 2010):

Gráfico 13: Relação ano/produção de 20 dissertações em ESFTBC no Brasil entre 2003 e 2010



Quadro 60: Relação ano/produção de 20 dissertações em ESFTBC no Brasil entre 2003 e 2010

	ano 2010	ano 2009	ano 2008	ano 2007	ano 2006	ano 2005	ano 2004	ano 2003
2003, Osilene Cruz (UFMG);								1
2003, Cristina Mauri (UFMG);								1
2004, Silvana de Jesus							1	
2004, Roberto Assis (UFMG);							1	
2005, Letícia Bueno (UFMG);						1		
2005, Roberta Rodrigues						1		
2005, Eliza Morinaka (UFSC);						1		
2005, Tassiani Cançado						1		
2005, Marcos Feitosa (UFMG);						1		
2005, Viviane Paquilin (UFSC);						1		
2006, Gleimara Zuniga					1			
2006, Lillian Fleuri (UFSC); O					1			
2006, Ariana Zanella (UFSC);					1			
2007, Daniel Alves (UFMG);				1				
2007, Giacomo Figueiredo				1				
2007, Cristiano Araújo				1				
2007, Joanna Filgueiras				1				
2009, Alinne Fernandes		1						
2009, Thiago Pires (UFSC);		1						
2010, Leonardo Nunes	1							
total	1	2	0	4	3	6	2	2

Observa-se que o ano de maior produção foi 2005 com um total de 6 dissertações defendidas, representando 30% das dissertações em ESFTBC, seguido do ano de 2007, com 4 dissertações representando 20% e o ano de 2006, com 3 dissertações representando 15%. Somente em 2005, 2006 e 2007 produziu-se 65% das dissertações em ESFTBC entre 2003 e 2010. Esses anos coincidem com a atividade do PROCAD, que incentivou o vínculo entre UFSC e UFMG no projeto de pesquisa em ESFTBC, iniciado pelo CORDIAL. Ao longo desses 7 anos de publicação de pesquisas em ESFTBC, no que tange as suas metodologias, observo que foi criado e testado um código de rotulação sistêmico funcional para anotação de corpus em pesquisas em ESFTBC (Feitosa, 2005), sendo utilizado nas pesquisas de Paquilin (2005); Morinaka (2005); Fleuri (2006); Zuniga (2006); Fernandes (2009); Pires (2009). Foram ainda aplicadas diferentes formas de alinhamento e foram utilizados modos variados de quantificar os dados. No entanto, não se observam formas diferentes de compilar o corpus, de modo a se gastar menos tempo na compilação do corpus, como: os processos de escaneamento, a revisão e correção dos erros de leitura do escâner, de preparação do corpus para um alinhamento mais automático, de anotação do corpus, de quantificação dados e de revisão desta rotulação.

PP4: Como pode ser desenvolvida uma alternativa econômica, acessível e dinâmica de compilação de corpus que responda às necessidades apontadas pela maioria das dissertações em ESFTBC?

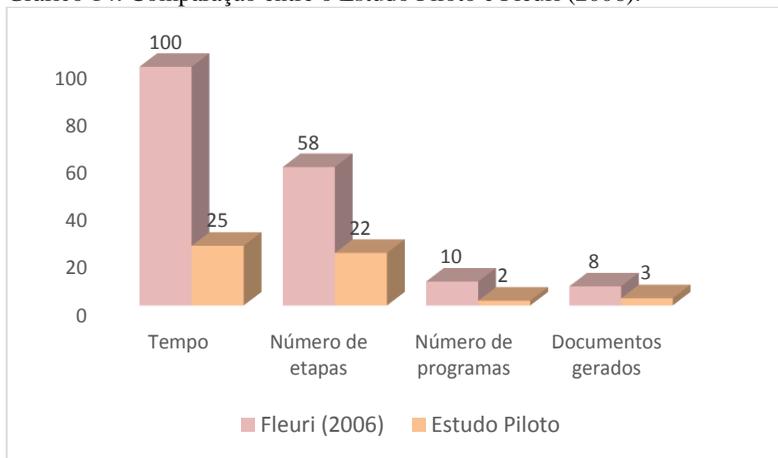
Ao analisar tais dissertações, observou-se que muitos pesquisadores apresentaram dificuldades durante a compilação do corpus. Esse fato corroborou o que eu já havia constatado durante minha pesquisa de mestrado, apontando à latente necessidade de se obter um possível método que dinamize o processo de compilação de corpus paralelo bilíngues de pequena dimensão. O fator tempo é, por exemplo, mencionado em diversas pesquisas em ESFTBC como um obstáculo ao aprofundamento da análise textual, levando muitas dissertações a se manterem essencialmente no campo da descrição linguística, ou micro textual e deixar o aprofundamento da análise macro textual (cf. Tymoczko, 1998), para futuras pesquisas. Ou então, observa-se que tal fator obriga o pesquisador a enxugar seu corpus ou o raio de investigação para que esta possa ser realizada dentro do tempo previsto, cumprindo com a proposta de pesquisa (i.e. Fleuri, 2006).

A partir da identificação dessa necessidade, propõe-se um método de compilação e processamento de corpus, passível de ser adotado e adaptado por outras pesquisas tanto em ESFTBC como em ETBC. Tal método não apenas visa a oferecer uma alternativa de se trabalhar eletronicamente com corpus multilíngues de pequena dimensão, como também tornar acessível economicamente a manipulação automática de corpus e de dados. Para tanto, crio *Templates* em um Processador de Texto e um Aplicativo de Planilha, da *MS Word* e *MS Excel*, que desempenham a função de (i) corrigir e formatar os textos para serem alinhados (realizado no *Template Editor*, no *MS Word*), (ii) alinhar e corrigir pares desalinhados semiautomaticamente em corpus paralelo (no *Template Corpus Paralelo*, no *MS Excel*), (iii) organizar a classificação dos objetos de análise, fornecendo (iv) a quantificação dos dados automática e simultaneamente e (v) sugerir a organização de uma tabela que auxilia no processo de revisão dos elementos rotulados, arranjando automática e paralelamente os termos e os rótulos correspondentes.

Tal método foi testado em um Estudo Piloto formado com o mesmo corpus da dissertação de Fleuri (2006). Foram comparados, entre Fleuri (2006) e o Estudo Piloto, o tempo gasto, o número de etapas executadas, o número de transições entre programas e o número de documentos gerados em cada processo de compilação: de preparação dos textos para alinhamento, de alinhamento do corpus, de anotação do

corpus, de correção das etiquetas e de quantificação dos dados. Observou-se que em Fleuri (2006) cada processo demandou mais tempo e mais etapas que no Estudo Piloto. A comparação entre a compilação em Fleuri (2006) e no Estudo Piloto é ilustrada no Gráfico a seguir:

Gráfico 14: Comparação entre o Estudo Piloto e Fleuri (2006).



Em Fleuri (2006) o processo de compilação de corpus demandou mais de um mês de trabalho contínuo para sua conclusão, tendo sido executadas mais de 58 etapas, transitando 10 vezes entre programas diferentes (*MS Word*, *WordSmith Tools*, *Notepad*, *MS Excel*) e gerando mais de oito documentos diferentes. O Estudo Piloto levou menos de duas semanas, executando um total de 22 etapas, duas transições entre programas e gerando 3 documentos (dois no *Processador de Texto* e um no *Aplicativo de Planilha*). Isso representou uma significativa economia de recursos e de tempo, resultando em maior eficiência do trabalho de pesquisa.

O Aplicativo de Planilha que antes era utilizado essencialmente para desempenhar funções básicas de organização de textos e de dados em tabelas e gráficos, demonstrou-se também útil no trabalho com corpora paralelos bilíngues de pequena dimensão, para alinhar textos e quantificar dados. As pesquisas que utilizaram o *WST* relataram a praticidade de se realizar concordâncias automáticas e de obterem dados estatísticos do corpus em segundos. Entretanto, também relataram dificuldades em alinhar textos bilíngues de modo completamente automático e a necessidade de se passar o alinhamento gerado no *WST* a outro documento (em geral a um *Processador de Texto*). Ambos os tipos

de programas demandam um tempo de aprendizagem. O *WST* e o *AntConc* requerem um estudo e uma prática de suas funções, enquanto o *Aplicativo de Planilha* demanda estudo de programação de *Macros* e de fórmulas para a elaboração de um novo *Template* ou a adaptação deste oferecido. A utilização dos *Templates* não significa, entretanto, abandonar completamente o uso do *WST* ou de outros programas (como o *AntConc*). Tais *Templates* visam a criar alternativas possíveis de pesquisa com corpus de pequena dimensão. Pois, tanto o *WordSmith Tools* como o *AntConc* oferecem as ferramentas de *WordList* e de *Keyword* que não são substituíveis em nenhum dos *Templates*. Também não significa abandonar a rotulação do corpus com o CROSF, pois a proposta é outra. Enquanto a forma de rotulação proposta nesta tese visa a flexibilizar e dinamizar o processo de classificação do pesquisador em sua própria pesquisa, o CROSF propõe uma sistematização de codificação, abrindo para a possibilidade de que sejam adicionados códigos provindos de pesquisas diferentes em um só texto, e estes textos podendo formar um grande corpus etiquetado com categorias da LSF.

Quanto à forma de classificar os elementos analisados, o *Template Corpus Paralelo* se mostrou uma alternativa flexível, visualmente organizada, facilitando o processo de revisão dos elementos classificados e automatizando o processo de quantificação. Adotar um código de rotulação pode ser uma alternativa de se trabalhar com Metafunções, visto que já há um código criado para anotação desse aspecto da Gramática Sistemico-Funcional (GSF), como o CROSF. Entretanto, em pesquisas que lidam com outros aspectos da GSF, o CROSF já se torna uma opção não praticável, como ocorre com Filgueiras (2007), que trabalha com prosódia semântica ou Alves (2007) que trabalha com aspectos do discurso (Alves, 2007). Além do mais, apesar de as tabelas do protótipo CROSF-14 serem autoexplicativas alguns detalhes acabam passando e a anotação, que deveria ser única, passa a ser realizada de modos diversos em pesquisas diferentes (i.e. Zuniga, 2006, Paquilin, 2005) ou então adaptados para abarcar certos detalhes condizentes aos objetivos das pesquisas (i.e. Morinaka, 2005, Fleuri, 2006).

O estudo piloto destes *Templates* realizado no corpus da dissertação de Fleuri (2006) demonstrou que os mesmos agilizam o processo de preparação dos textos para o alinhamento, de alinhamento do corpus, de anotação do corpus, de quantificação dos dados e de exposição do material analisado, organizando semiautomaticamente as tabelas e gráficos existentes. Os *Templates* representam, assim, uma alternativa para se trabalhar com corpora paralelos bilíngues de pequena dimensão.

Além disso, por manterem flexível a forma de classificação, permitem que o pesquisador escolha o nome da categoria dos elementos que analisa, sem que tal escolha comprometa o processamento de quantificação dos dados.

Os *Templates* criados nos programas adaptáveis não se propõem a substituir os programas prontos, pois estes são muito mais requintados e apresentam ferramentas difíceis ou impossíveis de serem reproduzidas naqueles programas. Esses *Templates* se propõem a oferecer uma nova perspectiva de trabalho com corpora descartáveis e de pequena dimensão, de modo a evitar a conversão de formatos de textos e a transição de um programa a outro. Além disso, os programas adaptáveis oferecem uma alternativa econômica de trabalho em um ambiente computacional familiar a todos os usuários de *Windows*, *Macintosh* e *Linux*.

A proposta do *Template Corpus Paralelo*, assim como do *Template Editor*⁸⁷ foi de oferecer um modelo básico de trabalho com textos em paralelo que pudesse ser adaptado aos objetivos de cada pesquisa. A intenção da criação destes *Templates* não é oferecer algo fixo e rígido, que demanda um enquadramento da pesquisa a este *Template*, mas o contrário. Uma vez compreendida a forma de elaboração dos *Templates*, suas fórmulas e *macros* podem ser modificados, servindo aos novos propósitos das futuras pesquisas em ESFTBC. Assim, o Capítulo 3 e 5 desta tese cumprem a função de explicar o modo de elaboração e de funcionamento dos *Templates* para que possam ser não só reproduzidos, reestudados e adaptados ou recriados em futuras pesquisas. Enfim, um estudo mais aprofundado dos recursos do *Excel* ou a obtenção de respostas em fóruns online sobre *Excel* podem permitir a criação de novas formas de manuseio de corpus, de tal forma a atender às necessidades individuais das subseqüentes pesquisas.

Desta forma, a presente pesquisa cumpriu com seus objetivos de (i) listar todas as dissertações e teses em ESFTBC publicadas no Brasil até 2010, a partir de uma pesquisa bibliométrica; (ii) descrever as características metodológicas de todas as dissertações em ESFTBC no Brasil até 2010; (iii) identificar as necessidades mais emergentes de cada dissertação em relação à compilação do corpus e (iv) criar, propor e testar um método que responda à maiorias das necessidades encontradas, realizado nos Capítulos 5 e 6. O processo de criação e a descrição dos

⁸⁷ Disponíveis para download gratuito no site:
<https://sites.google.com/site/lilianjfleuri/>

Macros e fórmulas explicados no Capítulo 3 e detalhados nos anexos 9 e 10.

Acredito que, como sugestão para futuras pesquisas, resta ainda realizar um aperfeiçoamento e redesenho dos *Templates*. Para isso, recomendo que testes nesses *Templates* sejam realizados por pesquisadores em ETBC e suas respostas aos comandos programados e ao *layout* sugerido sejam observadas e consideradas no aprimoramento. São infinitas as possibilidades de criação de manipulação com textos e números em programas como o do *MS Office*. Este trabalho também pode ser explorado em Sistemas Operativos Livres como o *Linux*. Imagino que as funções, como as de *Concordância* e de *WordList* presentes no *WST* e no *AntiConc* possam ter sua criação explorada em Aplicativos de Planilhas, através de elaborações de fórmulas e de *Macros*. Acredito que o desenho de cada planilha também possa ser repensado e reformulado de modo a ser ainda mais automatizado e autoexplicativo.

No que tange ao estudo bibliométrico de coleta de pesquisas em ESFTBC, não posso negar que inúmeras vezes cogitei em expandir o mapeamento destas investigações para o nível internacional. Tendo de ser deixado para ser tratado em futuras pesquisas, dado o limite temporal para a conclusão desta pesquisa e o enfoque desta tese ao contexto Brasileiro.

8 BIBLIOGRAFIA

- Ahmad, K. (2008). Being in Text and Text in Being: Notes on Representative Texts. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 60-94). Toronto: Multilingual Matters.
- Alexa, M., & Zuell, C. (2002). Text Analysis Software: Commonalities, Differences and Limitations: The Results of a Review. *Quality and Quantity*, 299-321.
- Allen, J., & Core, M. (1997). *Draft of DAMSL: Dialog Act Markup in Several Layers*. Retrieved from <http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/>
- Alves, D. (2006). *Aspectos da Representação do Discurso em Textos Traduzidos: Os Verbos de Elocução Neutros*. Belo Horizonte: Faculdade de Letras UFMG.
- Alves, D. (2006). *Aspectos da representação do discurso em textos traduzidos: verbos de elocução neutros* (Vol. (Dissertação de Mestrado)). (O. A. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- Alves, D. (2012). Interfaces between Corpus Linguistics and Translation Studies in Brazil (2006-2010). *III Congresso Internacional da ABRAPUI: Language and Literature in the Age of Technology*. Florianopolis: Universidade Federal de Santa Catarina.
- Alves, D., & Vasconcellos, M. (no prelo). A LINGUÍSTICA DE CORPUS E OS ESTUDOS DA TRADUÇÃO: UMA ANÁLISE BIBLIOMÉTRICA DE PESQUISAS BRASILEIRAS PUBLICADAS ENTRE 2006-2010.
- Alves, F. (2000). Unidades de Tradução: o que são e como operá-las. In A. Pagano, C. Magalhães, & F. Alves, *Traduzir com autonomia: estratégias para o tradutor em formação* (pp. 29-38). Sao Paulo: Contexto.
- Alves, F. (2005). Ritmo Cognitivo, Meta-Reflexão e Experiência: Parâmetros de Análise Processual no Desempenho de Tradutores Novatos e Experientes. In A. Pagano, C. Magalhaes, & F. Alves, *Competência em Tradução -- Cognição e Discurso* (pp. 109 - 169). Belo Horizonte: UFMG.
- Anderman, G., & Rogers, M. (2008). *Incorporating Corpora: The Linguist and the Translator*. Clevedon, Buffalo, Toronto: Multilingual Matters.
- Anthony, L. (2012). *Laurence Anthony's Website*. Retrieved from <http://www.antlab.sci.waseda.ac.jp/software.html>

- Araújo, C. (2007). *O Sistema Semântico de Projeção e sua Dispersão Gramatical em Português Brasileiro: uma Descrição Sistêmico-Funcional Orientada para os Estudos Linguísticos da Tradução*. Belo Horizonte: Faculdade de Letras/ UFMG.
- Araújo, C. (2007). *O sistema semântico de PROJEÇÃO e sua dispersão gramatical em português brasileiro: uma descrição sistêmico-funcional orientada para os estudos linguísticos da tradução* (Vol. (Dissertação de Mestrado)). (O. A. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- Arrojo, R. (2007). *Oficina de Tradução: a teoria na prática*. São Paulo: Ática.
- Assis, R. (2009). *A representação de europeus e de africanos como atores sociais em Heart of Darkness (O coração das trevas) e em suas traduções para o português: uma abordagem textual da tradução* (Vol. (tese de doutorado)). (O. D. Magalhães, Ed.) Belo Horizonte: UFMG/FALE.
- Assis, R. (2012). A interface Tradução e Linguística Sistêmico-Funcional no Brasil. *Traduzires, 1 (1)*, 61-71.
- Assis, R. C. (2004). *A Transitividade na Representação de Sethe no Corpus Paralelo Beloved-Amada* (Vol. agosto). (UFMG, Ed.) Belo Horizonte: Programa de Pós-Graduação da Faculdade de Letras.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing 7(1)*, 1-16.
- Aubert, F. (1993). *As (in)fideliades da Tradução*. Campinas: Editora da Unicamp.
- Baker, M. (1992). *In Other Words: A coursebook on Translation*. London: Routledge.
- BAKER, M. (1992). *In Other Words: a Coursebook on Translation*. London: Routledge.
- Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis, & E. Tognini-Bonelli, *Texts and Technology: in Honour of John Sinclair* (pp. 233-250). Amsterdam and Philadelphia: John Benjamins.
- Baker, M. (1995). *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*. Amsterdam: John Benjamins.
- BAKER, M. (1995). *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*. Amsterdam: John Benjamins B.V.
- Baker, M. (1998). *Encyclopedia of Translation Studies*. London: Routledge.
- Baker, M. (n.2 de v.4 de 1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics*, pp. 281-298.
- BAKER, M. (1999, v.4 n.2). The Role of Corpora in Investigation the Linguistic Behavior of Professional Translators. pp. 281-298.

- Baldissera, E. (2010). *Illuminated te Analysis of the Translation is: Systemic Functional Linguistics Strikes Yoda Back* (Vol. (tese de doutorado)). (O. D. Vasconcellos, Ed.) Florianopolis: UFSC/PPGI.
- Barbosa, H. (2004). A Formação do Tradutor/Intérprete e sua Inserção no Mercado. *IX Encontro Nacional de Tradutores e III Encontro Internacional de Tradutores (ABRAPT)*. Fortaleza: (mesa redonda "Perspectivas para o Ensino da Tradução").
- Barlow. (2003). *Paraconc.pdf*. Retrieved from ParaConc: A Concorde for Parallel Texts: <http://www.athel.com/paraconc.pdf>
- Barlow, M. (s/d). *Analysing Parallel Texts with ParaConc*. Retrieved from <http://gandalf.aksis.uib.no/allc/barlow.pdf>
- Barnbrook, G. (1996). *Language and Computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Bartholamei Junior, L. (2010). *O Novum e os Padrões Preferenciais nas Traduções de Speaker for the Dead de Orson Scott Card: Um Estudo Baseado em Corpus*. (O. D. Fernandes, Ed.) Florianopolis: UFSC (Dissertação de Mestrado).
- Bell, R. (1991). *Translation and Translating: Theory and Practice*. London: Longman.
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, 257-269.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 243-257.
- Biber, D., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Borja, A. (2008). Corpora for Translators in Spain: The CDJ-GITRAD Corpus and the GENTT Project. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 243-265). Clevedon, Buffalo, Toronto: Multilingual Matters.
- Borja, A. (2008). Corpora for Translators in Spain: The CDJ-GITRAD Corpus and the GENTT Project. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 243-265). Clevedon, Buffalo, Toronto: Multilingual Matters.
- Bowker, L. (2001). Towards a Methodology for Corpus-Based Approach to Translation Evaluation. *Meta: journal des traducteurs/ Meta: Translator's Journal*, 46, 345-364.
- Bowker, L. (2002). *Computer-Aided Translation Technology: a practical introduction*. Ottawa: University Ottawa Press.
- Braga, C. (2007). *Indagando o perfilde tradutores em formação: um estudo de caso*. Belo Horizonte: UFMG/FALE - Orientadora A.S. Pagano (Dissertação de Mestrado).

- Bueno, L. (2005). *Transitividade, Coesão e Criatividade Lexical no Corpus Paralelo Macunaíma, de Andrade e Macunaíma, de Goodland* (Vol. Dissertação de Mestrado). (O. C. Magalhães, Ed.) Belo Horizonte: UFMG/FALE.
- Campestrato, L. (2002). *The traffic of English words in the Brazilian translation context: a case study* (Vol. (Dissertação de Mestrado)). (O. M. Vasconcellos, Ed.) Florianópolis: UFSC/PPGI.
- Campestrato, L. (2002). *Thematic Structure in Brazilian portuguese abstracts in English translation: steps towards a discussion of textual competence*. Florianópolis: UFSC/PPGI.
- Cançado, T. (2005). *Transitividade e Representação do Discurso no Corpus Paralelo "Interview with the Vampire"/ "Entrevista com o Vampiro"* (Vol. (Dissertação de Mestrado)). (O. C. Magalhães, Ed.) Belo Horizonte: Faculdade de Letras UFMG.
- Chestermann, A. (2009). The Name and Nature of Translator Studies. *Journal of Language and Communication Studies*(42).
- Confluências, R. (2010, 11 27). Retrieved from <http://www.confluencias.net/>
- CORDIAL, P. (2010, 03 24). *Letras UFMG*. Retrieved from <http://www.lettras.ufmg.br/net/cordial/portugues/projeto.htm>
- COSTA, W. (1992). *A Linguistic Approach to the Analysis and Evaluation of Translated Texts with Special Reference to Selected Texts by J.L. Borges*. University of Birmingham.
- Costa, W. (1992, v.28). The Translated Text as (Re)Textualization. *Ilha do Desterro*.
- Coulthard, M. (1987). Evaluative Text Analysis. In S. & Treadgold, *Language Topics - Essays in Honour of Michael Halliday*. Amsterdam: John Benjamins.
- CRUZ, O. M. (2003). *"Harry Potter and the Chamber of Secrets" e sua Tradução para o Português do Brasil: uma Análise dos Verbos de Elocução com Base na Linguística Sistêmica e nos Estudos de Corpora*. Belo Horizonte: Programa de Pós-Graduação da Faculdade de Letras, UFMG.
- Cruz, O. M. (2003). *"Harry Potter and the Chamber of Secrets" e sua tradução para o português do Brasil: uma análise dos verbos de elocução com base na linguística sistêmica e nos estudos de corpora*. Belo Horizonte: Faculdade de Letras da UFMG/PosLing.
- Daves, M. (2005). The Advantage of Using Relational Databases for Large Corpora: Speed, Advanced Queries, and Unlimited Annotation. *International Journal of Corpus Linguistics*, 307-334.
- Deslile, J., & Woodsworth, J. (1995). *Translators Through History*. Montreal: John Benjamins.
- DESLILE, J., & WOODSWORTH, J. (1995). *Translators Through History*. Montreal: John Benjamins.

- Deslile, J., & Woodsworth, J. (2003). *Os Tradutores na História*. (S. Bath, Trans.) São Paulo: Ática.
- DESLILE, J., & WOODSWORTH, J. (2003). *Os Tradutores na História*. (trad. Sérgio Bath) São Paulo: Atica.
- Egbert, J., & Petrie, G. (2005). *CALL Research Perspectives*. New Jersey: Lawrence Erlbaum Association Publisher.
- Egbert, L., & Hanson-Smith, E. (1999). *CALL Environments: Research, Practice, and Critical Issues*. Virginia: TESOL.
- Eggins, S. (1994). *An Introduction to Systemic-Functional Linguistics*. London: Pinter.
- Feitosa, M. (2005). *Uma Proposta de Anotação de Corpora Paralelos com Base na Linguística Sistemico-Funcional*. Belo Horizonte: Programa de Pós-Graduação da Faculdade de Letras/UFMG.
- Feitosa, M. (2009). *Legendagem comercial e legendagem pirata: um estudo comparado* (Vol. (tese de Doutorado)). (O. A. Pagano, Ed.) Belo Horizonte: Universidade Federal de Minas Gerais.
- Fernandes, A. (2009). *Black into white and preto no branco: can you tell one's colour by the company one keeps?* (Vol. (Dissertação de Mestrado)). (O. D. Vascellos, Ed.) Florianópolis: UFSC/PPGI.
- Fernandes, L. (2004). *Translating Children Literature: A Corpus-Based Study of Proper Names*. Florianópolis: Programa de Pós-Graduação em Letras Inglês/UFSC.
- Fernandes, L. (2006). Corpora in Translation Studies: Revising Baker's Typology. *Fragmentos*, 30, 87-95.
- Figueredo, G. (2007). *Uma descrição sistêmico-funcional da estrutura do grupo nominal em português orientada para estudos linguísticos da tradução* (Vol. Dissertação de Mestrado). (O. D. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- Filgueiras, J. (2007). *Capoeira em Tradução: Representações discursivas em um Corpus paralelo bilíngue* (Vol. (Dissertação de Mestrado)). (O. M. Vasconcellos, Ed.) Florianópolis: UFSC/PGET.
- Filgueiras, J. d. (2007). *Capoeira em Tradução: Representações Discursivas em um Corpus Paralelo Bilingue*. Florianópolis : Pos-Graduação em Estudos da Tradução/UFSC.
- Fillmore, C. (1992). Corpus Linguistics or Computer-Aided armchair linguistics. In I. Svartvik, *Directions in Corpus Linguistics: Processings of Nobel Symposium 82, Stockholm 4-8 August 1991*. Berlin: Mouton de Gruyter.
- Firth, J. (1957). A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, 1-32.
- Fleuri, L. (2010). PROPOSTA DE UM MODELO SISTEMATIZADO DE METODOLOGIA PARA PESQUISAS NA INTERFACE ENTRE ESTUDOS DA TRADUÇÃO, LINGUISTICA SISTÊMICO-

- FUNCIONAL E LINGUISTICA DE CORPUS. *VI Congresso da Associação de Linguística Sistêmico-Funcional da América Latina (ALSFAL)*. Fortaleza. Retrieved from www.6alsfal-uece.com.br.
- Fleuri, L. (2011). Methodological Proposal for Research on Translation Studies in interface with Systemic-Functional Linguistics and Corpus-Based Methodology. *International Conference on Translation and Cross-Cultural Communication*. Brisbane: University of Queensland. Retrieved from http://www.slccs.uq.edu.au/docs/ConferenceProgram_Final.pdf
- FLEURI, L. J. (2006). *O Perfil Ideacional dos Itens Lexicais Tradutor/Tradutor e "Translators Through History" e "Os Tradutores na História"*. UNIVERSIDADE FEDERAL DE SANTA CATARINA. FLORIANOPOLIS: UFSC.
- Fleuri, L. J. (2006). *O Perfil Ideacional dos Itens Lexicais Tradutor/Tradutor em "Translator Through History"*. Florianópolis: Programa de Pós-Graduação em Estudos da Tradução/UFSC.
- Fleuri, L., & Vasconcellos, M. (2011). Proposta de Sistematização Metodológica para Pesquisas em Análise Textual e Tradução: uma Interface com a Linguística Sistêmico-Funcional e Linguística de Corpus. *I Seminário de Pesquisa do Grupo Tradução e Corpora*. Florianópolis: Universidade Federal de Santa Catarina. Retrieved from <http://www.tracor.ufsc.br/seminario/texto>
- Fleuri, L., Pagano, A., & Vasconcellos, M. (2009). A REPRESENTAÇÃO DO PARTICIPANTE “TRADUTOR/TRANSLATOR” EM TRANSLATOR THOUGH HISTORY E OS TRADUTORES NA HISTORIA. *Cadernos de Tradução*, 159-192.
- Freire, P. (1989). *A importância do ato de ler: em três artigos que se completam*. São Paulo: Cortez.
- Frota, M. (2007). UM BALANÇO DOS ESTUDOS DA TRADUÇÃO NO BRASIL. *Cadernos de Tradução*, 135- 169.
- Frota, M. P., Martins, M., & Cristina, R. (1994). Um Breve Histórico do GT de Tradução e sua Importância para o desenvolvimento da área em nosso país. *Revista da ANPOLL*, 1, 60-70. Retrieved from http://letra.letras.ufmg.br/gttrad/breve_historico.html
- Gallina. (1992, 4). Cohesion and the Systemic-Functional Approach to Text: Application to Political Speeches and the Significance to Simultaneous Interpretation. *Interpreter's News Letter*, pp. 62-71.
- Garretson, G. (2008). Desiderata for Linguistic Software Design. *Internatinal Journal of English Studies (IJES)*, 8 (1), 67-94.
- Genette, G. (1982/1998). *Palimpsests: Literature in the Second Degree (Stages)*. Nebraska: University of Nebraska Press.

- Ghadessy, M., & Gao, Y. (2001). Small Corpora and Translation: Comparing Thematic Organization in Two Languages. In M. G. al., *Small Corpus Studies and ELT: Theory and Practice* (pp. 335-359). Amsterdam: John Benjamins.
- Ghadessy, S., Henry, A., & Roseberry, R. (2001). *Small Corpus Studies and ELT: Theory and Practice*. Amsterdam/Philadelphia: John Benjamins.
- Gorp, H. V. (2006). *On Describing Translation. Functional Approaches to Culture and Translations: selected papers by José Lambert*. (D. Delabastita, L. D'hulst, & R. Meylaerts, Eds.) Amsterdam/Philadelphia.
- Goveia, C., & Barbara, L. (2003). *Tema e Estrutura Temática em PE e PB: um Estudo Contrastivo das Traduções portuguesa e brasileira*. São Paulo/Liverpool: LAEL - Pontífica Universidade Católica/ AESLU - University of Liverpool.
- Gragger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam/Philadelphia: John Benjamin.
- Guerini, A., & Torres, M.-H. (2006). A pós-graduação em Estudos da Tradução da UFSC: Trajetórias e Projeções. *Cadernos de Tradução*, 19-24.
- Gysel, E. (2009). *COMPARATIVE ANALYSIS OF THE LEXICAL ITEMS TRANSLATION/TRADUÇÃO AND TRANSLATOR/TRADUTOR IN A BILINGUAL CORPUS: PARATEXTS AND INTRODUCTIONS OF ACADEMIC BOOKS ON TRANSLATION STUDIES IN BRAZIL* (Vol. Trabalho de Conclusão de Curso (TCC)). (O. M. Vasconcellos, Ed.) Florianópolis, SC, Brasil: UFSC.
- Hale, S. (1997, v.3(1)). The Treatment of Register Variation in Court Interpreting. *The Translator*, pp. 39-54.
- Halliday, M. (2006). *Computational and Quantitative Studies*. (J. Webster, Ed.) New York: Continuum.
- Halliday, M. A. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.
- Halliday, M. A. (1985/1994/2004). *An Introduction to Functional Grammar*. London: Arnold.
- Halliday, M. A. (2001). Towards a Theory of a Good Translation. In E. Steiner, & Yallop, *Exploring Translation and Multilinguas Text Production: Beyond Content*. New York/Berlin: Mouton de Gruyter.
- Halliday, M. A. (2004b). *The Language of Science*. (J. J. Webster, Ed.) New York, MPG Bookes Ltd.
- Halliday, M. A., & Hassan, R. (1989). *Language, Context, and Text: Aspects of Language in a Social-Seciotic Perspective*. Hong Kong: Oxford University Press.

- Halliday, M. A., & Mathiessen, C. (2004). *An Introduction to Functional Grammar*. London: Arnold.
- Hatim, B., & Mason, I. (1990). *Discourse and the translator*. London: Longman.
- Hatim, B., & Munday, J. (2004). *Translation: An Advanced Resource Book*. London: Routledge.
- Herman, T. (1999). *Translation in Systems: Descriptive and Systemic approaches explained*. Manchester: St. Jerome.
- Hermans, T. (1985). Translation Studies and a new Paradigm. In T. Hermans, *The Manipulation of Literature: Studies in Literary Translation* (pp. 7-15). London: Croom Helm.
- Hermans, T. (1991). Translational Norms and Correct Translation. In K. Van Leuven-Zwart, & T. Naaijken, *Translation Studies: The state of the art. Proceedings of the First James S. Holmes Symposium on Translation Studies* (pp. 155-169). Amsterdam: Rodopi.
- Hirst, D. (1991). Intonation models: towards a third generation. *Actes du XIIeme Congres International des Sciences phonetiques* (pp. 305-310). Aix-en-Provence, France: Universite de Provence, Service des Publications.
- Holliday, O. J. (2006). *Para Sistematizar Experiências*. Brasília: MMA. Retrieved from <http://culturadigital.br/gepepi/files/2011/02/sistematizacao-jara.pdf>
- Holmes, J. (2000). The Name and Nature of Translation Studies. In V. Laurence, & M. Baker (Ed.), *The Translation Studies Reader* (pp. 172-185). London and New York: Routledge. Retrieved from http://www.academia.edu/3283416/Translation_studies_reader
- In-Traduções. (2010, 11 27). *PGET/UFSC*. Retrieved from <http://www.pget.ufsc.br/in-traducoes/>
- Jesus, S. (2004). *Representação do Discurso e Tradução: Padores de Textualização em Corpora Paralelos e Comparáveis*. Belo Horizonte: Faculdade de Letras UFMG/PosLin.
- Jesus, S. (2004). *Representação do Discurso e Tradução: padrões de textualização em corpora paralelo e comparável* (Vol. (Dissertação de Mestrado)). (O. A. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- Jesus, S. (2008). *Relações de tradução: SAY/DIZER em corpora de textos ficcionais* (Vol. (Tese de Doutorado)). (O. A. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- Jesus, S. (2008). *Relações de Tradução: Say/Dizer em Corpora de Textos Ficcionalis*. Belo Horizonte: Faculdade de Letras UFMG/PosLin.
- Johns, T. (1992, v.0 n.27). It is Presented Initially: Linear Dislocation & Inter-Language Strategies in Brazilian Academic Abstracts in English and Portuguese. *Ilha do Desterro*.

- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. New York: Logman.
- Kenny, D. (2001). *Lexis and Creativity in Translation: A Corpus-Based Study*. Manchester: St. Jerome.
- Knowles, G., Williams, B., & Taylor, L. (1996). *A corpus of formal British English Speech: The Lancaster/IBM Spoken English Corpus*. London: Longman.
- Knowles, M., & Moon, R. (2006). *Introducing Metaphor*. New York: 2006.
- Kobelinski. (2005). *A Corpus-Based Approach to ESP: EST Vocabulary in Information Technology*. Florianopolis: UFSC/PPGI - orientadora M.L. Vasconcellos (Dissertação de mestrado).
- Kovala, U. (1996). Translation, Paratextual Mediation and Ideological Closure. *Target*(8:1), pp. 119-147.
- Kruger, A. (2002). CORPUS-BASED TRANSLATION RESEARCH: Its development and implications for general, literary and Bible translation. *African Journals Online*, 70-106. Retrieved from <http://www.ajol.info/index.php/actat/article/view/5455>
- Lambert, J. (1996). Language and Translation as General Management Problems. In C. Dollerup, & V. Appel.
- Laviosa-Braithwaite. (1996). *The English Comparable Corpus (ECC): a resource and a methodology for the empirical study of translation*. ((. thesis), Ed.) Manchester: UMIST.
- Leech. (1992). Corpora and theories of linguistic performance. In *Startvik* (pp. 105-122).
- Leech, G. (1991). The State of the art in Corpus Linguistics. In K. Aijmer, & B. Altenberg, *English Corpus Linguistics: Study in Honour of Jan Svartvik* (pp. 8-29). London: Logman.
- Leech, G. (2005). Adding Linguistic Annotation. (M. Wynne, Ed.) *Developing Linguistic Corpora: A Guide to Good Practice*, 17-29. Retrieved 2012, from <http://ahds.ac.uk/linguistic-corpora/>
- Leech, G., & Weisser, M. (2003). Generic Speech Act Annotation for Task-Oriented Dialogue. (D. Archer, P. Rayson, A. Wilson, & A. McEnery, Eds.) *Proceedings of the Corpus Linguistics 2003 Conference*.
- Leuven-Zwart, V. (1990, 2(1)). Translation and Original: Similarities and Dissimilarities. *Target*, pp. 69-95.
- Levy, M. (1997). *CALL: Context and Conceptualization*. Oxford: Oxford Univesity Press.
- Lima, K. (2008). *Conhecimento Experto em Tradução: Orientaçõe e revisão em tarefas tradutórias executadas por pesquisadores expertos não tradutores* (Vol. (Dissertação de Mestrado)). (O. A. Pagano, Ed.) Belo Horizonte: UFMG/FALE.

- Magalhães, C. (2005). Coesão como Recurso de Continuidade do Discurso. In A. Pagano, C. Magalhães, & F. Alves, *Competência em Tradução: Cognição e Discurso* (pp. 209-245). Belo Horizonte: Editora da UFMG.
- Maia, B. (1998). Word Order and the First Person Singular in Portuguese and English. *Meta*, 43, 589-601.
- Malmkjaer, K. (1998). www.erudit.org/revue/meta/1998/v43/n4/003545.html. Retrieved November 8, 2007, from www.erudit.org.
- Marcu, M., Santori, B., & Marcinkiewicz, M. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19: 313-330.
- Martin, J. R., Mathiessen, C. M., & Painter, C. (1997). *Working with Functional Grammar*. New York: Arnold.
- Mason, O. (2008). Developing Software for Corpus Research. *International Journal of English Studies (IJES)*, 8 (1), 141-156.
- Mauranen, A. (2008). Universal tendencies in translation. In G. Anderman, & M. Rogers, *Incorporating Corpora: The linguist and the translator* (pp. 33-48). Toronto: Multilingual Matters.
- Mauri, C. (2003). *Um Estudo de Tradução Italiana de "Laços de Família", de Clarice Linspector, a Partir da Abordagem em Corpora: a Construção da Introspecção Feminina Através dos Verbos de Elocução*. Belo Horizonte: Faculdade de Letras PosLin/UFMG.
- Mauri, C. (2009). *Uma análise do Ponto de Vista em Hora da Estrela e Laços de Família, de Clarice Linspector, e nas Traduções Italianas L'ora della Stella e Legami Familiari* (Vol. (tese de doutorado)). (O. C. Magalhães, Ed.) Belo Horizonte: UFMG- FALE- POSLIN.
- McAndrew, J., & McAndrew, P. (2002, n.3). Systemic Functional Linguistics: An Introduction. *Journal of the Faculty of Global Communication Siebold University of Nagasaki*.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., & Xiao, R. (2008). Parallel and Comparable Corpora: What is Happening. In G. Anderman, & M. Rogers, *Incorporating Corpora: the Linguist and the Translator* (pp. 18-31). Clevedon, Buffalo, Toronto: Multilingual Matters.
- McEnery, T.; Xiao, R., Tono, Y. (2006) *Corpus-Based Language Studies: An Advanced Resource Book*. New York: Routledge.
- Meyer, C. (2002). *English Corpus Linguistics: an introduction*. Cambridge: Cambridge University Press.
- Moresi, E. (2003). *PDF UFES*. Retrieved from <http://www.inf.ufes.br/~falbo/files/MetodologiaPesquisa-Moresi2003.pdf>

- Morinaka, E. M. (2005). *Gabriela, Cravo e Canela and its (re)textualization in English: Representation Through Lexical Relation*. Florianópolis: Pós-Graduação em Letras Inglês/UFSC.
- Mossop, B. (1994). Goals and Methods for a Course in Translation Theory. In M. Snell-Hornby, F. PöchHacker, & K. Kaindl, *Translation Studies: An Interdiscipline*. Amsterdam/Philadelphia: John Benjamins.
- Munday, J. (1998, n.3). Problems of Applying Thematic Analysis to Translation Between Spanish and English. *Cadernos de Tradução*, pp. 183-214.
- Munday, J. (2001). *Introducing Translation Studies: Theories and Applications*. London e New York: Routledge.
- Munday, J. (2002). Systems in Translation. A Systemic Model for Descriptive Translation Studies. In T. Hermans, *Crosscultural Transgressions, Research Models in Translation Studies, II. Historical and Ideological Issues*. (pp. 76-92). Manchester: St. Jerome.
- Newmark, P. (1988). *A textbook of translation*. New York: Prentice-Hall International.
- Nord, C. (1997). *Translating as a purposeful activity: Functionalist Approaches Explained*. Manchester: St. Jerome.
- Nunes, L. (2010). *As conjunções but e mas em textos ficcionais originais e traduzidos: uma análise tridimensional com base na linguística sistêmico-funcional* (Vol. (Dissertação de Mestrado)). (O. D. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- O'Donnel, M. (2004, maio). *Syscoder 4.6*. Retrieved from Web: <http://www.wagsoft.com>
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London/New York: Routledge.
- Pagano, A., & Vasconcellos, M. (2003). ESTUDOS DA TRADUÇÃO NO BRASIL: REFLEXÕES SOBRE TESES E DISSERTAÇÕES ELABORADAS POR PESQUISADORES BRASILEIROS NAS DÉCADAS DE 80 E 90. *D.E.L.T.A.*, 1-25.
- Pagano, A., & Vasconcellos, M. L. (2005). Explorando Interfaces: Estudos da Tradução, Linguística Sistêmico-Funciona e Linguística de Corpus. *Competências em Tradução: Cognição e Discurso*.
- Paquillin, V. (2005). *The Various Facets of a Message: An Analysis of the Thematic Structure in Bridget Jones's Diary in the light of the Systemic-Functional Grammar, Corpus Linguistics and Translation Studies Interface*. Florianópolis: Pós-Graduação em Letras Inglês/UFSC.
- PeriódicosUFSC. (2010, 11 27). Retrieved from <http://www.periodicos.ufsc.br/index.php/scientia/issue/current/>

- Piotrowski, T. (2008). The Translator and Polish-English Corpora. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 117-132). Toronto: Multilingual Matters.
- Pires, T. (2006). *The Ideational Profile of Bishop in "Flores Raras e Banalíssimas" and "Rare and Commonplace Flowers": A corpus-based Translation Study*. Florianópolis: Pós-Graduação em Letras Inglês/UFSC.
- Pires, T. (2009). *The construal of Bishop's Ideational profile in Flores Raras e Banalíssimas and Rare and Commonplace Flowers: a corpus-based Translation Study* (Vol. (Dissertação de Mestrado)). (O. D. Vasconcellos, Ed.) Florianópolis: UFSC/PPGI.
- PROCAD. (2006, 04 04). *Programa de Pós-Graduação em Inglês*. Retrieved from http://ppgi.paginas.ufsc.br/?page_id=96
- Rambousek, J., & Chamonikolasová, J. (2008). The Existential There-Construction in Czech Translation. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 133-152). Toronto: Multilingual Matters.
- Rasinger, S. (2008). *Review of Quantitative Research in Linguistics*. ((. T. Hoffmann, Ed.) Retrieved from The Linguist List: <http://linguistlist.org/pubs/reviews/get-review.cfm?SubID=2230895>
- Ravelli, L. (1988). Grammatical Metapho: An Initial Analysis. In E. H. Steiner, & R. Veltman, *Pragmatics, Discourse and Text. Some Sistemically-Oriented Approaches* (pp. 133-147). London: Pinter.
- RevistaNotadoTradutor. (2010, 11 27). Retrieved from <http://www.notadotradutor.com/revista.html>
- Ribeiro, G. (2004). Tradução Técnica, terminologia e lingüística de corpus: a ferramenta Wordsmith Tools. *Cadernos de Tradução*, 159-174.
- Robinson, D. (2003). *Becoming a Translator: an Accelerated Course*. London and New York: Routledge.
- Rodrigues Junior, A. (2006). *A representação de personagens gays na coletânea de contos Stud e em sua tradução As Aventuras de um Garoto de Programa* (Vol. (tese de doutorado)). (O. E. Mendes, Ed.) Belo Horizonte: UFMG/FALE.
- Rodrigues, R. (2005). *A organização Temática em "A hora da Estrela" e "The hour of the Star"*. Belo Horizonte: Faculdade de Letras/UFMG.
- Rodrigues, R. (2010). *Tradução e Apresentação do Discurso: Um Estudo de "Bliss" de Katherine Mansfield* (Vol. (tese de Doutorado)). (O. A. Pagano, Ed.) Belo Horizonte: UFMG/FALE.
- Sampson, G. (1995). *English for the computer: The SUSANNE corpus and analytic scheme*. Oxford: Clarendon Press.

- Sandrelli, A., & Jerez, J. (2007). The Impact of information and Communication Technology on Interpreter Training. *The Interpreter and Translator Trainer*, 269-303.
- Santos, D. (2008). Perfect Mismatches: 'Result' in English and Portuguese. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 217-242). Toronto: Multilingual Matters.
- Sardinha, B. (2004, agosto). Linguísticade Corpus: um entrevista com Tony Berber Sardinha. (R. V. ReVel, Interviewer) Retrieved from http://www.revel.inf.br/files/entrevistas/revel_3_entrevista_tony_berber_sardinha.pdf
- Sardinha, T. (2000). Corpus linguistics: history and problematization. *Delta*, 16(2), 323-367.
- Sardinha, T. B. (1999). *Usando WordSmith Tools na Investigação da Linguagem*. Retrieved outubro 07, 2009, from <http://www2.lael.pucsp.br/direct/DirectPapers40.pdf>.
- Sardinha, T. B. (2004). *Linguística de Corpus*. Barueri (SP): Manole.
- Schopenhauer, A. (2008). *The World as Will and Presentation*. (R. E. Carus, Trans.) New York: Longman.
- SciELO, R. (2010, 11 27). Retrieved from http://www.scielo.br/scielo.php?script=sci_serial&pid=0102-4450&lng=en&nrm=iso
- Scott, M. (1999). *WordSmith Tools*. Oxford University Press.
- Scott, M. (2001). Comparing Corpora and Identifying keywords collocations, frequency distributions through the WordSmith Tools suite of computer programs. In M. Ghadessi, M. Henry, & R. Roseberry, *Small Corpus Studies and ELT: Theory and Practice* (pp. 47-67). Amsterdam/Philadelphia: John Benjamins.
- Scott, M. (2008). Developing WordSmith. *International Journal of English Studies (IJES)*, 8 (1) , 95-106.
- Scott, M. (2010). *WordSmith Tools: software for finding word patterns*. Retrieved from WordSmith Tools: software for finding word patterns: http://www.lexically.net/downloads/version5/HTML/?getting_start_ed.htm
- Scott, M. (2010b, junho 27). *WordSmith Tools*. Retrieved from <http://www.lexically.net/downloads/version6/HTML/?wshell.htm>
- Serpellet, N. (2008). A Study of the Mandative Subjunctive in French and its Translation in English: A Corpus-based Contrastive Analysis. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 187-216). Toronto: Multilingual Matters.
- Silva, I. (2007). 4. *Conhecimento experto em tradução: aferição da durabilidade de tarefas tradutórias realizadas por sujeitos não-*

- tradutores em condições empírico-experimentais*. Belo Horizonte: UFMG/FALE - Orientadora A. S.Pagano .
- Simov, K., Simov, A., Ganey, H., Ivanova, K., & Grigov, k. (2004). The Clark system: XML-based corpora development system for rapid prototyping. In M. Lino, M. Xavier, F. Ferreira, R. Costa, & R. Silva, *Proceedings of Language Resources and Evaluation Conference 2004* (pp. 235-238). Lisboa, Portugal.
- Sinclair, J. M. (2001). Preface. In M. Ghadessy, A. Henry, & R. Roseberry, *Small Corpus Studies and ELT. Theory and Practice* (pp. vii-xv). Amsterdam: John Benjamin.
- Smith, N., Hoffmann, S., & Rayson, P. (2008). Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists Manual Annotation. *Literary and Linguistic Computing* 23, 163-180.
- Souza, L. (2006). *O Modelo de Linguagem Avaliativa (appraisal framework) como Ferramenta para a Análise Descritiva do Texto Traduzido* (Vol. (Dissertação de Mestrado)). (O. M. Weininger, Ed.) Florianópolis: UFSC/PGET.
- Souza, L. (2010). *Interlingual Re-Instantiation: a model for a new and more comprehensive systemic functional perspective on translation* (Vol. (Tese de Doutorado)). (O. M. Vasconcellos, Ed.) Florianópolis: UFSC/PGL.
- Steiner, E., & Yallop, C. (2001). *Exploring Translation and Multilingual Text Production Beyond Content*. New York: Mouton de Gruyter. Retrieved from http://books.google.com.au/books?id=Vyh1ErgK7uIC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell Publisher.
- Taylor, C., & Baldry, A. (2001). Computer Assisted text analysis and translation: a functional approach in the analysis and translation of advertising texts. In E. Steiner, & C. Yallop, *Exploring Translation and Multilingual Text Production: Beyond Content* (pp. 277-303). New York: Mouton de Gruyter.
- Thompson, G. (2004, 11:1). *Introducing Functional Grammar*. Great Britain: Arnol.
- Thompson, G., & Hunston, S. (2006). *System and Corpus: Exploring Connections*. Oakville: David Brown Book Company.
- Timarová, S. (s/d). *Corpus Linguistics Methods in Interpreting Research: A case Study*. Retrieved from <http://www.openstarts.units.it/dspace/bitstream/10077/2471/1/05.pdf>
- Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.

- TRACOR. (2013). *Projeto Tradução e Corpora*. Retrieved from <http://www.tracor.ufsc.br/>
- TradTerm. (n.d.). *USP*. Retrieved from <http://www.fflch.usp.br/citrat/tradterm.htm>
- Tradução, R. (2010, 11 27). *Unibero*. Retrieved from http://www.unibero.edu.br/nucleosuni_revtraducao.asp
- Tymoczko, M. (1998, v.XLIII n.4). Computerized Corpora and the future of Translation Studies. *META*.
- Tymoczko, M. (2002). Connecting the Two Infinite Orders: Research Methods in Translation. In T. Hermans (Ed.), *Crosscultural Transgressions: Research Models in Translation Studies II: Historical and Ideological Issues*. St. Jerome Publishing.
- Vandenbergen, A., & Ravelly, M. (2003). *Grammatical Metaphor*. Amsterdam: John Benjamins.
- Váradi, T. (2008). A Study of the Mandative Subjunctive in French and its Translations in English: A Corpus-based Contrastive analysis. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 168-186). Toronto: Multilingual Matters.
- Varantola, K. (2002). DISPOSABLE CORPORA AS INTELLIGENT TOOLS IN TRANSLATION. *Cadernos de Tradução*, 171-189.
- Vasconcellos, M. (2004). (Re)(textualizing Joyce' Eveline: Translation from a Systemic Functional perspective. *Tradução e Comunicação*, 83-118.
- Vasconcellos, M. (2009). The Fuzzy Place of Linguistics in Translation Studies (TS). In T. F. N., & L. Diniz, *Itinerários: Homenagem a Solange Ribeiro de Oliveira* (pp. 353-374). Belo Horizonte: FALE/UFMG.
- Vasconcellos, M. L. (1997). *Retextualizing Dubliners: A Systemic Functional Approach to Translation Quality Assessment*. Florianopolis: Ph.D. Dissertation - Universidade Federal de Santa Catarina.
- Vasconcellos, M. L. (2009). Systemic Functional Translation Studies (SFTS): The Theory Travelling in Brazilian Environments. *Delta*.
- Venuti, L. (1995). Invisibility. In L. Venuti, *The Translator's Invisibility: A History of Translation* (pp. 1-43). London: Routledge.
- Venuti, L. (2000). *The Translation Studies Reader*. London and New York: Routledge.
- Vintar, S. (2008). Corpora in Translator Training and Practice: A Slovene Perspective. In G. Anderman, & M. Rogers, *Incorporating Corpora: The Linguist and the Translator* (pp. 153-167). Toronto: Multilingual Matters.
- Wiechmann, D., & Fuhs, S. (2006). Concordancing Software. *Corpus Linguistics and Linguistics Theory*, 107-127.

- Williams, J., & Chesterman, A. (2002). *The Map: A Beginners' Guide to Doing Research in Translation Studies*. Manchester e Kinderhook: St. Jerome Publishing.
- Wilss, W. (1999, 11:1). Interdisciplinarity in Translation Studies. *Target*, pp. 131-144.
- Zanella, A. (2006). *Mapeamento Macro e Micro estrutural da retextualização de resumos on-line: Estudo da Transitividade de abstracts biomédicos* (Vol. (Dissertação de Mestrado)). (O. R. Lima, Ed.) Florianópolis: UFSC/PGET.
- Zanettin, F. (2000). Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis. In M. Olohan, *Intercultural Faultlines. Research Models in Translation Studies I Textual and Cognitive Aspects*. Manchester: St. Jerome.
- Zhu, C. (1993). *Structure of Meaning (SOM): Towards a Three-Dimension Perspective on Translating Between Chinese and English*. Nottingham: University of Nottingham.
- Zuniga, G. (2006). *Construing the Translator in "Becoming a Translator" and "Construindo o Tradutor": a Case Study Based on Corpus and Systemic Linguistics*. Florianópolis: Pós-Graduação em Letras Inglês/UFSC.

9 ANEXOS

Anexo 1: Lista Assis (2012)

No.	Autor	Ano	Orientador	IFES	Tipo	
					Mestrado	Doutorado
01	Maria Lúcia B. Vasconcellos	1997	Rosa Weingold Konder	UFSC		X
02	Osilene M. S. S. Cruz	2003	Adriana Silvína Pagano	UFMG	X	
03	Cristina Mauri	2003	Célia Maria Magalhães	UFMG	X	
04	Silvana M. de Jesus	2004	Adriana Silvína Pagano	UFMG	X	
05	Roberto C. de Assis	2004	Célia Maria Magalhães	UFMG	X	
06	Leticia T. Bueno	2005	Célia Maria Magalhães	UFMG	X	
07	Roberta R. Rodrigues	2005	Adriana Silvína Pagano	UFMG	X	
08	Eliza M. Mormaka	2005	Maria Lúcia Barbosa de Vasconcellos	UFSC	X	
09	Tassiani M. L. Cançado	2005	Célia Maria Magalhães	UFMG	X	
10	Marcos P. Feitosa	2005	Adriana Silvína Pagano	UFMG	X	
11	Viviana Paquelin	2005	Maria Lúcia Barbosa de Vasconcellos	UFSC	X	
12	Lilian J. Fleuri	2006	Maria Lúcia Barbosa de Vasconcellos	UFSC	X	
13	Gleimara R. F. Zuniga	2006	Lincoln Paulo Fernandes	UFSC	X	
14	Ladjane M. F. de Souza	2006	Markus Johannes Weiningger	UFSC	X	
15	Daniel A. S. Alves	2007	Adriana Silvína Pagano	UFMG	X	
16	Giácomo P. Figueiredo	2007	Adriana Silvína Pagano	UFMG	X	
17	Cristiano Araújo	2007	Adriana Silvína Pagano	UFMG	X	
18	Joana de P. Filgueiras	2007	Maria Lúcia Barbosa de Vasconcellos	UFSC	X	
19	Silvana M. de Jesus	2008	Adriana Silvína Pagano	UFMG		X
20	Cristina Mauri	2009	Célia Maria Magalhães	UFMG	X	
21	Roberto C. de Assis	2009	Célia Maria Magalhães	UFMG	X	
22	Marcos P. Feitosa	2009	Adriana Silvína Pagano	UFMG	X	
23	Thiago B. Pires	2009	Maria Lúcia Barbosa de Vasconcellos	UFSC	X	
24	Roberta R. Rodrigues	2010	Adriana Silvína Pagano	UFMG		X
25	Elaine Espindola	2010	Maria Lúcia Barbosa de Vasconcellos	UFSC		X

Anexo 2: Lista Alves e Vasconcellos (no prelo)

Título	Autores	IES	Programa de Pós-graduação	Formato do Arq	Tamanho do Arq	Acesso	Resumo	Menciona Linguística de (palavra corpus) no título	Menciona Linguística de (palavra resumo)	Menciona Linguística de (palavra título)	Menciona Linguística de (palavra resumo)	Seretonado para análise final	Seretonado para análise final
1. A apresentação textual em A Hora da Estrela e The Hour of the Star	Roberta Rego Rodrigues	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	6,24 KB	104	Esta dissertação apresenta	Não	Sim	Sim	Sim	Sim	Sim
2. A apresentação de elementos de afecção como atores sociais em Heart of Darkness (O Coração das Trevas) e em suas traduções para o português, uma abordagem textual da tradução	Roberto Carlos de Assis	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	3,11 MB	92	Pesquisas desenvolvidas	Não	Sim	Sim	Sim	Sim	Sim
3. A apresentação de narrativas em A. E. no sistema de contos Sui e em sua tradução: as adaptações de um gênero de programa de televisão e de uma revista de televisão tridimensional com base na linguística sistêmico-funcional	Adail Sebastião Rodrigues Junior	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	1,08 MB	283	Seguindo a tradição de	Não	Sim	Sim	Sim	Sim	Sim
4. As adaptações de um gênero de programa de televisão e de uma revista de televisão tridimensional com base na linguística sistêmico-funcional	Leonardo Pereira Nunes	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	5,56, 64 KB	6	Esta dissertação se afilia	Não	Sim	Sim	Sim	Sim	Sim
5. Black invisible and not on the company website: all one's, colour by the company one keeps?	Almeida Balduino Pires Fernandes	UPSC	LETRAS (INGLÊS E LINGÜÍSTICA)	.pdf	3,02 MB	15	Esta dissertação foi	Não	Sim	Sim	Sim	Sim	Sim
6. Cultura em tradução, mapas sentados, iluminated the analysis of the translation system functional linguistic strikes, void, back	Joanna de Paula Figueiras	UPSC	ESTUDOS DA TRADUÇÃO	.pdf	1,69 MB	129	Este estudo integra o	Sim	Sim	Sim	Sim	Sim	Sim
7. Interlingual retranslation, a model for a new and more comprehensive systemic functional perspective on translation	Elaíne Espíndola	UPSC	LETRAS (INGLÊS E LINGÜÍSTICA)	.pdf	3,83 MB	2	Na área de Estudos da	Não	Sim	Sim	Sim	Sim	Sim
8. Interlingual retranslation, a model for a new and more comprehensive systemic functional perspective on translation	Ladaine Maria Farias de Souza	UPSC	LETRAS (INGLÊS E LINGÜÍSTICA)	.pdf	2,40 MB	3	Esta tese propõe um	Não	Sim	Sim	Sim	Sim	Sim
9. Leituras de tradução, SAU/QUIZ em somatório de textos funcionais	Marcos Pereira Felosa	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	1,11 MB	68	Este trabalho situa-se	Não	Sim	Sim	Sim	Sim	Sim
10. Mapas de tradução, SAU/QUIZ em somatório de textos funcionais	Ariana Zanella	UPSC	ESTUDOS DA TRADUÇÃO	.pdf	1,48 MB	405	Os resumos como	Não	Sim	Sim	Sim	Sim	Sim
11. O sistema semiótico de tradução, SAU/QUIZ em somatório de textos funcionais	Lilian Jurkevitz Pleuri	UPSC	ESTUDOS DA TRADUÇÃO	.pdf	4,62 MB	50	A pesquisa parte dos	Não	Sim	Sim	Sim	Sim	Sim
12. O sistema semiótico de tradução, SAU/QUIZ em somatório de textos funcionais	Cristiano Gonçalves Araújo	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	1,01,795 KB	646	O presente estudo	Não	Sim	Sim	Sim	Sim	Sim
13. O sistema semiótico de tradução, SAU/QUIZ em somatório de textos funcionais	Silvana Maria de Jesus	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	1,35 MB	74	Esta tese insere-se	Sim	Sim	Sim	Sim	Sim	Sim
14. O sistema semiótico de tradução, SAU/QUIZ em somatório de textos funcionais	Thiago Bianchi Pires	UPSC	LETRAS (INGLÊS E LINGÜÍSTICA)	.pdf	1,41 MB	21	Esta pesquisa, realizada	Sim	Sim	Sim	Sim	Sim	Sim
15. Uma análise do ponto de vista em A Hora da Estrela e L'Esca de Famille de Gerce L'Esca, e nas traduções para o português de Della Stella e Legami Familiari	Cristina Mauri	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	1,26 MB	101	Esta tese está	Não	Sim	Sim	Sim	Sim	Sim
16. Uma descrição sistêmico-funcional da estrutura do grupo nominal em português, adaptada para os estudos linguísticos da tradução	Giacomio Patrocínio Piqueiro	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	2,18 MB	98	Diante da necessidade	Não	Sim	Sim	Sim	Sim	Sim
17. Uma descrição sistêmico-funcional da estrutura do grupo nominal em português, adaptada para os estudos linguísticos da tradução	Marcos Pereira Felosa	UPMG	ESTUDOS LINGÜÍSTICOS	.pdf	2,20 MB	104	Este trabalho situa-se	Sim	Sim	Sim	Sim	Sim	Sim

Anexo 3: Resumos de teses e dissertações coletadas

	Data	Pesquisa
1.	2010	<p>Ladjane Maria Farias de Souza. INTERLINGUAL RE-INSTANTIATION: A MODEL FOR A NEW AND MORE COMPREHENSIVE SYSTEMIC FUNCTIONAL PERSPECTIVE ON TRANSLATION. 01/12/2010</p> <p>1v. 216p. Doutorado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): James Martin; Maria Lúcia Barbosa de Vasconcellos</p> <p>Biblioteca Depositária: Biblioteca Central</p> <p>Email do autor:</p> <p>Palavras - chave: re-instanciação interlingual, acoplamento, calibragem.</p> <p>Área(s) do conhecimento: LÍNGUAS ESTRANGEIRAS MODERNAS</p> <p>Banca examinadora: Maria Ester Wollstein Moritz Markus J. Weininger Peter White Solange Coelho Vereza Tania Maria Granja Shepherd</p> <p>Linha(s) de pesquisa: Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta tese propõe um novo modelo sistêmico-funcional de tradução como re-instanciação interlingual. Tal modelo é posto no fulcro de uma perspectiva tridimensional baseada</p>

		<p>na relação de complementaridade entre três hierarquias sistêmico-funcionais – realização, instanciação e individuação. Voltando o foco para a hierarquia de instanciação, o modelo define a re-instantiação interlingual como a remissão de um texto-alvo (doravante TA) a um texto-fonte (doravante TF). Supõe-se que tal remissão origina uma gama de relações intertextuais não apenas entre TF e TA mas também entre eles e várias outras instâncias nos sistemas linguísticos/culturais envolvidos. Re-instanciar um TF torna-se então uma renegociação de tais relações. A fim de caracterizar tal renegociação, o modelo distingue entre três matrizes constitutivas do TF que são recriadas na tradução - 1) suas relações instanciais, isto é, suas escolhas e combinações particulares de significados entre aqueles disponíveis no potencial global da língua-fonte (doravante LF); 2) suas relações intertextuais intralinguais, isto é, suas relações com outros textos da LF enquanto pertencentes ao mesmo discurso, gênero/registo e tipo de texto; e 3) suas relações com as leituras que propicia enquanto manifestadas nos intertextos de chegada da LF. A recriação dessas matrizes é feita através de um processo de gerenciamento estratégico que visa obter do leitor da língua-alvo (doravante LA) um determinado tipo de leitura. São propostos três modos de gerenciamento – citação, paráfrase e recontagem. Na citação, as relações estabelecidas na matriz 1 (relações instanciais) são privilegiadas e o tradutor elege o polo da instância da escala de instanciação da LF como ponto focal para a convergência entre os dois sistemas (enquanto repertórios) e exerce sua criatividade recriando padrões linguísticos do TF. Os movimentos de distanciação atingem o nível do sistema global visto que o tradutor pode precisar compelir o sistema da LA a produzir significados que até então permaneciam virtuais. Os potenciais de significado do TF e do TA são apresentados como completamente sobrepostos. Na paráfrase e na recontagem, as relações estabelecidas na matriz 2 (relações interdiscursivas e intertextuais) são privilegiadas e o tradutor elege o tipo de texto como o ponto focal para a convergência entre os sistemas. Este ponto é posto entre as duas escalas de instanciação visto que nenhuma delas é favorecida. A criatividade do tradutor é exercida na criação de um TA visto como pertencendo ao mesmo tipo textual que o TF, ou seja, os movimentos de distanciação atingem o nível do tipo de texto. No que concerne aos potenciais de significado, na paráfrase, eles são apresentados como parcialmente sobrepostos e na recontagem tal sobreposição é ainda menor. Supõe-se que a diferença entre esses modos de</p>
--	--	--

		<p>gerenciamento de relações intertextuais é proporcional à diferença entre os acoplamentos e calibrações de significado feitos no TF e no TA. A tese oferece uma sondagem inicial do modelo proposto, investigando a re-instanciação de recursos de valoração em 11 trios de textos (TF + 2 TAs). Tal fonte de dados é usada para explorar: 1) o gerenciamento de matrizes; 2) os movimentos de distanciação feitos pelos tradutores e 3) diferenças entre as leituras propiciadas pelo TF e pelo TA.</p>
2.	2010	<p>Roberta Rego Rodrigues. Tradução e Apresentação do Discurso: um estudo de "BLISS" de Katherine Mansfield. 01/03/2010</p> <p>1v. 221p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: Biblioteca Universitária da UFMG</p> <p>Email do autor:</p> <p>Palavras - chave: Bliss, Estilística Lingüística , Tradução Literária</p> <p>Área(s) do conhecimento: LETRAS LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Cláudio Márcio do Carmo Célia Maria Magalhães Fabio Alves da Silva Junior Paulo Henrique Caetano Silvana Maria de Jesus</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - PROF</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p>

Resumo tese/dissertação:

Esta tese se vincula ao LETRA (Laboratório Experimental de Tradução) da Faculdade de Letras da Universidade Federal de Minas Gerais e investiga a apresentação do discurso (discourse presentation) (SEMINO e SHORT, 2004) em um corpus paralelo trilingüe composto pelo conto 'Bliss', de Katherine Mansfield, por três de suas traduções para o português do Brasil e por três de suas traduções para o espanhol europeu. A apresentação do discurso diz respeito à fala, ao pensamento e à escrita, geralmente estudada em textos literários, embora este recurso possa estar presente em outros tipos de texto (SEMINO e SHORT, 2004). Nesta pesquisa, a apresentação do discurso é abordada sob a perspectiva da Estilística Lingüística, que leva em consideração a investigação de textos literários a partir dos estudos lingüísticos (HALLIDAY, 1964; PROCHÁZKA, 1964; HASAN, 1989). Examina-se a configuração de modos de apresentação do discurso em cada texto, tomando-se a oração e a oração menor (HALLIDAY e MATTHIESSEN, 2004) como unidades de análise para a anotação do corpus. O corpus de textos objeto deste trabalho foi digitalizado, revisado e anotado manualmente, em arquivo XML, com categorias de apresentação do discurso. A combinação deste arquivo XML com a folha de estilos possibilitou a quantificação de dados, visualizada em um arquivo HTML. Resultados apontam que há predileção pelo uso do pensamento e principalmente da fala mediante parataxe no corpus. Resultados mostram que as cinco categorias mais frequentes em 'Bliss' são: Fala Direta, Pensamento Indireto Livre, Fala Direta Livre, Narração e Narração Interna. As três categorias menos frequentes neste texto correspondem à Fala Indireta, Pensamento Direto Livre e Pensamento Indireto. As traduções para o português do Brasil de Érico Veríssimo, Julieta Cupertino e Ana Cristina Cesar e os textos alvo para o espanhol europeu de Lucía Graves e Elena Lambea e de Juani Guerra têm como categorias mais recorrentes as mesmas de 'Bliss'. A configuração das cinco categorias mais recorrentes difere na tradução para o espanhol europeu de Esther de Andreis, i.e., Fala Direta, Pensamento Indireto Livre, Narração, Narração Interna e Fala Indireta Livre são as categorias mais frequentes nesta tradução, demonstrando que este texto alvo apresenta mais ocorrências do Discurso Indireto Livre. A Fala Indireta, o Pensamento Direto Livre e o Pensamento Indireto são as três categorias menos recorrentes nas traduções para o português do Brasil de Érico Veríssimo, Julieta Cupertino e Ana Cristina Cesar e nos textos alvo para o espanhol europeu de

		<p>Lucía Graves e Elena Lambea e de Juani Guerra, conforme o texto fonte ('Bliss'). Na tradução para o espanhol europeu de Esther de Andreis, as três categorias menos frequentes são Pensamento Direto Livre, Fala Indireta e Pensamento Indireto. Resultados também apontam que o co-texto e o contexto podem motivar algumas das mudanças encontradas entre os textos alvo e o texto fonte no tocante à apresentação do discurso, como, por exemplo, a Fala Direta do texto fonte traduzida como Pensamento Direto no texto alvo de Julieta Cupertino para o português do Brasil; e a Fala Direta Livre do texto fonte textualizada como Pensamento Direto no texto alvo de Esther de Andreis para o espanhol europeu. Nas categorias Fala Indireta Livre, Pensamento Indireto Livre e Narração Interna, os substantivos comuns se repetem em maior grau; a forma de dêixis mais recorrente é a específica; e o tempo verbal preferencialmente usado é o Pretérito Imperfeito do Indicativo (Past Simple no texto fonte e Pretérito Imperfecto del Indicativo nos textos alvo para o espanhol europeu), corroborando estudos anteriores sobre o Discurso Indireto Livre.</p>
3.	2010	<p>Elaine Espindola. ILLUMINATED THE ANALYSIS OF THE TRANSLATION IS: SYSTEMIC FUNCTIONAL LINGUISTICS STRIKES YODA BACK. 01/09/2010</p> <p>1v. 174p. Doutorado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Christian M. I. M. Matthiessen; Maria Lúcia Barbosa de Vasconcellos</p> <p>Biblioteca Depositária: Biblioteca Central</p> <p>Email do autor:</p> <p>Palavras - chave:</p> <p>Estudos da Tradução Lingüístico-Sistêmico Funcional</p> <p>Área(s) do conhecimento:</p> <p>LINGÜÍSTICA, LETRAS E ARTES</p> <p>Banca examinadora:</p> <p>Célia Maria Magalhães</p> <p>Kazuhiro Teruya</p> <p>Roberto Carlos de Assis</p> <p>Viviane Maria Heberle</p> <p>Linha(s) de pesquisa:</p> <p>Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos</p>

		<p>(literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>CAPES - DS</p> <p>Idioma(s):</p> <p>Inglês</p> <p>Dependência administrativa</p> <p>Federal</p> <p>Resumo tese/dissertação:</p> <p>Na área de Estudos da Tradução (ET), pesquisa em tradução audiovisual tem abordado a tradução para filmes de uma forma que reforça a parcialidade que tem sido dada a esta atividade. Estes estudos parecem negligenciar a natureza lingüística da tradução de filmes concentrando-se principalmente em aspectos relacionados a fatores culturais da atividade da legendagem (Espindola, 2005), o ambiente audiovisual da legendagem (Nobre, 2002), aspectos técnicos envolvidos na tradução de filmes e programas de TV bem como fatores de audiência (Luyken et al., 1991 e Dries, 1995). Por mais importante que estes estudos sejam, eles deixam alguns aspectos intocados uma vez que omitem a dimensão lingüística das legendas. Devido ao fato da linguagem como um sistema modelador auxiliar na construção de significados no ambiente do filme – junto com outros tipos de sistemas semióticos, atenção direcionada à linguagem das legendas parece ser uma lacuna a ser preenchida. A partir de uma perspectiva lingüística, ET está na agenda de pesquisa de lingüística sistêmico-funcional há um longo tempo, porém somente recentemente é que Estudos da Tradução Audiovisual foram adicionados aos Estudos da Tradução Lingüístico-Sistêmico. Neste contexto e como tentativa de superar a parcialidade mencionada anteriormente, a presente pesquisa de doutorado dá atenção direta à linguagem dos diálogos falados e das legendas da Saga Guerra nas Estrelas (dirigida por George Lucas) concentrando esforços de pesquisa no estudo do personagem selecionado: Mestre Yoda, através da investigação da estrutura temática em nível oracional em ambas as instâncias, com o intuito de observar o papel de Temas marcados e não-marcados na construção do discurso do personagem. Os resultados obtidos evidenciam a contribuição da estrutura temática para conceber a representação de Yoda nos diálogos falados como um personagem sábio e poderoso da comunidade galáctica de Guerra nas Estrelas. Uma contribuição similar foi</p>
--	--	---

		<p>encontrada nas legendas onde a estrutura temática provou desempenhar papel crucial na construção do comportamento lingüístico de Yoda, indicativo da natureza de relação de poder existente entre ele e os outros personagens da Saga. Em ambos os casos, o discurso marcado emergiu como uma característica constante com Complemento e Predicador ocorrendo mais freqüentemente em posição temática. Tipicamente, quando o Tema em inglês era realizado por um Predicador sozinho com o finito em posição remática para marcar o discurso, um Complemento era usado para causar o mesmo efeito em resposta as possibilidades permitidas pelo sistema lingüístico do português brasileiro. Finalmente, a necessidade de se dar atenção direta a complexidade lingüística das legendas tornou-se evidente, uma vez que este tipo de análise provou ser possível de iluminar aspectos outrora desconsiderados – a análise lingüístico-sistêmico funcional levou à descoberta do linguajar característico de Yoda que media uma visão de mundo distinta.</p>
4.	2010	<p>Leonardo Pereira Nunes. As conjunções but e mas em textos ficcionais originais e traduzidos: uma análise tridimensional com base na linguística sistêmico-funcional. 01/03/2010 1v. 103p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: Biblioteca Universitária da UFMG</p> <p>Email do autor:</p> <p>Palavras - chave: Conjunção externa e interna, corpus paralelo e comparável</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Célia Maria Magalhães Paulo Henrique Caetano Silvana Maria de Jesus</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - PROF</p>

		<p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta dissertação apresenta uma análise das conjunções but e mas em um corpus paralelo bilíngue e comparável monolíngue composto por textos ficcionais originalmente produzidos em inglês, suas traduções para o português brasileiro e textos originalmente escritos nesta língua. Com o propósito de contribuir para uma descrição sistêmico-funcional de mas a partir de uma perspectiva interlinguística e comparada, este estudo incorpora contribuições de Halliday e Hasan (1976) com relação ao sistema de CONJUNÇÃO construído através de significados ideacionais (relações externas) e interpessoais (relações internas). Desenvolvendo esta dualidade, promoveu-se uma separação sistemática das conjunções em externas e internas segundo categorias propostas por Martin e Rose (2003) e acentuou-se a diferença entre a natureza textual e interpessoal das relações internas de acordo com variáveis sugeridas por Thompson (2005). Desta forma, as relações conjuntivas foram assim abordadas por um prisma tridimensional (ideacional, textual e interpessoal). A partir deste arcabouço teórico, investigou-se o potencial de construção de relações externas ideacionais e internas textuais e interpessoais pela conjunção but em dez romances originais em inglês e pela conjunção mas em suas traduções ao português brasileiro. A mesma indagação foi feita considerando-se a conjunção mas em dez romances originais em português brasileiro. Os dados quantitativos apresentam ocorrências expressivas de but nos textos fonte e mas nos textos alvo em uma relação (i) adversativa/concessiva no nível ideacional, (ii) de adição, consequência-conclusão e consequência-contraposição no nível textual e (iii) de negação-contraposição e concessão-asserção no nível interpessoal. Observou-se este mesmo padrão no tocante a conjunção mas nos textos originais em português brasileiro, com exceção da relação de consequência-conclusão no nível textual. Concluiu-se então que, apesar desta diferença, a conjunção mas nesta língua possui, assim como but no inglês, similar potencial de construir significado interno textual e interpessoal em textos ficcionais.</p>
5.	2009	<p>Marcos Pereira Feitosa. Legendagem comercial e legendagem pirata: um estudo comparado. 01/12/2009</p>

	<p>1v. 162p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Adriana Silvina Pagano</p> <p>Biblioteca Depositária: Bilbiloteca Universitária da UFMG</p> <p>Email do autor:</p> <p>Palavras - chave: CROSF+ EFI; Fluxo da Informação; Explicitação</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Carlos Alberto Gohn Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos Paulo Henrique Caetano Vera Lúcia Santiago Araújo</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Este trabalho situa-se no campo disciplinar dos Estudos da Tradução, tendo como abordagens a Lingüística Sistemico-Funcional (Gramática Sistemico-Funcional de M.A.K. Halliday) e a lingüística de Corpus e é desenvolvido no âmbito do Núcleo de Estudos da Tradução (NET) da Faculdade de Letras (FALE) da Universidade Federal de Minas Gerais (UFMG). Visando uma anotação de corpora mais eficiente e menos propensa a erros de digitação, os quais interferem nos dados de pesquisas, propõe-se aqui um modelo de anotação instrumentalizado através de um código numérico para a Rotulação de corpora com base na</p>
--	--

		<p>Gramática Sistemico-Funcional de Halliday. Esse código, denominado ζCROSFζ (Código de Rotulação Sistemico-Funcional), é desenvolvido através de sucessivos protótipos testados com um corpus paralelo bilíngüe composto de um fragmento do romance <i>The blind assassin</i> de Margaret Atwood (2000) em inglês canadense e sua tradução para o português brasileiro feita por Lea Viveiros de Castro (Atwood, 2001). O estudo de caso aplicando-se a elipse do sujeito em português, bem como escolhas idiossincráticas da tradutora. O código numérico facilitou as buscas no corpus eletrônico, permitindo maior visibilidade do texto anotado e buscas de diferentes combinações de escolhas temáticas</p>
6.	2009	<p>Alinne Balduino Pires Fernandes. Black into White and Preto no Branco: Can You Tell One's Colour by the Company One Keeps?. 01/04/2009</p> <p>1v. 127p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Maria Lúcia Barbosa de Vasconcelos Biblioteca Depositária: Biblioteca Universitária - UFSC</p> <p>Email do autor:</p> <p>Palavras - chave: Functional Translation Studies (SFTS);</p> <p>Área(s) do conhecimento: LETRAS</p> <p>Banca examinadora: Célia Maria Magalhães Fernando Simão Vugman Meta Elisabeth Zipser</p> <p>Linha(s) de pesquisa:</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação:</p>

		<p>This Thesis was carried out at the interface between Corpus-based Translation Studies (CTS) and Systemic Functional Translation Studies (SFTS). CTS informs the methodological apparatus which framed this Thesis. SFTS provides conceptual tools to look at how the Brazilian entity mestiço is construed in North American English in the textualisation (T) Black into White, and how it is construed in Brazilian Portuguese in the re-textualisation (RT) Preto no Branco by means of the categories of the experiential component of the ideational metafunction of language, and by means of the categories of lexical cohesion. The book was written by the Brazil-specialist Thomas Skidmore, and published for the first time in 1974. In 1976, the book was published in Brazil translated by the retired Brazilian diplomat Raul de Sá Barbosa. The book, which proposes to describe Brazilian racial thought, narrates Brazilian political and social turnabouts from the abolition of slavery until the late thirties of the twentieth century. In this vein, the research questions which guide this Thesis are: (i) How are the mestiços construed in terms of processes in the ideational metafunction of language both in BW (T) and PB (RT)? Do ideational patterns arise in the two representations? If so, which? (ii) Is the cohesive chain of the T different from that of the RT? (iii) Do the T and RT reaffirm the racist thought of 1880-1930? T and RT do not diverge considerably in terms of ideational constructions, though interesting patterns of new language construction have arisen in the RT. Predominant material and relational processes seem to realise the mestiço in agreement with racist thought that influenced Brazilian political posture from late nineteenth to late twentieth centuries. The RT lexical cohesive chain follows pretty much the same patterns of the T.</p>
7.	2009	<p>Cristina Mauri. Uma análise do ponto de vista em A Hora da Estrela e Laços de Família, de Clarice Lispector, e nas traduções italianas L'ora Della Stella e Legami Familiari. 01/03/2009</p> <p>1v. 260p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Célia Maria Magalhães Biblioteca Depositaria: Biblioteca Universitária da UFMG</p> <p>Email do autor:</p> <p>Palavras - chave: Corpora; traduções; pontos de vista.</p> <p>Área(s) do conhecimento:</p>

	<p>LINGÜÍSTICA LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcelos Patrizia G. E. Collina Bastianetto</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta tese está afiliada à subárea de abordagens textuais da tradução, que integra o projeto Corpus Discursivo para Análises Lingüísticas e Literárias (CORDIALL), desenvolvido no Laboratório Experimental de Tradução (LETRA) da Universidade Federal de Minas Gerais (UFMG). Tal subárea do CORDIALL compreende trabalhos cujo foco principal é o estudo da linguagem da tradução, e, mais especificamente, a construção de personagens e a apresentação do discurso em textos ficcionais. Continuando essa tradição, o presente trabalho aborda o ponto de vista narrativo na tradução a partir da investigação de um corpus paralelo constituído pelo romance A hora da estrela (AHDE) e pela coletânea de contos Laços de família (LDF), ambos de Clarice Lispector, e suas respectivas traduções italianas L'ora della stella (LODS) e Legami familiari (LF), realizadas por Adelina Aletti. Visando contribuir para os estudos do ponto de vista na tradução, foram estabelecidos três objetivos: (i) analisar o ponto de vista narrativo nos textos-fonte (TFs) para detectar padrões de escrita da autora no romance e nos contos; (ii) observar padrões de escrita no que diz respeito à construção do ponto de vista narrativo nos textos-alvo (TAs); e (iii) averiguar as mudanças entre os TFs e os TAs com relação à construção do ponto de vista. O</p>
--	---

		<p>referencial teórico adotado na investigação são estudos sobre a tradução do ponto de vista narrativo e, como pressuposto destes, trabalhos de estilística e de apresentação do discurso. Recorre-se ainda à fortuna crítica de Lispector para observar se são apontados traços do ponto de vista como características típicas da escrita clariceana. Foram também utilizados recursos descritivos da Gramática Sistemico-Funcional (GSF) como suporte na observação de algumas categorias de análise e ferramentas da Linguística de Corpus na metodologia de análise. Para cumprir os objetivos de pesquisa, foram analisados quatro realizadores lingüísticos da focalização narrativa, a saber: dêixis, discurso indireto livre, Processos materiais e modalidade. Os resultados da análise apontam padrões de escrita similares nos pares AHDE-LDF e LODS-LF, porém com algumas diferenças em termos de frequência das ocorrências em cada categoria. As mudanças observadas nas traduções se referem a todas as quatro categorias de análise em ambos os TAs, embora algumas mudanças específicas tenham sido detectadas apenas em LF. Foi constatado que tais mudanças acarretam diferenças na construção do ponto de vista narrativo, especialmente no que diz respeito à focalização nos personagens. Ao mesmo tempo, a análise contrastiva entre os TFs e os TAs apontou questões que não constavam nos objetivos iniciais da investigação, como a provável explicitação por meio do uso de exclamações e interrogações. Palavras-chave: abordagens textuais da tradução; estudos da tradução baseados em corpus; ponto de vista na tradução; Linguística Sistemico-Funcional; traduções italianas da obra de Clarice Lispector.</p>
8.	2009	<p>Roberto Carlos de Assis. A representação de europeus e de africanos como atores sociais em Heart of Darkness (O coração das trevas) e em suas traduções para o português: uma abordagem textual da tradução. 01/03/2009</p> <p>1v. 254p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Célia Maria Magalhães</p> <p>Biblioteca Depositária: Biblioteca Universitária da UFMG</p> <p>Email do autor:</p> <p>Palavras - chave:</p> <p>Abordagens Textuais da tradução de viés sistemico-funcional</p> <p>Área(s) do conhecimento:</p> <p>LINGÜÍSTICA</p>

		<p style="text-align: center;">LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Diva Cardoso de Camargo Leila Barbara Maria Lúcia Barbosa de Vasconcellos Paulo Henrique Caetano</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS CNPq</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Pesquisas desenvolvidas no âmbito de um dos diversos projetos do LETRA - Laboratório Experiencial de Tradução vêm se consolidando no contexto nacional e internacional dos Estudos da Tradução ao descreverem padrões de registro de textos em relação tradutória, através do aparato de descrição textual oferecido pela Linguística Sistemico-Funcional. Além de dar continuidade aos estudos desenvolvidos pelos pesquisadores do LETRA, esta tese avança ao propor a descrição de aspectos sociosemânticos dos textos e ao estabelecer uma interface dos Estudos da Tradução com a Semiótica Social, por meio da comparação das representações de atores sociais em um corpus paralelo. Este é composto pelo romance Heart of Darkness, escrito por Joseph Conrad em 1899, e duas de suas traduções para o português do Brasil publicadas em 1984 e em 2002. Esta obra tornou-se polêmica, especialmente após a forte crítica de Chinua Achebe ([1977] [1988] 2006), que se referiu a Conrad como um "perfeito racista" (thoroughgoing racist) dado o tratamento dispensado aos africanos no romance. Tal afirmação repercutiu entre os críticos literários e de outras</p>
--	--	--

abordagens, no entanto, recebeu pouca atenção por parte de pesquisadores afiliados às abordagens lingüísticas, sobretudo na situação de línguas e culturas em contato, como acontece no caso de textos em relação de tradução. Abordando a tradução através de uma abordagem textual de viés sistêmico-funcional, para descrição dos textos, utiliza-se como referencial a teoria de representação de atores sociais (VAN LEEUWEN, 1993 e 1996). A investigação procura responder à perguntas relacionadas à realização das categorias sociosemânticas de representação de atores sociais em português, à representação de europeus e de africanos no texto de partida e nos de chegada e às formas como os tradutores lidaram com as polêmicas em torno do texto em inglês. Como passos metodológicos, o corpus foi alinhado, anotado e explorado com o auxílio de um programa de computador, o WordSmith Tools®; foi feito um recorte no inventário sociosemântico de van-Leeuwen (1996), concentrando-se em suas categorias de Personalização/Impersonalização e Ativação/Apassivação. Os resultados da pesquisa apontaram para a necessidade de expansão das categorias existentes e, entre as propostas, estão a Sistemização e a Transfiguração. A primeira relaciona-se às instâncias em que os atores sociais não são ativados ou apassivados, e a segunda, àquelas em que o traço +humano é subtraído do ator social através de recursos adversos da abstração ou da referência metonímica. Europeus e africanos têm representações desfavoráveis no texto de partida, sendo a dos africanos ainda mais desfavorável. Enquanto os primeiros são principalmente personalizados, funcionalizados, nomeados e classificados, os últimos são impersonalizados, classificados, primitivizados e somatizados. Essas, associadas a outras formas de representação, pode ser uma indicação do estabelecimento de uma linha divisória entre os dois grupos, em que os representantes (o europeus) parecem se servirem dessas estratégias para criar distanciamento e retirar a humanidade dos africanos, como forma de legitimação da subjugação. As análises das realizações lingüísticas das traduções revelaram uma semelhança na forma de representação dos dois grupos quando comparados ao texto de partida, ou seja, nas traduções os africanos tiveram uma representação mais desfavorável do que a representação dos europeus. Além de algumas mudanças no ambiente mais amplo, como a permutabilidade entre as categorias, que alteram as representações, a análise das escolhas dos tradutores para a tradução de alguns itens lexicais revelaram uma intensificação do uso de termos com conteúdos

		<p>interpessoais para se referir tanto aos europeus quanto aos africanos. Registrou-se também que a escolha de itens socialmente aceitos para a tradução de itens insultantes aos africanos, presentes no texto de partida, dilui, nas traduções, as questões de racismo apontadas pela crítica literária em inglês. Esta tese, apesar de suas limitações, constitui-se em uma contribuição para os Estudos da Tradução, para a Semiótica Social e para os estudos conradianos.</p>
9.	2009	<p>Thiago Blanch Pires. THE CONSTRUAL OF BISHOP'S IDEATIONAL PROFILE IN FLORES RARAS E BANALÍSSIMAS AND RARE AND COMMONPLACE FLOWERS: A CORPUS-BASED TRANSLATION STUDY. 01/12/2009</p> <p>1v. 102p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Maria Lúcia Barbosa de Vasconcelos</p> <p>Biblioteca Depositária: Biblioteca Central</p> <p>Email do autor:</p> <p>Palavras - chave: Flores Raras e Banalíssimas; Rare and Commonplace Flowers.</p> <p>Área(s) do conhecimento:</p> <p>Banca examinadora: Gloria Gil Lincoln Paulo Fernandes</p> <p>Linha(s) de pesquisa:</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta pesquisa, realizada na interface entre a Linguística Sistêmico-Funcional, os Estudos da Tradução e as metodologias dos Estudos baseados em Corpus, investiga o perfil Ideacional de Elizabeth Bishop em dois textos em relação tradutória: Flores Raras e Banalíssimas (Oliveira,</p>

		<p>1995) e Rare and Commonplace Flowers (trad. Besner, 2002). O objetivo da pesquisa, portanto, é examinar, por meio das categorias do Sistema da Transitividade, em seu componente experiencial, o Participante 'Bishop' e os Processos nos quais ele está inscrito. Analisa-se o padrão do uso do item lexical "Bishop" para observar que tipo de Participante é realizado pelos itens lexicais a ele relacionados dentro do complexo oracional e como esse Participante pode ser associado a representações da poetisa americana na textualização (texto em português brasileiro) e na retextualização (texto em inglês americano). Os métodos utilizados nessa investigação foram divididos em (i) compilação do corpus (desenho, construção e processamento) com o auxílio da suíte de programas WordSmith Tools (Scott, 1999) e (ii) análise manual do corpus complementando a análise automatizada na abordagem da gramática de Processos e Participantes. Embora os resultados colhidos na análise quantitativa mostrem que há semelhanças nos padrões de transitividade em ambos os textos (Bishop é construída enquanto Participante ativo envolvido em Processos Materiais (43%)), novas configurações de linguagem emergem na análise qualitativa. Em 10 casos, o Participante Bishop é representado como Experienciador na textualização e encontra-se em relações de atribuição (Portador) na retextualização. A escolha de Bishop como Participante Portador no texto em inglês americano constrói uma representação associada a Processos Relacionais de ser e estar, construindo um Participante passivo no contexto de chegada.</p>
10.	2008	<p>Silvana Maria de Jesus. Relações de tradução: SAY/DIZER em corpora de textos ficcionais.. 01/06/2008 1v. 211p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: FALE; BU</p> <p>Email do autor:</p> <p>Palavras - chave: textos ficcionais; corpus combinado; equivalência; gramática</p> <p>Área(s) do conhecimento: LINGÜÍSTICA</p> <p>Banca examinadora: Adriana Silvina Pagano</p>

	<p>Célia Maria Magalhães Fabio Alves da Silva Junior Maria Carmen Dayrell Gomes da Costa Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - PROF</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta tese insere-se nos estudos sistêmico-funcionais da tradução e aborda as relações de tradução de SAY/DIZER em textos ficcionais no par lingüístico inglês-português. Adotando uma perspectiva empírica de observação de dados em corpora paralelos e comparáveis, este trabalho desenvolve três abordagens de análise. A primeira abordagem, a partir da fundamentação teórica da lingüística sistêmico-funcional (Halliday e Matthiessen, 2004), descreve as funções das orações verbais realizadas por SAY/DIZER em textos ficcionais não-traduzidos, isto é, em textos originalmente escritos em inglês e português, respectivamente. A segunda abordagem, a partir da proposta de Catford (1965) de análise das probabilidades condicionadas e incondicionadas de tradução, observa as relações de tradução de SAY/DIZER em corpora paralelos bidirecionais constituídos por textos originais e suas respectivas traduções no par lingüístico inglês-português. A terceira abordagem investiga as relações de tradução de SAY/DIZER sob a perspectiva de análise das propriedades de textos traduzidos, segundo a proposta de Teich (2003). O corpus da pesquisa, 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, faz parte do CORDIAL (Corpus Discursivo para Análises Lingüísticas e Literárias) desenvolvido pelos pesquisadores do LETRA (Laboratório Experimental de Tradução) da</p>
--	--

		<p>Faculdade de Letras da UFMG. A metodologia da lingüística de corpus é aplicada na análise deste corpus, denominado corpus combinado, que, devido as suas características, amplia o campo de observação do pesquisador, gerando dados mais abrangentes. Os programas WordSmith Tools e SPSS (Statistical Package for the Social Sciences) são utilizados para a extração de dados quantitativos e estatísticos e para a construção do banco de dados da pesquisa. Os resultados apresentam uma visão abrangente das relações de tradução de SAY/DIZER em textos ficcionais. Verificou-se que as orações verbais com SAY/DIZER realizam funções experienciais e interpessoais na introdução do discurso, apresentando padrões distintos no inglês e no português. Foram apresentados 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. Por fim, a análise aponta resultados que se mostram produtivos para a descrição sistêmico-funcional do português e para os estudos da tradução.</p>
11.	2007	<p>Giacomo Patrocínio Figueiredo. Uma descrição sistêmico-funcional da estrutura do grupo nominal em português orientada para os estudos lingüísticos da tradução. 01/07/2007 1v. 291p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: tradução, descrição do português, grupo nominal</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcelos</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p>

	<p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Diante da necessidade apontada pelas pelos estudos sistêmico-funcionais aplicados à língua portuguesa, qual seja, a descrição sistêmico-funcional do português como instrumento de análise e estudo da tradução, esta dissertação adota uma visão lingüístico-funcional da tradução (TEICH, 1999; MATTHIESSEN, 2001), na qual a tradução é entendida a partir da dimensão da instanciamento como uma relação entre sistemas lingüísticos e a descrição lingüística se constitui como uma das etapas dos estudos tradutórios no contexto desta visão. Objetivando promover a identificação da ordem do grupo em português e a descrição dos sistemas componentes das estruturas experiencial e lógica da classe do grupo nominal (cf. HALLIDAY, 2002), esta dissertação tomou como base os processos sócio-semióticos (cf. COUCHMAN, 2006), para a compilação de um corpus (com um total de 12.000 palavras distribuídas em oito subcorpórea de 1.500 palavras, cada qual relativo a um dos oito processos sócio-semióticos) no qual as classes de grupos foram anotadas segundo suas funções, bem como os itens na ordem da palavra que serviram de elementos na estrutura dos grupos. Destas anotações foram retirados os padrões de funções destes elementos. Para a realização da análise, adotou-se como ponto de partida para a descrição a dimensão da estrutura. O exame “de baixo” revelou a forma como os elementos da ordem inferior se agrupam, formando, assim, classes. “De cima” foi possível observar quais são as classes de grupos que operam como elemento na estrutura da oração. Tal análise indica que são as ordens em português: morfema, palavra, grupo (e frase) e oração. Para a descrição da estruturação do grupo nominal, este foi examinado “de baixo” e nele foram identificadas as classes de palavras que servem como elemento na ordem do grupo (nominal). Examinando o grupo nominal “ao redor”, descreveu-se a forma como as classes de seus elementos componentes se organizam. Os resultados alcançados indicam que o grupo nominal em português se estrutura de forma experiencial e lógica. Experiencialmente, os elementos presentes nesta classe de grupo cumprem as</p>
--	--

		<p>funções de Qualidade: Dêiticos: não-seletivo (específico e não-específico), seletivo de proximidade, seletivo de pessoa, indefinido, interrogativo. Numerativo: ordenativo, quantitativo, interrogativo. Epíteto: experiencial, interpessoal. Classificador: várias classes (material, origem, característica, etc.). Além destas, há presente também a função do Ente, que pode ser classificado taxonomicamente como: consciente, animal, objeto material, substância, abstração material, instituição, objeto semiótico, abstração semiótica. Na estrutura lógica, o sistema lingüístico do português possui um elemento operando como Núcleo e outros elementos operando como Pré- e Pós-modificadores. Foi possível também observar que a posição de pós-modificação exerce dominância em relação de pré-modificação, permitindo que os elementos Pós-modificadores sejam modificados pelos Pré-modificadores. Em seguida, promoveu-se a comparação e o contraste entre o português e um outro sistema lingüístico que, no caso desta pesquisa, a seleção foi pelo inglês. Corroborando os estudos de Matthiessen (2001), Teich (2001) e Steiner (2002), esta descrição provou ser útil para o contraste tipológico com outros sistemas lingüísticos, podendo ser capaz de contribuir para análises futuras pautadas pela visão lingüístico-funcional da tradução.</p>
12.	2007	<p>Cristiano Gonçalves Araújo. O sistema semântico de PROJEÇÃO e sua dispersão gramatical em português brasileiro: uma descrição sistêmico-funcional orientada para os estudos lingüísticos da tradução. 01/07/2007 1v. 133p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: tradução, lingüística sistêmico-funcional</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa:</p>

		<p>ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: O presente estudo assume a lingüística sistêmico-funcional como parte integrante de um modelo de abordagem à tradução, e busca contribuir para este modelo a partir de um enfoque específico, qual seja, o da descrição do sistema semântico de PROJEÇÃO e de sua dispersão gramatical em português brasileiro. Baseando-se no trabalho de Halliday e Matthiessen (2004), e em outros trabalhos fundamentados na teoria sistêmica, este estudo aborda a descrição da projeção de locuções e idéias, tanto em relação aos modos de relato e citação quanto em relação à função discursiva (se proposta ou proposição) das orações projetadas. Para se proceder à descrição proposta neste trabalho, foi montado um corpus composto por textos em português brasileiro pertencentes a diferentes domínios de registros, somando-se cerca de 30.000 tokens. Este corpus foi anotado e os dados foram coletados com o auxílio do programa WordSmith Tools. A partir dos dados obtidos foram descritos padrões de ocorrência que possibilitaram caracterizar a configuração do sistema semântico de PROJEÇÃO em português brasileiro. Os resultados obtidos com esta análise podem informar outros estudos discursivos e contrastivos que, adotando os subsídios teóricos da teoria sistêmico-funcional de linguagem, tenham por interesse a tradução e a produção multilíngüe de textos.</p>
13.	2007	<p>Joanna de Paula Filgueiras. Capoeira em tradução: representações discursivas em um corpus paralelo bilíngüe. 01/12/2007</p> <p>1v. 111p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - ESTUDOS DA TRADUÇÃO Orientador(es): Maria Lúcia Barbosa de Vasconcelos Biblioteca Depositária: Biblioteca Universitária</p>

	<p>Email do autor:</p> <p>Palavras - chave: Abordagens discursivas as textos de tradução; linguística.</p> <p>Área(s) do conhecimento: LINGÜÍSTICA, LETRAS E ARTES</p> <p>Banca examinadora: Andréia Guerini Lincoln Fernandes Marco Antonio E da Rocha Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: Teoria, crítica e história da tradução O fenômeno tradutório em suas especificidades culturais, históricas e ideológicas. Abordagens teóricas, críticas e analíticas de tradução e diferentes propostas de estudo dos percursos históricos da tradução. Exame de corpora de textos traduzidos.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Este estudo integra o Projeto Interinstitucional UFMG/UFSC intitulado Corpora, Cognição e Discurso: uma proposta interdisciplinar para os Estudos da Tradução a partir de bancos eletrônicos de dados. A proposta desta dissertação é identificar representações discursivas sobre a capoeira a partir da análise do perfil da Prosódia Semântica dos itens lexicais capoeir*, malícia, mandinga e malandr* no corpus paralelo bilíngüe composto pelos livros: Capoeira. Pequeno Manual do Jogador (2002), de Nestor Capoeira e sua re-textualização The little capoeira book (2003), feita por Alex Ladd. Para fins dessa pesquisa, são consideradas palavras-chaves os vocábulos que expressam idéias e valores no corpus de estudo (conforme Williams, 1976) e que ocorrem com frequência muito alta ou baixa quando comparado a um corpus de referência (com auxílio da ferramenta keywords do software WS Tools). A pesquisa é desenvolvida a partir das Abordagens Discursivas aos</p>
--	--

		<p>Estudos da Tradução (ADET), com enfoque na noção na linguagem enquanto semiótica social, conforme a Linguística Sistemico Funcional (LSF), e no fenômeno colocacional - com exploração do conceito de Prosódia Semântica (PS) - identificado no corpus em estudo a partir de ferramentas computacionais. Os resultados finais apontam para diferenças na observação do fenômeno colocacional, que influenciam no perfil da PS dos itens lexicais analisados - na textualização (o texto de partida) e a re-textualização (o texto de chegada) - e mostram a importância de se relacionar os dados obtidos eletronicamente aos contextos em que as obras analisadas estão inseridas.</p>
14.	2006	<p>Lilian Jurkevicz Fleuri. O perfil ideacional dos itens lexicais translator/tradutor em "Translators. 01/02/2006 1v. 177p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - ESTUDOS DA TRADUÇÃO Orientador(es): Maria Lúcia Barbosa de Vasconcellos Biblioteca Depositária: Biblioteca da UFSC</p> <p>Email do autor:</p> <p>Palavras - chave: Abordagem discursivas aos Estudos da Tradução.</p> <p>Área(s) do conhecimento: LETRAS</p> <p>Banca examinadora: Célia Magalhães Marco Antonio E da Rocha Maria Lúcia Barbosa de Vasconcellos Marie Helene Catherine Torres</p> <p>Linha(s) de pesquisa: Lexicografia, tradução e ensino de línguas Estudo e análise de ferramentas lexicográficas (dicionário, recursos informatizados); pesquisa do seu aprimoramento para o processo tradutório e aprendizagem de uma língua estrangeira. Relação entre aprendizagem e uso dos instrumentos lexicográficos.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Outro</p> <p>Dependência administrativa</p>

		<p>Federal</p> <p>Resumo tese/dissertação:</p> <p>Esta pesquisa parte dos conceitos de linguagem e de texto desenvolvidos pela Lingüística Sistêmico-Funcional (LSF) que interpreta a língua como um sistema modelador de realidades. Partindo deste conceito, sugere-se, nesta pesquisa de caráter interdisciplinar, a interface entre a Lingüística Sistêmico-Funcional, os Estudos da Tradução e a Lingüística de Corpus para análise descritiva da obra Translators Through History e sua tradução Os Tradutores na História. Neste estudo busca-se examinar o perfil ideacional dos itens lexicais "translator/tradutor" em ambos os textos, identificando-se os padrões emergentes na textualização e na retextualização e os novos significados construídos no processo de retextualização. Observa-se que, apesar de os padrões de transitividade serem análogos em ambos os textos - os resultados da análise apontam para uma construção similar do "tradutor" - há especificidades no perfil ideacional do tradutor sendo tal item lexical mais representado como Portador na textualização (texto em inglês) e mais como Ator, na retextualização (texto em português), configurando-se, portanto, em uma representação que enfatiza um papel ativo do tradutor no contexto de chegada.</p>
15.	2006	<p>Gleimara Regina Ferreira Zuniga. CONSTRUING THE TRANSLATOR: A META-REFLECTION GROUNDED IN CORPUS-BASED TRANSLATION STUDIES AND SYSTEMIC FUNCTIONAL LINGUISTICS. 01/02/2006</p> <p>1v. 114p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Lincoln Paulo Fernandes; Maria Lúcia Barbosa de Vasconcellos</p> <p>Biblioteca Depositária: Biblioteca Central da UFSC</p> <p>Email do autor:</p> <p>Palavras - chave:</p> <p>Ideational profile, translator corpus-based.</p> <p>Área(s) do conhecimento:</p> <p>LINGÜÍSTICA APLICADA</p> <p>Banca examinadora:</p> <p>Adriana Silvina Pagano</p> <p>Lincoln Paulo Fernandes</p> <p>Maria Lúcia Barbosa de Vasconcellos</p>

		<p>Markus Johannes Weininger</p> <p>Linha(s) de pesquisa: Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: A natureza interdisciplinar de Estudos da Tradução permite que diferentes abordagens investiguem o fenômeno da tradução. Neste sentido, este estudo, informado por estudos de corpus e lingüística sistêmica, se propõe a investigar como a entidade “o tradutor” é construída em um corpus paralelo de pequena dimensão composto por <i>Becoming a Translator</i> (1997), um livro acadêmico textualizado por Douglas Robinson e o seu correlativo <i>Construindo o Tradutor</i> (2002), retextualizado por Jussara Simões, através da investigação de padrões de transitividade emergentes na modelagem do tradutor nos dois textos. A análise é assistida pelo programa de computador WordSmith Tools (Scott, 1998) para coleta de dados, e pelo código numérico para anotação de corpus CROSF 14 (Feitosa, 2005). A análise das escolhas de transitividade na linguagem usada para construir a imagem de “o tradutor” mostrou que tal entidade foi quantitativamente e proporcionalmente retextualizada. No entanto, a análise qualitativa revelou uma ligeira divergência no que diz respeito à construção da imagem de “o tradutor” na retextualização.</p>
16.	2006	<p>Ladjane Maria Farias de Souza. A licação do modelo de Linguagem Avaliativa à Tradução: Análise de um Texto Seminal da Filosofia do Software Livre.. 01/02/2006</p> <p>1v. 241p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - ESTUDOS DA TRADUÇÃO</p> <p>Orientador(es): Markus Johannes Weininger</p> <p>Biblioteca Depositária: Biblioteca Universitária</p> <p>Email do autor:</p>

	<p>Palavras - chave: Modelo de Linguagem Avaliativa (Appraisal Framework),</p> <p>Área(s) do conhecimento: LINGÜÍSTICA, LETRAS E ARTES</p> <p>Banca examinadora: José Luiz Meurer Lincoln Paulo Fernandes Markus Johannes Weininger Werner Ludger Heidermann</p> <p>Linha(s) de pesquisa: Lexicografia, tradução e ensino de línguas Estudo e análise de ferramentas lexicográficas (dicionário, recursos informatizados); pesquisa do seu aprimoramento para o processo tradutório e aprendizagem de uma língua estrangeira. Relação entre aprendizagem e uso dos instrumentos lexicográficos.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: RESUMO Esta dissertação descreve e compara o uso da linguagem avaliativa em dois textos argumentativos, um sendo a tradução do outro, tomando como mapa de referência o Modelo de Linguagem Avaliativa, uma abordagem lingüística desenvolvida no âmbito da Lingüística Sistemico-Funcional para descrever o modo pelo qual a língua é usada para se adotar pontos-de-vista, construir personae textuais e negociar relações interpessoais. Para melhor situar a dimensão interpessoal em cada texto, a pesquisa leva em consideração não apenas diferenças de significado concernentes ao sistema de Avaliação, mas também aos sistemas de Negociação e Envolvimento, além de diferenças relativas às semânticas ideacional e textual. São realizados 3 tipos de análise: uma quantitativa, na qual as diferenças de significado são classificadas e contabilizadas; uma análise da interação dos recursos de Avaliação na produção de significados interpessoais; e uma análise da construção de personae textuais. Tais análises mostram que as diferenças de Avaliação entre os textos</p>
--	---

		<p>constituem a maioria das diferenças de significado Interpessoal, caracterizando-se como diferenças de Comprometimento, de Atitude e de Gradação. Mostram também que elas interagem com os outros tipos de diferença, interferindo na relação interpessoal construída em cada texto, visto que contribuem para a construção de diferentes personae textuais e diferentes tipos de apelo ao leitor. O TC reduz o impacto do texto sobre o leitor, dada a inconsistência do posicionamento da voz textual com relação aos valores negociados com o leitor. A pesquisa demonstra, assim, que o MLA pode ser usado de modo produtivo na análise do texto traduzido.</p>
17.	2006	<p>Ariana Zanella. MAPEAMENTO MACRO E MICRO ESTRUTURAL DA RETEXTUALIZAÇÃO DE RESUMOS ON-LINE: Estudo da transitividade de abstracts biomédicos.. 01/01/2006</p> <p>1v. 110p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - ESTUDOS DA TRADUÇÃO Orientador(es): Ronaldo Lima Biblioteca Depositaria: Biblioteca Universitaria</p> <p>Email do autor:</p> <p>Palavras - chave: Tradução, organização retórica.</p> <p>Área(s) do conhecimento: LETRAS</p> <p>Banca examinadora: Adriana Kuerten Dellagelo Ana Cláudia de Souza Ronaldo Lima Viviane Herble</p> <p>Linha(s) de pesquisa: Lexicografia, tradução e ensino de línguas Estudo e análise de ferramentas lexicográficas (dicionário, recursos informatizados); pesquisa do seu aprimoramento para o processo tradutório e aprendizagem de uma língua estrangeira. Relação entre aprendizagem e uso dos instrumentos lexicográficos.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Outro</p>

		<p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação:</p> <p>Os resumos como gênero acadêmico impresso têm sido foco de vários estudos. No entanto, há uma carência de pesquisas que focalizem os resumos de periódicos indexados e principalmente sua retextualização (versões) para a língua de chegada (abstract), por representarem uma possibilidade de comunicação internacional rápida e eficaz na pesquisa acadêmica. Esta dissertação permeia os estudos da tradução em interface com a lingüística sistêmico-funcional. Com o objetivo de compreender como a retextualização de resumos da área biomédica se realiza, a macro e a micro estrutura foram quantitativamente investigadas a cada estágio retórico: introdução, objetivo, método, resultado e conclusão. O propósito foi identificar como a linguagem está em uso através de significados ideacionais/experienciais do sistema de transitividade (participantes e processos de experiências externas do mundo: materiais, relacionais, verbais, e processos de experiências internas: mental, existencial e comportamental) realizados nos resumos e em suas retextualizações. Os corpora de 15 resumos e respectivos abstracts foram coletados aleatoriamente no sítio dos periódicos CAPES, e inseridos no software SYSCODER. A seguir, cada corpus foi segmentado e codificado para a análise estatística comparativa. Quanto à macroestrutura, os resultados sugeriram que a comunidade discursiva biomédica faz uso de estágios estruturados de forma prescritiva. Por sua vez, os resultados da análise do sistema de transitividade demonstraram que resumos são compostos pelas características externas como normas comunicativas desenvolvidas pela atividade profissional que afetam as características internas ao texto. Além disso, demonstrou diferenças significativas de tipos de processos na passagem para o texto de chegada. Em conclusão, este estudo sugere que o português, língua-mãe de textualização, tem sofrido modificações léxico-gramaticais devido ao contexto situacional, aproximando-se cada vez mais do inglês, língua franca de pesquisa acadêmica, conseqüentemente construindo novas realidades lingüísticas. Por fim, o software SYSCODER é recomendado como ferramenta eficaz e precisa, também se salienta a importância da gramática sistêmico-funcional como ferramenta de análise de contextos mais abstratos da língua.</p>
18.	2006	<p>Adail Sebastião Rodrigues Júnior. A representação de personagens gays na coletânea de contos Stud e em sua tradução As Aventuras de um Garoto de Programa. 01/03/2006</p>

	<p>1v. 255p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Eliana Amarante de Mendonça Mendes</p> <p>Biblioteca Depositaria: FALE e CENTRAL</p> <p>Email do autor:</p> <p>Palavras - chave: contos gays, estudos da tradução, transitividade</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Carlos Alberto Gohn Diva Cardoso de Camargo Eliana Amarante de Mendonça Mendes João Azenha Júnior Sônia Maria de Oliveira Pimenta</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Seguindo a tradição das abordagens discursivas aplicadas à tradução, sobretudo um modelo hallidayano para análises textuais, esta tese investiga como as personagens gays são representadas na coletânea de contos intitulada Stud, escrita originalmente nos anos 1960 no contexto norte-americano, e em sua tradução As Aventuras de um Garoto de Programa, traduzida cerca de trinta anos depois no contexto brasileiro. O sistema de transitividade foi eleito como recurso lingüístico para a análise das experiências de mundo das personagens gays por meio de suas ações, comportamentos, sentimentos, relações e falas. A gramática sistêmico-funcional tem sido usada por um grupo considerável de</p>
--	---

		<p>pesquisadores no campo dos Estudos da Tradução interessados, sobretudo, na compreensão da tradução literária de um ponto de vista discursivo. No entanto, no que tange à literatura gay, o sistema de transitividade ainda não foi aplicado na investigação de como personagens gays são representadas em narrativas de primeira pessoa. Esta pesquisa, pois, além de analisar o sistema de transitividade de um corpus Inglês/Português de contos gays, também utiliza o software WordSmith Tools para quantificar os dados e analisá-los discursivamente. A metodologia usada teve como principais etapas os pontos seguintes: (i) todo o corpus foi digitalizado e corrigido; (ii) em seguida, o corpus foi anotado manualmente a fim de mostrar que papéis discursivos o narrador e as outras personagens gays assumiam nas histórias – ou seja, se as personagens estavam agindo, se comportando, sentindo, pensando, falando, dentre outros aspectos discursivos, de acordo com o processo (verbo) vinculado a cada uma delas; (iii) por fim, as personagens gays foram investigadas através do ponto de vista do narrador, admitindo-se que os contos são narrados em primeira pessoa, com narrador participativo. As análises demonstram, especialmente, que há uma frequência significativa de processos que representam participantes humanos agindo sobre o mundo em seu redor, bem como partes dos corpos das personagens, vistas como agentes ou participantes abstratos, que igualmente agem sobre o mundo ao seu redor. Os resultados também apontam para a aplicação de outras possibilidades teóricas para a investigação realizada nesta pesquisa, principalmente os valores e interesses ideológicos que subjazem à publicação de Stud nos E.U.A. e sua tradução, cerca de trinta anos depois, no Brasil.</p>
19.	2006	<p>Daniel Antônio de Souza Alves. Aspectos da representação do discurso em textos traduzidos: os verbos de elocução neutros. 01/03/2006</p> <p>1v. 111p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Adriana Silvina Pagano</p> <p>Biblioteca Depositária: FALE e CENTRAL</p> <p>Email do autor:</p> <p>Palavras - chave: estudos da tradução, lingüística de corpus, retextualização</p> <p>Área(s) do conhecimento: LINGÜÍSTICA</p>

		<p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta dissertação se insere na interface entre os Estudos da Tradução e a Linguística de Corpus, afiliando-se teoricamente à Gramática Sistemico-Funcional e obras baseadas na teoria sistêmica que exploram aspectos da representação do discurso. Investiga-se aqui a representação do discurso em citações paratáticas em um corpus paralelo bilíngüe, composto pelas obras <i>The adventures of Huckleberry Finn</i> (de Mark Twain), <i>As aventuras de Huck</i> (de Monteiro Lobato) e <i>As aventuras de Huckleberry Finn</i> (de Sérgio Flaksman). Tendo como foco os verbos de elocução neutros, esta pesquisa busca verificar uma tendência (apontada por Assis, 2004) de não retextualização dos Processos Verbais em romances traduzidos para o português brasileiro. Os resultados da pesquisa não apontam exclusivamente para um movimento de não retextualização de tais Processos Verbais. Verifica-se, na retextualização de Lobato, tanto um movimento no sentido de não retextualização de verbos de elocução neutros quanto um movimento de diversificação dos processos empregados e verifica-se na retextualização de Flaksman o que parece ser uma preocupação em se aproximar dos padrões da textualização de Twain.</p>
20.	2005	<p>Leticia Taitson Bueno. Transitivity, cohesion and creativity lexical in the parallel corpus Macunaíma, de Andrade e Macunaíma, de Goodland. 01/03/2005</p>

	<p>1v. 176p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Célia Maria Magalhães</p> <p>Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: transitividade, coesão, criatividade lexical</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Leila Bárbara</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Este trabalho aborda a transitividade, a coesão e a criatividade lexical em um corpus composto pela rapsódia Macunaíma, de Mário de Andrade, e sua tradução, Macunaíma, por E. A. Goodland, a partir da perspectiva dos estudos discursivos da tradução baseados em corpora de pequena dimensão. O arcabouço teórico fundamental para esta pesquisa foi a gramática sistêmico-funcional de Halliday, aliada a princípios metodológicos da lingüística de corpus. Partindo de itens (isolados ou em colocação) considerados lexicalmente criativos na proposta de Kenny (2001), a transitividade e a coesão (através dos processos e participantes e da coesão lexical, respectivamente) foram analisados com base nos dados levantados principalmente com o auxílio do programa de computador WordSmith Tools. Para a análise das colocações criativas, foi escolhido</p>
--	--

		<p>um nódulo específico, fundamental na construção do domínio experiencial relacionado ao sexo, em Macunaíma, e suas redes coesivas. Foram constatadas representações bastante distintas na textualização e na retextualização, embora os perfis da transitividade fossem semelhantes, o que aponta para a necessidade de estender as análises processuais para além dos processos propriamente ditos e do sistema experiencial da função ideacional; e para a importância de se tentar relacionar estas diferenças a aspectos culturais de cada textualização.</p>
21.	2005	<p>Roberta Rego Rodrigues. A Organização Temática em A Hora da Estrela e The Hour of The Star. 01/03/2005 1v. 172p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: Estudos da Tradução, Lingüística de Corpus, Transitividade</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - PROF</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta dissertação apresenta um estudo da organização temática em A hora da estrela de Clarice Lispector e sua</p>

		<p>tradução para o inglês, <i>The hour of the star</i>, por Giovanni Pontiero. A organização temática é abordada enquanto recurso coesivo estrutural (Halliday, 1994), sensível a especificidades de gênero do discurso e passível de ser mapeada a partir da análise e classificação de todos os Temas dentro de um texto, verificando-se como escolhas em nível local da oração possuem impacto no nível global da construção textual. No âmbito dos estudos da tradução, a organização temática é reconhecida como problema de tradução (Baker, 1992; Vasconcellos, 1997, 1998) e pode ser abordada a partir de critérios que diferenciem as escolhas temáticas do tradutor motivadas por restrições sistêmicas daquelas que evidenciam idiossincrasias do mesmo (Munday, 1998). A presente pesquisa parte do aparelho teórico da lingüística sistêmica hallidayana, complementada pelos trabalhos de Downing (1991), Eggins (1994), Martin et al. (1997), Thompson (2002), para examinar um corpus composto por um texto original e sua tradução, utilizando recursos da metodologia da lingüística de corpus. Para efeito de coleta de dados, o corpus foi digitalizado e revisado. Após esta etapa, houve a anotação manual do mesmo e posterior extração de dados mediante o programa WordSmith Tools. Os dados quantitativos foram analisados com vista à identificação de padrões de escolhas temáticas em ambos os textos. Resultados apontam que o texto alvo apresenta uma alta frequência de Temas ideacionais participantes enquanto que o texto fonte apresenta um alto índice de Temas ideacionais processo-participantes. Resultados também apontam para padrões distintos na correlação dos Temas interpessoais com os ideacionais participantes e ideacionais processo-participantes, os quais revelam representações diferentes do narrador da novela. A análise do narrador e dos processos em posição temática relacionados a ele dentro do sistema da transitividade também corrobora as diferenças nas representações observadas por meio da análise da organização temática. Pode-se concluir que o tradutor segue padrões do texto fonte em sua re-textualização ao mesmo tempo em que modifica a organização temática em muitos momentos na sua tradução, o que pode ser tomado como evidência de um estilo próprio.</p>
22.	2005	<p>Eliza Mitiyo Morinaka. GABRIELA, CRAVO E CANELA AND THE ENGLISH (RE)TEXTUALIZATION: REPRESENTATION THROUGH LEXICAL RELATIONS. 01/09/2005</p> <p>1v. 115p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p>

		<p>Orientador(es): Maria Lúcia Barbosa de Vasconcellos Biblioteca Depositaria: Biblioteca Central UFSC</p> <p>Email do autor:</p> <p>Palavras - chave: (Re)textualization of Gabriela, cravo e canela into English</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Maria Lúcia Barbosa de Vasconcellos Viviane Maria Heberle</p> <p>Linha(s) de pesquisa: Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Essa pesquisa insere-se no campo dos Estudos da Tradução, sob uma perspectiva interdisciplinar abarcando Lingüística Sistêmico-Funcional e Estudos da Tradução baseados em Corpus, com enfoque no estudo comparativo da protagonista Gabriela no romance Gabriela, cravo e canela de Jorge Amado (1958) e sua (re)textualização para o inglês Gabriela, clove and cinnamon traduzido por James Taylor e William Grossman (1962). Visando-se verificar como relações lexicais estabelecem coesão textual para construir a representação da personagem em ambos os textos, parte-se do princípio que relações lexicais e Transitividade pertencem ao mesmo nível na semântica-discursiva (Eggin, 2000). O arcabouço teórico de coesão lexical é de Halliday e Hasan (1976) e o conceito de função Ideacional é baseado em Halliday (1994). As ferramentas metodológicas dos Estudos da Tradução com base em Corpus contribuem para agilizar o processo em busca de padrões emergentes, os</p>
--	--	---

		<p>quais são posteriormente anotados, quantificados e analisados. Os resultados quantitativos indicam que ambos os textos usam relações de sinonímia e meronímia como ferramentas de coesão lexical para representar Gabriela. Os resultados qualitativos apontam algumas alterações na representação da personagem na (re)textualização, particularmente no uso de relações de meronímia, onde os tradutores tentam manter uma certa restrição em relação às conotações sensuais e sexuais referentes à personagem. Espera-se poder contribuir para um abordagem interdisciplinar dos Estudos da Tradução, pois uma teoria sistêmico-funcional facilita a descrição de corpus oferecendo ferramentas para estudar sistemas lingüísticos distintos de várias perspectivas.</p>
23.	2005	<p>Tassiani Miguel Lopes Caçado. Transitividade e Representação do Discurso no Corpus Paralelo Interview with the Vampire/ Entrevista com o Vampiro. 01/03/2005 1v. 142p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Célia Maria Magalhães Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: tradução, lingüística sistêmica, lingüística de corpus</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Stella Esther Ortweiler Tagnin</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - PROF</p> <p>Idioma(s): Português</p> <p>Dependência administrativa</p>

		<p>Federal</p> <p>Resumo tese/dissertação:</p> <p>Este trabalho busca analisar a representação do discurso de dois personagens no corpus paralelo Interview with the vampire, de Anne Rice, e Entrevista com o vampiro, de Clarice Lispector. Como foco de análise, destacam-se os verbos de elocução e os processos de transitividade, observando-se os padrões de (re)textualização que representam o discurso dos dois personagens envolvidos na entrevista narrada no romance. A investigação baseou-se nos pressupostos teóricos dos Estudos da Tradução em diálogo com a lingüística sistêmica e a lingüística de corpus. A metodologia consistiu na classificação dos verbos introdutórios dos discursos diretos do entrevistador e do entrevistado no romance, através das categorias de verbos de elocução e de processos de transitividade, com o auxílio dos recursos disponibilizados pelo programa de computação WordSmith Tools. Os resultados das análises revelaram que, embora a distribuição dos verbos e processos apresentem poucas diferenças entre os dois textos, os padrões na retextualização do romance constataam o uso mais acentuado de elementos característicos do gênero “entrevista”, em comparação com os padrões utilizados na sua textualização.</p>
24.	2005	<p>Viviane Paquilin. THE VARIOUS FACETS OF A MESSAGE: AN ANALYSIS OF THE THEMATIC STRUCTURE IN BRIDGET JONES'S DIARY IN THE LIGHT OF THE SYSTEMIC FUNCTIONAL GRAMMAR, CORPUS LINGUISTICS AND TRANSLATION STUDIES INTERFACE. 01/02/2005</p> <p>1v. 101p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Maria Lúcia Barbosa de Vasconcellos</p> <p>Biblioteca Depositária: Biblioteca Central UFSC</p> <p>Email do autor:</p> <p>Palavras - chave:</p> <p>Estrutura Temática; Corpus paralelo bilíngüe</p> <p>Área(s) do conhecimento:</p> <p>LINGÜÍSTICA APLICADA</p> <p>Banca examinadora:</p> <p>Gloria Gil</p> <p>Lincoln Paulo Fernandes</p> <p>Maria Lúcia Barbosa de Vasconcellos</p>

		<p>Linha(s) de pesquisa: Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CNPq</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: A interface entre Lingüística Sistemica Funcional, Lingüística de Corpus e Estudos da Tradução tem fornecido ferramentas prolíficas para o treinamento, ensino e pesquisa em tradução. Nesta interface, é de interesse especial a noção de escolha e sua direta relação com o significado da mensagem, que permite ao tradutor modelar a realidade de acordo com as decisões que ele (a) toma a fim de retextualizar um texto escrito em um cenário sócio-cultural diferente. Neste contexto, a Estrutura Temática desempenha um papel crucial para a construção de significado de uma mensagem, uma vez que sua manutenção ou não, no texto de chegada, pode ser influenciada por restrições lingüísticas, culturais ou ideológicas. As organizações temáticas dos textos de partida (TP) e de chegada (TC) podem alterar a função comunicativa da mensagem, bem como o estilo do autor do texto original. O corpus analisado consiste na versão inglesa (TP) e de sua versão na língua portuguesa brasileira (TC) de O Diário de Bridget Jones, escrito por Helen Fielding e traduzido por Beatriz Horta. Os conceitos de tema marcado e não-marcado são explorados a fim de investigar o comportamento da tradução com base nos padrões temáticos emergentes. A descrição dos dados foi feita com o auxílio do software Wordsmith Tools.</p>
25.	2005	<p>Marcos Pereira Feitosa. Uma proposta de anotação de corpora paralelos com base na Lingüística Sistemico-Funcional. 01/03/2005</p> <p>1v. 178p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: FALE e BU</p>

	<p>Email do autor:</p> <p>Palavras - chave: Linguística de Corpus, Tradução, Organização Temática</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Este trabalho situa-se no campo disciplinar dos Estudos da Tradução, tendo como abordagens a Linguística Sistemico-Funcional (Gramática Sistemico-Funcional de M. A. K. Halliday) e a Linguística de Corpus e é desenvolvido no âmbito do Núcleo de Estudos da Tradução (NET) da Faculdade de Letras (FALE) da Universidade Federal de Minas Gerais (UFMG). Visando uma anotação de corpora mais eficiente e menos propensa a erros de digitação, os quais interferem nos dados das pesquisas, propõe-se aqui um modelo de anotação instrumentalizado através de um código numérico para a rotulação de corpora com base na Gramática Sistemico-Funcional de Halliday. Esse código, denominado “CROSF” (Código de Rotulação Sistemico-Funcional), é desenvolvido através de sucessivos protótipos testados em um corpus paralelo bilíngüe composto de um fragmento do romance <i>The blind assassin</i> de Margaret Atwood (2000) em inglês canadense e sua tradução para o português brasileiro feita por Léa Viveiros de Castro (Atwood, 2001). O estudo de caso aplicando-se o protótipo</p>
--	--

		<p>final do CROSF ao corpus referido apontou aspectos da Tematização da informação decorrentes de diferenças sistêmicas entre as línguas (destacando-se a elipse do sujeito em português), bem como escolhas idiossincráticas da tradutora. O código numérico facilitou as buscas no corpus eletrônico, permitindo melhor visibilidade do texto anotado e buscas de diferentes combinações de escolhas Temáticas.</p>
26.	2004	<p>Silvana Maria de Jesus. Representação do Discurso e Tradução: padrões de textualização em corpora paralelo e comparável.. 01/02/2004</p> <p>1v. 128p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Adriana Silvina Pagano</p> <p>Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: tradução, linguística sistêmico-funcional</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Antônio Paulo Berber Sardinha José Luiz Vila Real Gonçalves</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CNPq</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Esta dissertação insere-se nos Estudos da Tradução baseados em corpora, numa abordagem discursiva da tradução. O objeto de estudo é a representação do discurso na ficção, abordada a partir dos pressupostos teóricos da lingüística</p>

		<p>sistêmico-funcional de Halliday (1994), utilizando-se a metodologia da Lingüística de Corpus aplicada em um corpus de pequenas dimensões, analisado de forma automática, semi-automática e manual. Os romances Point counter point (1928), Contraponto (1933) e Caminhos cruzados (1935) formam o corpus da pesquisa, classificados, sob a perspectiva dos estudos de corpora, em um subcorpus paralelo bilíngüe composto por Point counter point, um romance inglês de Aldous Huxley, e Contraponto, sua tradução para o português do Brasil, e um subcorpus comparável formado por Caminhos cruzados, um romance de Érico Veríssimo, e Contraponto, tradução feita por Veríssimo, ambos em português do Brasil. O corpus foi analisado sob a perspectiva do sistema da transitividade proposto por Halliday, dentro da metafunção ideacional da linguagem, com enfoque nos processos mentais e no sistema lógico-semântico da projeção. Foram analisadas as formas de representação do pensamento das personagens, focalizando-se dois verbos prototípicos de processos mentais, THINK, em inglês, e PENSAR, em português, visando-se observar especificidades nos padrões de textualização das três obras em análise. As formas e funções de THINK e PENSAR foram investigadas no corpus da pesquisa, a partir da análise dos tipos de fenômeno projetados por THINK e PENSAR, segundo as categorias propostas por Halliday (1994) e Thompson (1994). Também foram analisadas as formas e funções dos diferentes tipos de fenômeno encontrados no corpus, sob a perspectiva da gramática sistêmico-funcional de Halliday, segundo a qual forma e função se complementam na construção do significado. Foram observadas, ainda, as formas de re-textualização de THINK e PENSAR no corpus paralelo bilíngüe. O software WordSmith Tools foi utilizado para o levantamento de dados quantitativos e o processo de anotação para a análise qualitativa. A pesquisa revela padrões lingüísticos utilizados pelos autores no uso dos verbos de elocução THINK e PENSAR, bem como diferenças na forma de textualização de PENSAR por Veríssimo tradutor e Veríssimo autor. Corroborar, também, o potencial de aplicação da lingüística sistêmico-funcional de Halliday e da Lingüística de Corpus aos estudos da tradução, especialmente na análise de obras literárias, em português do Brasil.</p>
27.	2004	<p>Roberto Carlos de Assis. A Transitividade na Representação de Sethe no Corpus Paralelo Beloved-Amada. 01/08/2004 1v. 122p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p>

	<p>Orientador(es): Célia Maria Magalhães Biblioteca Depositária: FALE e BU</p> <p>Email do autor:</p> <p>Palavras - chave: tradução, gramática sistêmico-funcional</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Este trabalho se insere no campo de estudos discursivos da tradução baseados em corpus de pequena dimensão. O corpus da pesquisa é composto pelo romance <i>Beloved</i>, de Toni Morrison, e a tradução <i>Amada</i>, de Evelyn Kay Massaro. O objeto de estudo é a transitividade como um dos recursos para a análise da construção de uma personagem literária. Foi de interesse particular observar se a construção da personagem Sethe se dava da mesma forma no texto em inglês e em sua retextualização para o português brasileiro. Para tanto, foram utilizados os pressupostos teóricos da gramática sistêmico-funcional de Halliday, com o suporte dos instrumentos da Linguística de Corpus. Foram analisados os tipos de processos nos quais a personagem está envolvida como participante no texto original e comparados aos resultados encontrados no texto traduzido. O programa de computador WordSmith Tools foi o programa de computador utilizado para a manipulação do</p>
--	--

		<p>corpus que foi tratado semi-automaticamente. Os resultados mostraram a importância do sistema de transitividade para análise da representação de personagens e apontaram para padrões distintos de retextualização dos diferentes tipos de processos. Embora a distribuição dos tipos de processos referentes à personagem analisada seja semelhante nos dois textos, observações detalhadas apontaram para mudanças na sua representação no texto traduzido.</p>
28.	2003	<p>Osilene Maria de Sá e Silva da Cruz. Harry Potter and the Chamber of Secrets e sua Tradução para o Português do Brasil: uma análise dos verbos de elocução, com base na lingüística sistêmica e nos estudos de corpora. 01/05/2003 1v. 207p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS Orientador(es): Adriana Silvina Pagano Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG</p> <p>Email do autor: osileneacruz@bol.com.br</p> <p>Palavras - chave: Harry Potter, texto traduzido</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação:</p>

		<p>Esta dissertação tem por objetivo investigar os verbos de elocução e os processos de transitividade que os mesmos configuram, a partir de um corpus bilíngüe paralelo, de textos escritos em inglês e em português do Brasil, a fim de indagar, através do estudo de corpus de pequena dimensão, como se comportam os verbos de elocução e de que forma os mesmos evidenciam padrões de textualização no texto fonte e no texto traduzido. O corpus desta pesquisa é constituído pelo romance Harry Potter and the chamber of secrets (1999) e sua tradução para o português do Brasil, Harry Potter e a câmara secreta (2000). A partir da abordagem da Collins Cobuild English Grammar (1990) e da Gramática Sistemica Funcional de Halliday, e utilizando os subsídios da Lingüística de Corpora, a pesquisa examina as escolhas da autora e da tradutora, relativas aos verbos de elocução e processos, sob a perspectiva do aspecto lógico da função ideacional, mais especificamente em projeções paratáticas e sob a perspectiva do aspecto experiencial, observando-se os padrões evidenciados na textualização realizada por cada uma delas. Foi utilizado o programa WordSmith Tools, especificamente, as ferramentas Wordlist, Aligner e Concord para a obtenção de dados quantitativos que permitiram verificar a presença de padrões nos textos analisados, os quais podem estar associados a convenções da própria língua (de origem ou de chegada) ou a escolhas da autora e da tradutora. Os resultados corroboram a relevância de se utilizar os processos no sistema da transitividade, enquanto configuradores de personagens no mundo ficcional, e o papel dos verbos de elocução como elementos significativos na construção de um romance e em sua textualização para uma outra língua.</p>
29.	2003	<p>Cristina Mauri. Um estudo da tradução italiana de Laços de Família, de Clarice Lispector, a partir da abordagem em "corpora": a construção da introspecção feminina através dos verbos de elocução. 01/05/2003</p> <p>1v. 133p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS</p> <p>Orientador(es): Célia Maria Magalhães</p> <p>Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG</p> <p>Email do autor: mauricristina@hotmail.com</p> <p>Palavras - chave: Clarice Lispector, personagens femininas</p>

		<p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA</p> <p>Banca examinadora: Adriana Silvina Pagano Célia Maria Magalhães Maria Lúcia Barbosa de Vasconcellos</p> <p>Linha(s) de pesquisa: TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Português</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Este estudo busca realizar uma análise comparativa da coletânea de contos <i>Laços de família</i>, de Clarice Lispector, e sua tradução em italiano <i>Legami familiari</i>, por Adelina Aletti. O foco dessa análise são os verbos de elocução, precisamente os que indicam processos mentais das personagens femininas, e suas colocações. Trata-se de uma análise lingüístico-discursiva que se insere dentro dos estudos realizados em corpora de pequenas dimensões. O suporte teórico é fornecido pela Lingüística de Corpora, Estudos da Tradução e Análise do Discurso. A metodologia de pesquisa se baseia na aplicação do Wordsmith Tools que permite uma análise detalhada dos verbos de elocução em seu contexto de ocorrência no corpus. Os resultados alcançados apontam algumas diferenças na tradução que na maioria dos casos são determinadas pela língua de chegada, ou por outros fatores que contribuem para criar um estilo próprio na tradução. As escolhas da tradutora não determinadas pelas instâncias de uso do italiano contribuem para uma retextualização de um nível distinto de introspecção para as personagens femininas.</p>
30.	1997	<p>Maria Lúcia Barbosa de Vasconcellos. <i>Retextualizing Dubliners: A Systemic Functional Approach to Translation Quality Assessment</i>. 01/12/1997</p>

	<p>1v. 273p. Doutorado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLES E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Rosa Weingold Konder</p> <p>Biblioteca Depositaria: Biblioteca Central da Universidade Federal de Santa Catarina</p> <p>Email do autor:</p> <p>Palavras - chave: tradução; cultura; tradução literária no contexto receptor</p> <p>Área(s) do conhecimento: LINGÜÍSTICA APLICADA TEORIA E ANÁLISE LINGÜÍSTICA</p> <p>Banca examinadora: Rosa Weingold Konder Solange Ribeiro de Oliveira Sonia Zyngier Stella Esther Ortweiler Tagnin Walter Carlos Costa</p> <p>Linha(s) de pesquisa: Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - a tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação: CAPES - DS</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: Tendências recentes em estudos em tradução literária têm demonstrado interesse em contatos entre literaturas e culturas. Assim, a pesquisa na área tem se concentrado na íntima relação entre tradução e cultura e na função da tradução literária no contexto receptor. Embora de grande importância, tais estudos enfatizam os aspectos culturais, desconsiderando os aspectos lingüísticos da tradução. Esta tese propõe um tipo especial de análise, descrição lingüística e avaliação de textos literários traduzidos, baseando-se no</p>
--	---

		<p>referencial teórico da Linguística Sistemico-Funcional, uma teoria de linguagem que toma como ponto de partida as significações das expressões lingüísticas. Operando com o conceito de tradução como retextualização, este estudo compara dois contos de Joyce, 'Araby' e 'Eveline' com algumas traduções publicadas no Brasil. O texto de partida e os textos traduzidos são descritos em termos de escolha sistêmicas, com especial atenção às configurações que emergem para a codificação de significados interpessoais e ideacionais. A sensibilidade dos tradutores para tais significados selecionados e realizados no texto de partida e sua resposta a eles em termos das opções retextualizadas informam a avaliação dos textos traduzidos. Argumenta-se que a análise de textos literários e de suas retextualizações em termos de sua forma de significar, ou seja, estilística translacional baseada nos conceitos da Linguística Sistemico-Funcional, oferece um sólido referencial para a avaliação da qualidade de textos literários traduzidos.</p>
--	--	--

Anexo 4: Orientações de Dra. Maria Lúcia Vasconcellos até 2011

Titulo	Autor	Ano	Orientador	IFES	Tipo
1) Domesticating, Foreignizing, and Omitting Culture-Specific Items: A Systemic-Functional Linguistics Approach to Subtitling Analysis of the TV Series Heroes	Rafael Matielo	2011	Maria Lúcia B. Vasconcellos	UFSC	MA RESumo não se encontra no Banco de Teses da Capes
2) The use and abuse of subtitling in Cidade de Deus and Boy'z in the Hood	Elaine Espindola	2005	Maria Lúcia B. Vasconcellos	UFSC	MA
3) A Corpus-Based Approach to ESP: EST Vocabulary in Information Technology	Sara Geane Kobelinski	2005	Maria Lúcia B. Vasconcellos	UFSC	MA
4) The traffic of English words in the Brazilian translation context: a case study	Sinara de Oliveira Branco	2002	Maria Lúcia B. Vasconcellos	UFSC	MA
5) The traffic of English words in the Brazilian translation context: a case study	Lucila Augusta Campesato	2002	Maria Lúcia B. Vasconcellos	UFSC	MA
6) Friends: The Translation of Humor in a Polysemiotic Text	Daniela Maria Ioppi Chile	2001	Maria Lúcia B. Vasconcellos	UFSC	MA

7) The Interdependence of Extratextual and Intratextual Factors in Translated Texts: A Sample of Ícaro Brasil, Varig's Bilingual In-Flight Magazine.	Eliana Maria Limongi	2000	Maria Lúcia B. Vasconcellos	UFSC	MA
8) A Janus-Faced Manipulation: an Investigation of Narrative Construction on the Amazon in the Target and Source Context	Christine Bareño Etges	2000	Maria Lúcia B. Vasconcellos	UFSC	MA
9) Character, Language and Translation: a Linguistic Study of the Construction of Characters in a Cinematic Version of William's A Streetcar Named Desire	Luciany Margarida da Silva	1999	Maria Lúcia B. Vasconcellos	UFSC	MA
10) Tim Burton's the Nightmare Before Christmas and its Subtitled Counterpart in the Brazilian Context: an Analysis of Genre, Register and Lexicogrammatical Features	Lincoln Paulo Fernandes	1998	Maria Lúcia B. Vasconcellos	UFSC	MA

11) Brazilian Practices of translating names in children's fantasy literature: a corpus-based study	Lincoln Paulo Fernandes	2004	Maria Lúcia B. Vasconcellos	UFSC	PHD
12) The Abusive Alternative: Translation of Culture-bound Terms / Brazilian Films for American Audiences: Subtitling of Culture-bound Terms.	Ritalice Ribeiro de Medeiros	2003	Maria Lúcia B. Vasconcellos	UFSC	PHD

Resumos

Elaine Espindola. THE USE AND ABUSE OF SUBTITLING AS A PRACTICE OF CULTURAL REPRESENTATION: CIDADE DE DEUS AND BOYZ 'N THE HOOD. 01/02/2005

1v. 92p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca Central UFSC

Email do autor:

Palavras - chave:

legendagem; representação cultural;

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Fernando Simão Vugman

Maria Lúcia Barbosa de Vasconcellos

Viviane Maria Heberle

Linha(s) de pesquisa:

Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CNPq

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

No que diz respeito a legendagem de filme, TV e vídeo, os países do mundo globalizado podem ser divididos, de um modo geral, em dois blocos distintos, a saber, (i) países legendados, i.e. países que falam a língua inglesa; (ii) países legendadores, i.e. países que não falam a língua inglesa (Gottlieb, 1997). Levando-se em consideração o fato de que a prática de assistir filmes é uma forma de entretenimento comum no meio internacional de comunicação audiovisual e considerando-se as desigualdades de vários níveis existentes entre estes dois grandes blocos de países, a legendagem torna-se uma forma de prática cultural de dimensões econômicas e políticas, implantada no cenário intercultural mais amplo. Neste contexto, o presente estudo investiga o tratamento dado aos termos culturais presentes em legendas do par de línguas – Português/Inglês – em ambas as direções, nos filmes “Cidade de Deus” (2001) and Boys ’N the Hood”(1991). Esta análise parte das hipóteses iniciais de que (i) elementos culturais tendem a ser representados ou adulterados nas legendas, dependendo da direção da tradução, e (ii) o tratamento dado aos termos culturais inevitavelmente afeta a maneira como as duas culturas envolvidas são representadas. A investigação é realizada no âmbito do arcabouço teórico localizado na interface entre Estudos da Tradução (ET) e Estudos Culturais (EC), que se mostra capaz de fornecer ferramentas para iluminar a análise dos chamados termos culturais e o tratamento dado a eles na prática de legendagem envolvendo o Brasil e os Estados Unidos nos filmes selecionados. Mais especificamente, os conceitos de domesticação e estrangeirização propostos por Venuti (1995), o conceito de, representação cultural proposto por Hall (1997) e, finalmente, o conceito de legendagem abusiva proposto por Nornes (1999), são explorados da descrição e análise do dados levantados. Os resultados obtidos apontam para um processo de exclusão dos termos culturais do filme do país periférico, conforme percebidas no processo de legendagem dos filmes em estudo.

Sara Geane Kobelinski. A CORPUS-BASED APPROACH TO ESP: EST VOCABULARY IN INFORMATION TECHNOLOGY. 01/05/2005

1v. 116p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca Central UFSC

Email do autor:

Palavras - chave:

Ensino de vocabulário técnico; Lingüística de Corpus

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Leda Maria Braga Tomitch

Lincoln Paulo Fernandes

Maria Lúcia Barbosa de Vasconcellos

Linha(s) de pesquisa:

Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Outro

Dependência administrativa

Federal

Resumo tese/dissertação:

O objetivo desta pesquisa é oferecer uma contribuição para a investigação de instrução de vocabulário. No contexto do ensino de inglês como língua estrangeira, a tradição foi, por bastante tempo, ensinar gramática e estruturas da língua. Com o advento da Lingüística Aplicada, nos anos noventa, a importância do ensino de vocabulário vem obtendo uma posição proeminente, pois gramática e léxico foram entendidos como assuntos inseparáveis. Neste novo cenário, vocabulário especializado é estudado como uma sub-linguagem, na qual até mesmo simples palavras lexicais adquirem novos significados. Esta sub-linguagem que está presente em contextos técnicos, faz parte das habilidades de um aluno que estuda inglês com propósitos específicos ou acadêmicos. Nesta área, apesar do fato de o material didático estar estabelecido como uma ferramenta indispensável, nota-se que esta ferramenta tem algumas limitações por razões práticas, a saber: (i) O uso de material não autêntico; (ii) não ser constantemente atualizado, por ser impresso; (iii) ser, ainda, baseado em gramática; (iv) não explorar fontes criativas além do material impresso. Em uma tentativa de remediar esta situação, algumas metodologias estão avançando em estudos de Lingüística de Corpus (LC), a qual nos permite estudar fatos de isolação semântica das palavras em

áreas específicas. O interesse crescente em metodologia LC, em uma visão pedagógica, está relacionada à possibilidade de coletar por via eletrônica textos autênticos e atualizados. Há três métodos que usam metodologia de LC, "Lexical Syllabus", "Lexical Approach" e "Data-Driven Learning". Mesmo importantes e inovadores, estes métodos não exploram uma área específica, vocabulário em inglês para Ciência e Tecnologia. Esta pesquisa pretende expandir o alcance de tal metodologia investigando como ela pode desenvolver a aprendizagem de vocabulário técnico. Neste contexto as seguintes perguntas de pesquisa são lançadas: (i) Como o vocabulário técnico se comporta em textos de Análise de Sistemas? (ii) CL é uma metodologia adequada para produção de material pedagógico em Análise de Sistemas? (iii) Como aprendizes respondem a abordagens baseadas em corpus? A metodologia de pesquisa inclui compilação de um corpus de estudo, estudo de padrões de linguagem submetidas a análise do programa WordSmith tools (Scott, 1996), elaboração de material pedagógico baseado pelos métodos de ensino baseados em corpus, bem como, a aplicação e análise do material desenvolvido.

Sinara de Oliveira Branco. The Traffic of English Words in the Brazilian Translation Context: A Case Study. 01/02/2002

1v. 90p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca Universitária da UFSC

Email do autor:

sbranco@brasilnet.net

Palavras - chave:

Tradução; Identidade Cultural; Empréstimos Lingüísticos

Área(s) do conhecimento:

Banca examinadora:

Adriana Silvina Pagano

José Roberto Basto O'Shea

Maria Lúcia Barbosa de Vasconcellos

Linha(s) de pesquisa:

Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

De todas as taxonomias utilizadas como procedimentos de tradução, a mais completa e também a mais criticada é a de Vinay e Darbelnet - *Stylistique Comparée du Français et de l'Anglais* (1958). Apesar das críticas, esta taxonomia continua exercendo uma posição de destaque que tem sido comprovada por reedições da mesma. Entre as categorias propostas, o empréstimo é considerado o mais simples de todos os procedimentos de tradução. Dessa forma, explica-se o uso de tal técnica por razões diferentes, desde à introdução da cultura do texto de partida na cultura do texto de chegada até à criação de efeitos estilísticos. Entretanto, dimensões de poder estão presentes neste contexto e estas não são consideradas no modelo de Vinay & Darbelnet. Se o fossem, o empréstimo passaria do mais simples ao mais complexo procedimento de tradução, considerando-se as conseqüências e implicações derivadas de seu uso. Nesta dimensão de poder da tradução, esta pesquisa visa problematizar o uso de empréstimos lingüísticos como uma técnica simples e de fácil utilização, mostrando que várias forças se fazem presentes, particularmente na direção primeiro-terceiro mundo. As fontes de dados desta dissertação são formadas por: (i) Um livro traduzido para o português; (ii) Um livro originalmente escrito em português e (iii) Artigos de revistas e jornais norte-americanos traduzidos e publicados no Brasil. O objetivo é analisar o material selecionado verificando com que frequência empréstimos lingüísticos são utilizados e observar que forças influenciam a utilização de tal procedimento. Depois da apresentação e investigação das fontes de dados, a conclusão apresenta pontos de reflexão sobre o uso de empréstimos no contexto brasileiro

Lucila Augusta Campesatto. THEMATIC STRUCTURE IN BRAZILIAN PORTUGUESE ABSTRACTS IN ENGLISH TRANSLATION: STEPS TOWARDS A DISCUSSION OF TEXTUAL COMPETENCE. 01/05/2002

1v. 99p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca Central da UFSC

Email do autor:

clucila@uri.com.br

Palavras - chave:

tradução, competência textual de tradutores

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano
 Andreia Guerini
 Maria Lúcia Barbosa de Vasconcellos

Linha(s) de pesquisa:

Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

Ao submeter uma pesquisa para publicação, acadêmicos brasileiros enfrentam o problema de traduzir os resumos escritos em Português Brasileiro para o Inglês. A fim de produzir textos gramaticalmente e estruturalmente corretos na língua de chegada, tradutores de abstracts usam algumas estratégias de tradução que, algumas vezes, perturbam a estrutura temática do texto de chegada. O objetivo desta dissertação é analisar a competência textual dos tradutores, através do estudo de estratégias visando identificar como tradutores de abstracts lidam com a estrutura temática do texto de partida no texto de chegada. Dois tipos de dados fazem parte do estudo, (i) multilingual data source, que servirá de base para a análise dos textos traduzidos, e (ii) parallel data source, para a análise de estratégias de tradução.

Daniela Maria Ioppi Chile. "Friends": The Translation of humor in a polysemiotic text. 01/02/2001

1v. 132p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos
 Biblioteca Depositária: Biblioteca da PGI

Email do autor:

Palavras - chave:

Tradução

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Maria Lúcia Barbosa de Vasconcellos

Stella Esther Ortweiler Tagnin

Walter Carlos Costa

Linha(s) de pesquisa:

Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - a tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - DS

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

No contexto da tradução de formas audio-visuais de humor, esta dissertação examina a legendação, em português, de alguns episódios da série americana "Friends", produzida pela NBC e veiculada no Brasil pela Sony Entertainment Television. A dissertação tem por objetivo investigar os tipos de humor encontrados no episódios selecionados e examinar sua (re)produção, em termos lingüísticos e culturais, para o português do Brasil. As limitações técnicas do modo da escrita são examinados, face à característica poli-semiótica do texto e às diferentes pressuposições culturais do contexto de chegada.

Eliana Maria Limongi. The Interdependence of Extratextual and Intratextual Factors in Translated Texts: A Sample of Ícaro Brasil, Varig's Bilingual In-Flight Magazine. 01/05/2000

1v. 1p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca Central da UFSC

Email do autor:

Palavras - chave:

não informada

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Fábio Alves da Silva Jr.

José Roberto Basto O'Shea

Maria Lúcia Barbosa de Vasconcellos

Linha(s) de pesquisa:

Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - a tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Dentro do contexto do funcionalismo na tradução de textos, esta dissertação explora a contribuição de Christiane Nord aos Estudos de Tradução, particularmente de seu modelo "Análise de Textos voltada para a Tradução". Assim fundamentada, a dissertação apresenta um estudo detalhado da maneira como os fatores extratextuais (i.e., que 'cercam' o ato tradutório) e intratextuais (i.e., inseridos 'no' texto a ser traduzido) se interrelacionam para a produção de um novo texto (Texto de Chegada - TC). Com este fim, é feita uma análise de um texto selecionado de Ícaro Brasil, revista de bordo da Varig, tanto em sua versão original como em sua versão traduzida, sendo Português-Inglês as línguas em questão. A análise de tais fatores permite a identificação dos elementos do texto de partida (TP) que são relevantes para a produção do TC, assim como que o(a) tradutor(a) estabeleça os procedimentos a serem adotados no processo de produção de um novo texto - para uma audiência específica e para fins específicos - para uma nova situação comunicativa, i.e., de chegada. Quanto ao crítico que avalia a tradução, os critérios de Christiane Nord são voláteis no sentido de possibilitar que a crítica seja feita em bases funcionais, levando-se em conta as circunstâncias que condicionam o ato tradutório.

Título: Janus-faced manipulation [dissertação] : an investigation of narratives constructed on the Amazon in the target and source contexts, A / 2000 - Dissertações - Acervo 180336

Número de Chamada CETD UFSC PLLE 0220

Autor Principal Etges, Christine Bareño

Demais autores Vasconcellos, Maria Lucia Barbosa de, orientadora

Autor(es) Entidade(s)	<u>Universidade Federal de Santa Catarina. Centro de Comunicação e Expressão</u>
Título Principal	A janus-faced manipulation [dissertação] : an investigation of narratives constructed on the Amazon in the target and source contexts / Christine Bareño Etges ; orientadora, Maria Lúcia B. de Vasconcellos. -
Publicação	Florianópolis, SC, 2000.
Descrição Física	[100] f. : tabs. ; 30 cm
Notas	Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão. Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente. Bibliografia : f. 79-83.
Notas de Resumo	Resenha : Narrativas de viagem são especialmente interessantes em tradução uma vez que seus autores já 'traduziram' a realidade do país que visitaram, tanto literalmente quanto figurativamente. O tradutor tem, então, a desafiadora tarefa de traduzir o traduzido. Quando acontece de o tradutor ser um nativo do país visitado, a tarefa torna-se ainda mais complexa, pois ele está retornando a seu contexto nacional uma perspectiva construída a partir de olhos estrangeiros. Este trabalho investiga uma dessas narrativas de viagem intitulada 'The world is burning - murder in the rain forest' (Alex Shoumatoff, 1990) e sua tradução 'O mundo em chamas - a devastação da Amazônia e a tragédia de Chico Mendes'. A metodologia adotada consiste em uma investigação descritiva do texto traduzido a fim de identificar os shifts (mudanças) em relação ao texto original, seguindo o modelo adotado por Kitty van Leuven-Zwart (1989,1990). As unidades de tradução são analisadas olhando-se os shifts que ocorrem em áreas como: escolha de itens lexicais individuais, transitividade, modalidade e estrutura temática (ordem de palavras). A análise move-se, então, para descrever como, através de 'shifts' na narrativa, a tradução brasileira reescreve a relação entre narrador / leitor e a imagem construída da Amazônia.
Numero de Controle local	180336

Número controle Bibliodata	(BIBLIODATA) SC000567125
Assuntos	<u>Literatura</u> <u>Lingua inglesa</u> <u>Traduções</u> <u>Narrativa (Retórica) -- Descrições e viagens</u>
<p>Luciany Margarida da Silva. Character, Language and Translation: a Linguistic Study of the Cosntruction of Characters in a Cinematic Version of Williams' A Streetcar Named Desire. 01/03/1999</p> <p>1v. 123p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)</p> <p>Orientador(es): Maria Lúcia Barbosa de Vasconcellos</p> <p>Biblioteca Depositaria: Biblioteca Central da UFSC</p> <p>Email do autor:</p> <p>Palavras - chave: Tradução; Legendação; análise de textos</p> <p>Área(s) do conhecimento:</p> <p>Banca examinadora: David Shepherd Maria Lúcia Barbosa de Vasconcellos Rosa Weingold Konder</p> <p>Linha(s) de pesquisa: Teoria e Prática de Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - a tradução de obras literárias, científicas e técnicas.</p> <p>Agência(s) financiadora(s) do discente ou autor tese/dissertação:</p> <p>Idioma(s): Inglês</p> <p>Dependência administrativa Federal</p> <p>Resumo tese/dissertação: No contexto de tradugáo de filmes, legendagáo é um aspecto importante que merece atengáo. Tradicionalmente, esta forma de tradugáo tem sido</p>	

estudada do ponto de vista dos aspectos técnicos envolvidos em sua produção. Entretanto, estes estudos têm ignorado aspectos lingüísticos cruciais, em suas análises dos textos traduzidos. Esta dissertação investiga a legendagem através da análise da linguagem usada em uma produção cinematográfica de *A Streetcar Named Desire* (1951) de Tennessee Williams e sua versão legendada em português brasileiro, intitulada *Uma Rua Chamada Pecado* (1951). Este estudo se propõe a investigar a maneira com que a personagem principal, Blanche DuBois, é construída na linguagem através da forma com que experiências de seus mundos interior e exterior são modelados por Stanley e por ela mesma. O arcabouço teórico informando este estudo é a *Linguística Sistemática Funcional* sugerida por Halliday (1985/1994), principalmente o componente ideacional, realizado pelo sistema da transitividade. O conceito de tradução usado neste estudo é o conceito de tradução como 'retextualização' sugerido por Costa (1992). Este conceito é baseado na noção de que um texto traduzido é uma retextualização de um outro texto, textualizado anteriormente em uma outra língua. Esta dissertação concentra-se na análise dos tipos de processos (verbos) nos quais Blanche DuBois se inscreve em quatro cenas (cenas um, dois, dez e onze), no texto original e no texto traduzido, de maneira a comparar os efeitos das escolhas feitas no texto de partida com aqueles efeitos produzidos no texto de chegada em termos da construção da personagem. A análise demonstra que, embora ocorra algumas diferenças locais na retextualização, estas diferentes realizações dos significados ideacionais na- o parecem afetar as configurações gerais da construção da protagonista na linguagem.

Lincoln Paulo Fernandes. "Tim Burton's The Nightmare before Christmas vs. O Estranho Mundo de Jack": A systematic Perspective of Subtitling. 01/09/1998

1v. 94p. Mestrado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LINGÜÍSTICA E LINGÜÍSTICA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca Central da Universidade Federal de Santa Catarina

Email do autor:

lico.fernandes@gmail.com

Palavras - chave:

Tradução, análise sistêmica, gênero, registro, linguagem

Área(s) do conhecimento:

Banca examinadora:

José Luiz Meurer

Maria Lúcia Barbosa de Vasconcellos

Walter Carlos Costa

Linha(s) de pesquisa:

Teoria e Prática de Tradução A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - a tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - DS

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

Legendação é uma importante área nos Estudos de Tradução. Entretanto, pesquisas feitas nesse campo são sempre voltadas aos aspectos técnicos que fazem parte desta atividade (legendagem), conseqüentemente, negligenciando aspectos lingüísticos e contextuais durante sua produção e recepção. Esta dissertação se propõe a examinar o film Tim Burton's The Nightmare before Christmas e sua versão legendada em Português do Brasil - O Estranho Mundo de Jack - objetivando investigar dois pontos lingüísticos: os contextos de cultura e situação, sugeridos por Steiner (1998) e Vasconcellos (1997), tanto no ambiente de partida quanto no de chegada e variação da linguagem (registro). Este estudo analisa as implicações da mudança de registro falado para registro escrito durante o processo de legendação e, também investiga as implicações e conseqüências de um tratamento diferente do gênero do filme no contexto de chegada. A pesquisa foi desenvolvida dentro de uma perspectiva sistemiótica da língua, fundamentada em Ventola (1988) e Eggins (1994). Argumenta-se que a mudança no canal de comunicação ("mode of discourse") afeta a produção de significados no texto traduzido e que o tratamento diferenciado do gênero no contexto de chegada afeta as dimensões do registro, que, em última instância, vem por afetar as escolhas lexicogramaticais na legendação.

Lincoln Paulo Fernandes. Brazilian Practices of Translating Names in Children's Fantasy Literature: A Corpus-based Study. 01/12/2004

1v. 189p. Doutorado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Central UFSC

Email do autor:

lico.fernandes@gmail.com

Palavras - chave:

Tradução, Literatura Infante-Juvenil, Procedimentos, Nomes Próprios

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA**Banca examinadora:**

Adriana Silvina Pagano

Cassio Rodrigues

Marco Antonio Esteves da Rocha

Maria Lúcia Barbosa de Vasconcellos

Markus Johannes Weininger

Stella Esther Ortweiler Tagnin

Linha(s) de pesquisa:

Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - DS

CNPq

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

O objetivo geral desta pesquisa é o de iluminar uma área que ainda permanece amplamente inexplorada dentro dos Estudos da Tradução (ET): Tradução de Literatura Infanto-Juvenil (TLIJ). Mais especificamente, este estudo investiga as práticas de traduzir nomes na literatura de fantasia infanto-juvenil dentro do contexto brasileiro durante o período 2000 e 2003. Para tal, o estudo se baseia num arcabouço teórico e metodológico extraído de três domínios dos ET: Tradução de Literatura Infanto-Juvenil, Estudos Descritivos da Tradução (EDT), e Estudos de Corpora em Tradução (ECT), os quais delimitam o escopo da pesquisa, seu corpus, sua linha de argumentação e seu desenvolvimento. TLIJ localiza o estudo dentro de uma área específica dentro dos ET, EDT fornece os conceitos e noções básicas informando o estudo, e ECT oferece as ferramentas metodológicas utilizadas na investigação de procedimentos tradutórios. O corpus paralelo eletrônico no qual o estudo se baseia consiste de 12 textos traduzidos e seus respectivos textos fonte, tirados de quatro séries de fantasia traduzidas no período, resultando um total de 24 textos (? 1,6 milhões de palavras). Argumenta-se que a análise de procedimentos é um dos necessários passos preliminares em se desvendar práticas tradutórias. Os resultados da análise de procedimentos revelaram a presença de duas principais tendências emergindo do corpus: a primeira em que as traduções privilegiam a legibilidade dos nomes, e a segunda em

que as traduções não priorizam a tradução deste específico elemento narrativo.

Ritalice Ribeiro de Medeiros. Subtitling as culture planning and representations of foreign lands: Rotten suburbs, bucks and raw sugar in the English subtitles of Terra Estrangeira, Central do Brasil and Abril Despedaçado. 01/06/2003

1v. 256p. Doutorado. UNIVERSIDADE FEDERAL DE SANTA CATARINA - LETRAS (INGLÊS E LITERATURA CORRESPONDENTE)

Orientador(es): Maria Lúcia Barbosa de Vasconcellos

Biblioteca Depositária: Biblioteca da UFSC

Email do autor:

Palavras - chave:

translation studies, equivalence

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

João Azenha Júnior

Maria Lúcia Barbosa de Vasconcellos

Marie Helene Torres

Viviane Maria Heberle

Linha(s) de pesquisa:

Tradução: Teoria e Prática A linha de pesquisa em tradução visa a: 1 - realização de pesquisas em teorias da tradução, meios e técnicas de tradução, e tipologia de textos (literários, científicos e técnicos), 2 - tradução de obras literárias, científicas e técnicas.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CNPq

Idioma(s):

Inglês

Dependência administrativa

Federal

Resumo tese/dissertação:

Várias pesquisas em legendagem são voltadas para determinados aspectos inerentes à atividade, tais como os aspectos mecânicos de sua produção e as limitações técnicas que aí operam. Sem desconsiderar a importância de tais aspectos, este trabalho propõe uma visão de legendagem como prática cultural, sempre inevitavelmente inserida em contextos interculturais mais amplos. Por meio de uma metodologia soft, esta tese investiga

procedimentos de legendagem adotados na tradução para o inglês de termos culturais presentes nos diálogos dos filmes brasileiros Terra Estrangeira, Central do Brasil e Abril Despedaçado. Em um primeiro momento, a Teoria da Relevância fornece as ferramentas teóricas para se analisar possíveis suposições dos legendadores em relação às expectativas das audiências no que diz respeito à legendagem de termos culturais. Em seguida, outros conceitos teóricos advindos dos Estudos Culturais, bem como de outros campos correlatos, são utilizados como ferramentas para uma análise crítica dos procedimentos de legendagem previamente identificados na análise baseada na Teoria da Relevância. No contexto desta pesquisa, os legendadores são vistos como planejadores de cultura em potencial, à medida que interferem nas representações que as audiências têm de componentes culturais estrangeiros, por meio de, por exemplo, procedimentos abusivos de legendagem. Conseqüentemente, entende-se que os legendadores podem também interferir em relacionamentos interculturais de maior âmbito.

Anexo 5: Orientações de Dra. Adriana Pagano até 2011

Titulo	Autor	Ano	Orientador	IFES e Tipo	Area
1) Conhecimento de domínio e expertise em tradução: contribuições de um estudo comparado entre tradutores profissionais e pesquisadores juniores e seniores no desempenho de tarefas de tradução	Maria Luísa de Oliveira Nao encontrado no Portal da Capes	2009	Adriana Pagano	UFMG MA	
2) Conhecimento experto em tradução: orientação e revisão em tarefas tradutórias executadas por pesquisadores expertos não tradutores	Kelen Cristina Sant Anna de Lima	2008			MA
3) Indagando o perfil de tradutores em formação: um estudo de caso	Camila Nathália de Oliveira Braga	2007			
4) Conhecimento experto em tradução: aferição da durabilidade de tarefas tradutórias realizadas por sujeitos não-tradutores em condições	Igor Antônio Lourenço da Silva	2007			

empírico-experimentais					
5) Academic writing: estudo de livros didáticos para o ensino/aprendizagem de escrita acadêmica em inglês sob a perspectiva de gêneros do discurso	Tatiana de Macedo	2003			MA
6) O discurso terapêutico e a análise crítica do discurso	Marilene Tavares Cortez	2003			MA
7) O gênero notícia científica no jornal televisivo brasileiro	Janaína Minelli de Oliveira	2001			
8) As Vozes da Ciência: a representação do discurso nos gêneros artigo acadêmico e de divulgação científica	Janaína Minelli de Oliveira	2005			PhD
9) O diálogo nos livros de ensino de língua inglesa: uma representação dos gêneros do discurso oral natural?	Kátia Honório do Nascimento	2000			MA
10) . Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues	Giacomo Patrocínio Figueredo Nao encontrado no Portal da Capes	2011			PHD

11) Escrita acadêmica e avaliação: o uso de reforços e atenuadores em artigos científicos publicados em inglês por pesquisadores brasileiros	Mônica Soares de Araújo Guimarães	2009			Phd
12) O metadiscorso interpessoal em artigos acadêmicos: espaço de negociações e construção de posicionamentos	Cibele Gadelha Bernardino	2007			Phd
13) A citação enquanto recurso de afiliação acadêmica	Tatiana do Socorro Chaves Lima de Macedo	2006			Phd

Resumos

Kelen Cristina Sant'Anna de Lima. Conhecimento experto em tradução: orientação e Conhecimento experto em tradução: orientação e pesquisadores expertos não-tradutores.. 01/10/2008

1v. 187p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: FALE; BU

Email do autor:

Palavras - chave:

tradução; conhecimento experto

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Fabio Alves da Silva Junior

José Luiz Vila Real Gonçalves

Linha(s) de pesquisa:

ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta dissertação está inserida no projeto Expert@ - Conhecimento experto em tradução: modelagem do processo tradutório em altos níveis de desempenho (CNPq 479340/2006-4), em desenvolvimento no Laboratório Experimental de Tradução (LETRA), da Faculdade de Letras da Universidade Federal de Minas Gerais. Apresenta um estudo das fases de orientação inicial e revisão final incluindo pausas de orientação em tempo real e de revisão em tempo real (e respectivos apoios externos e internos) dos processos tradutórios de quatro sujeitos pesquisadores expertos em medicina, quando da realização de duas tarefas tradutórias, em condições empírico-experimentais. Dando-se continuidade à descrição do perfil de expertos não-tradutores e à investigação acerca do impacto da variável independente conhecimento de domínio (SCARDAMALIA; BEREITER, 1991) sobre a realização de tarefas tradutórias, ambas iniciadas por Silva e Pagano (2007), esta pesquisa analisa os dados obtidos por esses autores para lançar luz sobre esse perfil tradutório sob a perspectiva de questões relativas à implicação cognitiva na orientação (MACHADO; ALVES, 2007) e na revisão (BATISTA; ALVES, 2007) desses sujeitos. A coleta de dados, descrita pormenorizadamente em Silva e Pagano (2007), consistiu em: (i) um questionário prospectivo; (ii) duas traduções, por sujeito, de introduções de artigos acadêmicos, da língua portuguesa para a língua inglesa (sendo que, em uma delas, o texto era correlato ao conhecimento de domínio vinculado à sua subárea de atuação dos sujeitos e, na outra, o conhecimento de domínio demandado pelo texto de partida não era congruente com a subárea do conhecimento em que o sujeito atua); e (iii) relatos retrospectivos coletados imediatamente após a realização das tarefas tradutórias. Essas tarefas foram registradas eletronicamente a partir da utilização dos programas Translog© e Camtasia©, havendo também o registro manual por meio de planilhas de observação direta, com base naquelas já aplicadas pelo grupo PACTE (2005). O procedimento para análise adotado nesta pesquisa consistiu na replicação parcial da metodologia de análise das pausas iguais ou superiores a cinco minutos identificadas no arquivo log gerado pelo Translog©, a qual fora

introduzida por Machado e Alves (2007), no que toca à orientação inicial e à orientação em tempo real, e por Batista e Alves (2007), com relação à orientação inicial e à orientação em tempo real. O processo de tradução dos sujeitos sob escrutínio foi estudado considerando-se as seguintes variáveis dependentes, que podem ser correlacionadas com a noção de durabilidade (ALVES; GONÇALVES, 2007): pausas (ALVES, 2005) (número de ocorrências; duração por ocorrência e duração total, juntamente com as médias e os desvios padrões, além dos valores máximos e mínimos) e recursividade (BUCHWEITZ; ALVES, 2006) atrelada a essas pausas (considerando-se o número de ocorrências de recursão para correção ortográfica, para substituição de decisões provisórias imediatamente descartadas e para a alteração de soluções duráveis), além de verificação qualitativa do(s) tipo(s) de apoio interno e/ou externo utilizado(s) para cada ocorrência de pausa. Os resultados desta pesquisa apontam características do perfil estudado que corroboraram dados de Silva e Pagano (2007) em relação a pausas indicativas de orientação, verificando-se impacto da variável conhecimento de domínio em relação ao número de ocorrências e à duração das pausas de orientação. Os resultados também indicam que os expertos não-tradutores apresentaram tendências comparáveis àquelas do perfil dos tradutores profissionais analisados por Machado e Alves (2007) e Batista e Alves (2007) no que diz respeito às fases do processo tradutório e, mais especificamente, à alocação de pausas de orientação e de revisão em tempo real associadas a mecanismos de apoio interno e/ou externo. A pesquisa evidenciou ainda a presença de teclas ou movimentos de recursão (sobretudo indicativas de correções ortográficas durante ou imediatamente após uma solução provisória) atrelados a pausas de orientação (e não de revisão) entre os sujeitos sob escrutínio. Alguns dos achados - especialmente o tempo dedicado à fase de orientação inicial, a variedade de apoios internos/externos de orientação em tempo real e o número de pausas na fase de revisão final - também apontam, tal qual observado por Silva e Pagano (2007) para a mesma amostra, o perfil diferenciado de um dos sujeitos, em razão de uma tarefa mais durável e de um nível superior de desempenho.

Camila Nathália de Oliveira Braga. Indagando o perfil de tradutores em formação: um estudo de caso. 01/03/2007

1v. 140p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

metalinguagem, tradução, linguística aplicada

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano
Fabio Alves da Silva Junior
José Luiz Vila Real Gonçalves

Linha(s) de pesquisa:

ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - PROF

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta dissertação apresenta um estudo de caso que visa a indagar o perfil de tradutores em formação com vistas a caracterizar seu desempenho no continuum novato-experto. Para tanto, cinco alunos da disciplina Tradução II, da Faculdade de Letras da UFMG, foram os sujeitos desta pesquisa que acompanhou seu desempenho na realização de três tarefas de tradução: uma no início, uma no meio e uma no final da disciplina. Para cada tarefa foi selecionado um texto originalmente escrito em língua inglesa, completo e/ou independente, pertencente a um tipo de texto específico, e selecionados com base em uma progressão de complexidade em relação às relações lógico-semânticas dos complexos oracionais (HALLIDAY & MATTHIESSEN, 2004), progressão esta definida pelo número de projeções nos três textos (0, 3 e 12, respectivamente). A fim de investigar o perfil dos sujeitos, a primeira pergunta de pesquisa teve como objetivo verificar que dados podem ser aportados pela análise da produção textual com foco nos complexos oracionais que se apresentam como problemáticos, dado o número e a extensão de pausas registradas durante a execução da tarefa tradutória. Essa verificação foi feita através da análise de algumas unidades que suscitaram maiores pausas quando de sua tradução, devido ao fato de serem configurados por complexos oracionais com múltiplas relações lógico-semânticas e de taxa intrincadas (HALLIDAY & MATTHIESSEN, 2004). A segunda pergunta de pesquisa teve como objetivo verificar dados relativos aos tempos despendidos em cada fase do processo tradutório (JAKOBSEN, 2002, 2003), as instâncias de metalinguagem e meta reflexão (ALVES, 2005; ALVES & GONÇALVES, no prelo) nos protocolos verbais e as estratégias de apoio externo utilizadas por cada sujeito. Por último, a terceira pergunta formulada neste estudo visou observar o impacto da instrução formal no processo e no produto tradutórios dos sujeitos, com base em dados obtidos

nos protocolos verbais, relacionados à percepção do papel dos conteúdos da disciplina e sua aplicação na prática efetiva de análise textual durante as tarefas tradutórias. Os resultados permitiram localizar os sujeitos em diferentes pontos do continuum tradutor novato-tradutor experiente de acordo com a caracterização do perfil dos mesmos: três sujeitos apresentaram um perfil que pode ser localizado num estágio intermediário do continuum, enquanto que dois deles apresentaram perfil claramente apontado pela literatura como de tradutores novatos. A observação da produção textual com base na resolução de problemas vinculados aos complexos oracionais evidenciou escassa meta reflexão e ausência de metalinguagem com base em uma teoria lingüística que permita aos sujeitos formular os problemas enfrentados. Dados relativos ao ritmo cognitivo dos sujeitos não foram facilmente correlacionáveis aos perfis dos dois grupos, mostrando que tempo de orientação e revisão não podem ser tomados como critérios necessariamente ser associados a um determinado tipo de desempenho. No que diz respeito ao impacto da instrução formal centrada em tópicos de análise textual, a pesquisa mostra que alguns dos conceitos discutidos durante o curso são mencionados nas verbalizações dos sujeitos nos protocolos retrospectivos, embora o desempenho dos sujeitos não tenha apresentado mudanças significativas ao longo das três coletas efetuadas.

Igor Antônio Lourenço da Silva. Conhecimento Experto em Tradução: aferição da durabilidade de tarefas tradutórias realizadas por sujeitos não-tradutores em condições empírico- experimentais. 01/11/2007

1v. 267p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

tradução; conhecimento experiente, linguística aplicada

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

Fabio Alves da Silva Junior

José Luiz Vila Real Gonçalves

Ronaldo Teixeira Martins

Linha(s) de pesquisa:

ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - PROF

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta dissertação está vinculada ao projeto EXPERT@ – Conhecimento experto em tradução: modelagem do processo tradutório em altos níveis de desempenho (CNPq 479340/2006-4), desenvolvido no LETRA (Laboratório Experimental de Tradução), na Faculdade de Letras da UFMG. São apresentadas três abordagens distintas e complementares ao mesmo objeto de estudo, qual seja: duas tarefas tradutórias realizadas em ambiente Translog©, sob condições experimentais, por quatro pesquisadore(a)s experto(a)s da área da medicina. Na primeira abordagem, que parte da proposta de Shreve (2006) de interface entre os estudos da tradução e os estudos sobre expertise e desempenho experto, o impacto da variável independente do experimento, conhecimento de domínio (SCARDAMALIA; BEREITER, 1991), sobre a realização das tarefas tradutórias é aferido a partir dos seguintes parâmetros (variáveis dependentes): pausas e recursividade (ALVES, 2005; BUCHWEITZ; ALVES, 2006), segmentação (DRAGSTED, 2004, 2005), ritmo cognitivo (JAKOBSEN, 2002), representação da tarefa (CHI, 2006b) e durabilidade (ALVES; GONÇALVES, 2007). Nessa abordagem, verifica-se que, para a amostra sob escrutínio, (i) há impacto significativo do conhecimento de domínio sobre a recursividade do processo tradutório, a fase de orientação desse processo e a extensão dos segmentos textuais e (ii) a durabilidade da tarefa tradutória do(a)s pesquisadore(a)s está vinculada à representação da tarefa tradutória e ao projeto tradutório implementado pelos sujeitos, os quais podem ser correlacionados com o tipo de segmentação observado no processo tradutório. Na segunda abordagem, que busca coadunar o conceito de durabilidade (ALVES; GONÇALVES, 2007) com o arcabouço teórico-metodológico da Teoria da Estrutura Retórica (RST) (TABOADA; MANN, 2006a, 2006b), constata-se que a coerência de um texto de chegada não está necessariamente correlacionada com a produção de relações retóricas com base naquelas encontradas nos níveis inferiores do texto de partida. Assim, escolhas nos níveis inferiores do texto de chegada vinculadas a realizações léxico-gramaticais que determinam os tipos de relações retóricas ou o componente lógico dos complexos oracionais não comprometem níveis superiores da estrutura retóricas. Os resultados obtidos apontam para uma configuração diferenciada na tarefa executada por um dos sujeitos do experimento, a qual pode ser correlacionada com maior durabilidade e um perfil de desempenho superior. Haja vista o destaque desse sujeito em particular, a sua produção textual é analisada na terceira abordagem desta pesquisa, em que se analisa o papel da (des)metaforização (STEINER, 2001a, 2002) como um processo de compreensão subjacente à tradução (entendida como produção

orientada de significado). Observa-se, nessa abordagem, que tanto o texto de chegada como o texto de partida são, do ponto de vista da Gramática Sistêmico-Funcional, altamente metafóricos, sendo a (des)metaforização pautada no conhecimento de domínio e/ou no conhecimento discursivo desse sujeito um importante processo para a produção de seus textos de chegada. Constatam-se ainda pausas de orientação ao início de cada complexo oracional e revisões que se dão concomitantemente com a fase de redação, de modo tal que o sujeito chega, ao final da fase de redação, com um produto tradutório altamente durável (como mostrado na abordagem anterior, em termos de estrutura retórica).

Tatiana do Socorro Chaves Lima de Macedo. Academic writing: estudo de livros didáticos para o ensino/aprendizagem de escrita acadêmica em inglês sob a perspectiva de gêneros do discurso.. 01/05/2003

1v. 183p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG

Email do autor:

Palavras - chave:

livros didáticos, língua inglesa

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

LÍNGUAS ESTRANGEIRAS MODERNAS

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Maria Cristina Faria Dalacorte

Linha(s) de pesquisa:

ENSINO/APRENDIZAGEM DE LÍNGUAS

ESTRANGEIRAS Ensino/aprendizagem de LE, incluindo o português como LE, envolvendo seus processos, sua mediação por computador e sua avaliação; o desenvolvimento e análise de abordagens, técnicas e materiais didáticos; discurso e ensino; formação de professor.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - PICDT

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta dissertação tem por objetivo analisar livros didáticos para ensino/aprendizagem de escrita acadêmica em língua inglesa sob a perspectiva do conceito de gêneros do discurso (Swales, 1991). Para esse fim, foram selecionados nove livros didáticos publicados ou reimpressos na década de 1990 e disponíveis no mercado, passíveis de serem utilizados em cursos de Letras em universidades brasileiras. Os referidos livros foram analisados focalizando-se a incorporação, explícita ou implícita, da noção de gênero do discurso em sua formulação; o tratamento dado aos gêneros do discurso presentes nos livros; o conceito de escrita acadêmica subjacente à elaboração das atividades; e a autenticidade dos materiais contidos nos referidos livros. Foi realizado um mapeamento de aspectos desenvolvidos por cada livro em sua abordagem da escrita acadêmica. Da mesma forma, foi feito um mapeamento dos pressupostos teóricos que fundamentam os livros didáticos em relação ao ensino/aprendizagem da escrita acadêmica, em especial ao ensino/aprendizagem baseados em gêneros do discurso. Os resultados sugerem que o impacto das teorias de gênero do discurso na elaboração de livros didáticos para o ensino/aprendizagem de escrita acadêmica é ainda incipiente e que o uso do termo gênero do discurso apresenta uma grande diversidade conceitual com afiliações teóricas também diversas. A análise evidencia, ainda, uma forte tendência em se considerar o ensino/aprendizagem da escrita acadêmica como o desenvolvimento de um único gênero, o ensaio, o que revela uma concepção da escrita acadêmica vinculada, sobretudo, aos discursos literário e jornalístico. Nas considerações finais deste trabalho, são apresentadas reflexões sobre a necessidade de se introduzir uma abordagem baseada em gêneros do discurso nos cursos de Letras que possibilite o desenvolvimento de uma consciência retórica por parte do aluno e que permita ampliar o conceito de escrita acadêmica, contemplando-se outros gêneros do discurso além do ensaio.

Marilene Tavares Cortez. O Discurso Terapêutico e a Análise Crítica do Discurso. 01/07/2003

1v. 231p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG

Email do autor:

lenacortez@uol.com.br

Palavras - chave:

discurso terapêutico, Sociolingüística Interacional

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Maria Izabel Santos Magalhães

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO Abordagens variadas do discurso, refletindo a diversidade de fontes, de instrumentais teóricos, de metodologias, bem como a natureza diversa do objeto-discurso e os múltiplos interesses que nele são projetados.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - DS

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta dissertação examina abordagens do discurso terapêutico à luz da Análise Crítica do Discurso proposta por Norman Fairclough. A partir da releitura de dois modelos de análise do discurso terapêutico - o primeiro, desenvolvido por Labov e Fanshel (1977) no viés da Análise do Discurso e o segundo proposto por Ribeiro (1993; 2002), na vertente da Sociolinguística Interacional, são discutidas questões em torno da constituição da identidade na modernidade tardia. Para a discussão a partir do referencial teórico da ACD, alguns de seus conceitos iniciais (Fairclough, 1992) foram reexaminados, para, num segundo momento, mostrar como estes conceitos foram modificados a partir de uma maior aproximação desse campo disciplinar com os estudos histórico-culturais (Fairclough e Chouliaraki, 1999). Este refinamento teórico possibilitou o emprego da metodologia estabelecida pela ACD para revisar as análises propostas por Labov e Fanshel e Ribeiro. A reanálise do modelo proposto por Labov e Fanshel aponta para algumas contribuições fundamentais destes autores, principalmente a recuperação da importância do papel da entonação para a produção de significado no discurso, bem como na constituição da identidade. A revisão da análise proposta por Ribeiro permite que se fundamente a necessidade do desenvolvimento de um enfoque histórico, crítico, sobre a análise discursiva, para que ela não reproduza discursos hegemônicos, como o da psicanálise. Como os modelos examinados foram construídos com base numa abordagem terapêutica psicanalítica, evidenciou-se, também, na reanálise, o caráter hegemônico desta proposta terapêutica, que passou a exercer, na modernidade tardia, a função de normatização e controle das abordagens sobre a constituição da identidade.

Janaina Minelli de Oliveira. O gênero informação científica transmitida através do jornal televisivo brasileiro. 01/05/2001

1v. 156p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: Biblioteca da FALE - UFMG

Email do autor:

Palavras - chave:

informação científica, jornal, televisão

Área(s) do conhecimento:

EDUCAÇÃO

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Desirée Motta Roth

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO As pesquisas examinam a produção, recepção e circulação do discurso, sob os enfoques enunciativo, organizacional e interacional. Estas pesquisas orientam-se pelas vertentes teóricas francesa e anglo-americana.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

A dissertação "O gênero informação científica transmitida através do jornal televisivo brasileiro" representa um esforço de caracterização da informação científica veiculada por telejornais a partir dos subsídios teóricos da Análise de Gêneros e Análise do Discurso, em sua vertente anglo-americana. Fazem parte de seu corpus notícias sobre ciência veiculadas pelos jornais Jornal da Noite e Jornal da Bandeirantes, da rede Bandeirantes de Televisão, Jornal Nacional e Jornal da Globo, da Rede Globo de Televisão. Tomando como ponto de partida as reflexões sobre gênero discursivo propostas por SWALES (1993), este trabalho examina a rescrição intersemiótica do saber científico. Fundamentando-se nas reflexões traçadas por SCOLLON (1998) sobre o discurso jornalístico, o presente estudo analisa a audiência alvo do gênero ora descrito e a especificidade dos contextos de produção e recepção da informação sobre ciência. Para uma melhor caracterização das estratégias de apropriação de tal informação, buscou-se em ELLIS (1992) e THOMPSON (1995) embasamento teórico para a análise do modo narrativo instaurado pela televisão e da especificidade da semi-interação, característica da comunicação de massa. Os resultados apontaram que a tessitura do gênero informação científica transmitida através do jornal televisivo brasileiro

revela como seu objetivo sócio-retórico a modalização de informação científica para uma comunidade de prática não especialista em ciência, idealizada em um espectador potencial. Os eventos comunicativos do gênero são curtos, têm um tom informal e dialogam com outras formas discursivas, como o cinema, por exemplo, e com outros discursos, como o religioso e o bélico. Tais traços são consequência da aproximação entre o discurso jornalístico e o discurso do entretenimento, aproximação esta que se revela como uma tendência mundial de comunicação de massa.

Janaina Minelli de Oliveira. As Vozes da Ciência: a representação do discurso nos gêneros artigo acadêmico e de divulgação científica. 01/09/2005

1v. 244p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

gêneros, discurso, conhecimento científico,

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

Carmen Rosa Caldas-Coulthard

Célia Maria Magalhães

Leila Bárbara

Maria Carmen Aires Gomes

Linha(s) de pesquisa:

ESTUDOS DA LINGUAGEM, IDENTIDADE E REPRESENTAÇÃO Esta linha reúne pesquisas com base em corpora de textos monolíngües, em formato eletrônico e não eletrônico, a partir da noção de linguagem como representação e de linguagem como semiótica social. A fundamentação teórica adotada contempla a lingüística s

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - PROF

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta tese apresenta uma investigação sobre a representação do discurso em um corpus de aproximadamente 35 milhões de palavras, constituído por textos em português do Brasil, vinculados a dois gêneros do discurso: o

gênero artigo acadêmico e o gênero artigo de divulgação científica. Partindo de uma perspectiva sistêmico-funcional da língua, a presente tese se concentra na realização de sistemas léxico-gramaticais relacionados à representação do discurso direto, que encenam relações interpessoais e sociais. A revisão da literatura e a análise empírica, ora apresentadas, sugerem que o comportamento citacional dos Escritores dos textos, vinculados aos gêneros pesquisados, negocia distância discursiva e é influenciado pelo status cultural e historicamente atribuído ao conhecimento científico. Tanto no gênero artigo acadêmico como no gênero artigo de divulgação científica, os Escritores posicionam-se hierarquicamente inferiores aos Autores citados, mas nota-se, no gênero artigo acadêmico, a manipulação de aspectos léxico-gramaticais, para a redução da distância discursiva. Há indícios de que o sistema da interdependência oracional, na projeção, passa por um processo semogênico, que realiza significado interpessoal no âmago da comunidade acadêmica.

Kátia Honório do Nascimento. O diálogo dos livros didáticos de língua inglesa: uma representação dos gêneros do discurso oral natural. 01/09/2000

1v. 154p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: FALE / UFMG

Email do autor:

Palavras - chave:

Discurso, livro didático, lingüística

Área(s) do conhecimento:

EDUCAÇÃO

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Deise Prina Dutra

Maria Cristina Faria Dalacorte

Linha(s) de pesquisa:

ENSINO/APRENDIZAGEM DE LÍNGUAS ESTRANGEIRAS Esta linha integra projetos sobre aquisição de línguas estrangeiras modernas, incluindo o português com língua estrangeira. As pesquisas concentram-se em questões tais como: estratégias de aprendizagem, desenvolvimento e análise de métodos.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

A presente pesquisa visa desenvolver uma abordagem do ensino de inglês como língua estrangeira sob uma perspectiva interdisciplinar, utilizando-se dos subsídios da Linguística Aplicada ao Ensino/Aprendizagem de Inglês como Língua Estrangeira e da Análise de Gêneros Discursivos. Tomando como ponto de partida as pesquisas realizadas por Dalacorte (1991) e Chiaretti (1993), que comparam a conversação espontânea e os diálogos nos livros didáticos de ensino de inglês publicados na década de 80, este estudo analisou os diálogos dos livros didáticos publicados nos anos 90, com o objetivo de observar se houve mudanças significativas na elaboração dos mesmos num momento em que a abordagem comunicativa já se encontra consolidada e a abordagem de autonomia do aprendiz vem ganhando progressiva aceitação por parte de teóricos e elaboradores de materiais de ensino. Num segundo momento, foram utilizados subsídios teóricos advindos da Análise de Gêneros Discursivos para se obter uma melhor definição do diálogo do livro didático enquanto gênero discursivo híbrido, diferente da conversação espontânea e, portanto, passível de ser avaliado, utilizando-se uma perspectiva diferente no que diz respeito à autenticidade e ao seu suposto caráter de reprodução da conversação espontânea. Os resultados apontaram para uma melhoria na construção dos diálogos nos livros didáticos publicados na década de 90, especialmente no que diz respeito à incorporação de aspectos que levem o aprendiz a refletir sobre o discurso oral nas linhas teóricas desenvolvidas pela abordagem de autonomia da aprendizagem. Alguns dos diálogos, contudo, mostraram que sua elaboração ainda se pauta pela concepção dos mesmos enquanto atividades gramaticais. A adoção de subsídios da Análise de Gêneros Discursivos enquanto abordagem do diálogo do livro didático revelou ser de grande utilidade, uma vez que aponta para a necessidade de se considerar o diálogo como gênero discursivo que guarda uma especificidade própria, diferente daquela dos gêneros do discurso oral espontâneo. Esta pesquisa sugere que os diálogos do livro didático não devem ser considerados como única fonte de exposição do aprendiz à língua estrangeira oral e que os livros didáticos poderiam apontar diversas atividades além do diálogo para oferecer ao aprendiz formas de acesso ao discurso oral espontâneo em língua inglesa.

Mônica Soares de Araújo Guimarães. Escrita acadêmica e avaliação: o uso de reforços e atenuadores em artigos científicos publicados em inglês por pesquisadores brasileiros. 01/12/2009

1v. 130p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: Biblioteca Universitária da UFMG

Email do autor:

Palavras - chave:

Artigos Acadêmicos; , Discurso; Pesquisadores

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Daisy Rodrigues Vale

Sueli Maria Coelho

Vera Lúcia Santiago Araújo

Linha(s) de pesquisa:

ESTUDOS EM LÍNGUAS ESTRANGEIRAS:
ENSINO/APRENDIZAGEM, USOS E CULTURAS Esta linha integra projetos sobre o uso e a aquisição de línguas estrangeiras (LEs) modernas, incluindo o português como LE.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:**Idioma(s):**

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Estudos sobre recursos modalizadores como os atenuadores/evasivas (hedges) e reforços (boosters) mostram a sua relevância para a escrita acadêmica. Esta tese investiga a utilização, por parte de pesquisadores afiliados a centros nacionais de pesquisa, o uso de reforços e de evasivas em seus artigos acadêmicos publicados em inglês e os compara a textos publicados por pesquisadores afiliados a centros internacionais de pesquisa que publicam seus textos em inglês. Para tanto, com base em Swales (1993, 1996, 2000), Hyland (1996 a, 1996b, 2000, 2004), Salager-Meyer (1994), analisamos dois subcorpora compostos por trinta (30) artigos publicados por cada grupo de pesquisadores. As realizações léxico-gramaticais dos reforços e das evasivas enfocadas foram selecionadas de uma lista compilada por Hyland (2000). Estas realizações foram identificadas nas seções retóricas que compõem os artigos acadêmicos e categorizadas em verbos lexicais, verbos modais, advérbios, adjetivos e substantivos. Os reforços e as evasivas foram classificados seguindo a proposta de Salager-Meyer (1994) como escudos ou como aproximadores. Nossa investigação indica que os pesquisadores de ambos os subcorpora utilizam reforços e evasivas em seus textos; entretanto, os pesquisadores afiliados a centros internacionais de pesquisa utilizam mais esses recursos que os pesquisadores afiliados a centros nacionais de pesquisa. Quanto a análise da estrutura superficial, ou seja, a análise dos itens

léxicogramaticais nas sentenças, ao computar os dados conjuntamente nos dois subcorpora, temos as mesmas classes utilizadas em ambos. Percebemos que ao analisar as classes como evasivas os pesquisadores dos dois subcorpora também usaram os mesmos recursos. Entretanto, as classes utilizadas pelos pesquisadores em ambos os subcorpora foram diferentes quanto ao uso dos reforços. Os reforços e as evasivas apresentam a função de escudos e/ou aproximadores (cf. Salager-Meyer, 1994) em ambos os subcorpora. Concluímos que as escolhas de uso dos reforços e das evasivas estão relacionadas ao propósito comunicativo de cada uma das seções retóricas dos artigos, entretanto, os pesquisadores brasileiros afiliados a centros nacionais de pesquisa tendem a utilizar este recurso de forma divergente no que diz respeito às seções dos artigos de pesquisa, uma vez que a literatura aponta a seção dos Materiais e Métodos com poucas evasivas e nos dois subcorpora este dado é divergente.

Cibele Gadelha Bernardino. O Metadiscurso Interpessoal em Artigos Acadêmicos: espaço de negociações e construção de posicionamentos. 01/06/2007

1v. 243p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS
Orientador(es): Adriana Silvina Pagano
Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

Metadiscurso, retórica, linguística aplicada

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano
Célia Maria Magalhães
Daisy Rodrigues do Vale
Heliana Ribeiro de Mello
Vera Lúcia Santiago Araújo

Linha(s) de pesquisa:

ESTUDOS DA LINGUAGEM, IDENTIDADE E REPRESENTAÇÃO Esta linha reúne pesquisas com base em corpora de textos monolíngües, em formato eletrônico e não eletrônico, a partir da noção de linguagem como representação e de linguagem como semiótica social.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta tese investiga a construção do metadiscorso interpessoal em exemplares de artigos acadêmicos produzidos por autores(as) brasileiros(as), visando indagar como os marcadores metadiscursivos interpessoais (HYLAND, 2000) realizados por adjuntos modais (HALLIDAY, 1994) são utilizados pelos(as) autores(as) brasileiros(as) da área de lingüística em exemplares de artigos acadêmicos experimentais, teóricos e de revisão de literatura (SWALES, 2004). Para cumprir tal objetivo, investigamos um corpus composto por 10 exemplares de artigos experimentais, 10 exemplares de artigos de revisão de literatura e de 10 exemplares de artigos teóricos, compilados a partir de exemplares do periódico D.E.L.T.A. disponíveis no site <http://www.scielo.br>, e publicados no período compreendido entre os anos de 1997 e 2004. Para o levantamento e mapeamento dos adjuntos modais que funcionaram como marcadores interpessoais foi feita uma anotação manual do corpus que permitiu extrair dados quantitativos com o programa WordSmith Tools. No primeiro momento de análise, os advérbios que funcionaram como adjuntos modais foram manualmente anotados, depois fizemos uma classificação dos tipos de adjuntos segundo Halliday (1994) e Halliday & Matthiessen (2004) em diálogo com abordagens funcionais do português (NEVES, 2000). Em um terceiro momento, identificamos os adjuntos modais que funcionaram como marcadores interpessoais (HYLAND, 1998/2000). Por fim, uma análise mais detalhada de cada tipo de artigo e de suas unidades retóricas foi realizada para examinar os marcadores interpessoais em seus co-textos. A análise dos dados apontou para os seguintes resultados: (a) nos três corpora o marcador metadiscursivo mais utilizado pelos(as) autores(as) foi o marcador atributivo realizado por adjuntos modais de validade, intensidade e usualidade; (b) Os marcadores metadiscursivos de ênfase foram realizados por adjuntos modais de suposição, persuasão, probabilidade, obviedade e intensidade e o número de ocorrências desses marcadores foi muito próximo nas três categorias de artigos, sendo levemente maior em artigos teóricos (c) os marcadores metadiscursivos de atenuação foram realizados, nos três corpora, por adjuntos modais de probabilidade e suposição, com predomínio absoluto dos adjuntos de probabilidade. O número de marcadores de atenuação foi bastante superior nos exemplares dos artigos experimentais se comparado ao número desses marcadores em artigos teóricos e de revisão de literatura; e (d) os marcadores metadiscursivos atitudinais realizados pelos adjuntos modais de desejo, predição e intensidade foram os menos utilizados em todas as três categorias de artigos. É importante salientar que os(as) autores(as) de artigos experimentais utilizaram mais marcadores metadiscursivos que os(as) autores(as) de artigos teóricos e de revisão de literatura. Esta distinção deve-se, principalmente, às diferenças quanto ao uso dos marcadores metadiscursivos interpessoais de atenuação nos três corpora, uma vez que

esse tipo de recurso metadiscursivo foi, preferencialmente, utilizado nos artigos experimentais. Este fato deve-se à presença marcante desses marcadores na unidade retórica Resultados e Discussão, que é típica dos artigos experimentais. Isto indica que em contraste com artigos teóricos e de revisão de literatura, em artigos experimentais, os(as) autores(as) parecem ter uma maior preocupação em atenuar a força asseverativa de suas afirmações, assim como em apresentar-se como fonte de suas proposições, deixando um espaço de negociação mais amplamente aberto com seus pares da comunidade disciplinar.

Tatiana do Socorro Chaves Lima de Macedo. A citação como recurso de afiliação acadêmica. 01/12/2006

1v. 221p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Adriana Silvina Pagano

Biblioteca Depositária: FALE e CENTRAL

Email do autor:

Palavras - chave:

ensino/aprendizagem, retórica, escrita acadêmica

Área(s) do conhecimento:

LÍNGUAS ESTRANGEIRAS MODERNAS

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Desirée Motta Roth

Maralice de Souza Neves

Viviane M. Heberle

Linha(s) de pesquisa:

ESTUDOS EM LÍNGUAS ESTRANGEIRAS:

ENSINO/APRENDIZAGEM, USOS E CULTURAS Esta linha integra projetos sobre o uso e a aquisição de línguas estrangeiras (LEs) modernas, incluindo o português como LE.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta tese apresenta um estudo de citações em escrita acadêmica na perspectiva da Análise de Citações e do conceito de gêneros do discurso (Moravcsik & Murugesan, 1975; Thompson & Yiyun, 1991; Swales, 1986,

1991, 2004; Bhatia, 2004). A pesquisa enfoca a citação no contexto da produção de textos por membros expertos e membros novatos da comunidade acadêmica de Lingüística e consistiu na análise e comparação de citações em nove artigos acadêmicos e treze trabalhos finais de disciplina, da área de Lingüística. Os artigos foram publicados no periódico on-line DELTA, entre 1999 e 2004; os trabalhos finais de disciplina (neste caso, projetos de pesquisa) foram escritos por treze alunos de um curso de pós-graduação do Brasil, em 2004. Os resultados foram comparados entre si para que fossem observadas diferenças e semelhanças na ocorrência de citações em ambos os gêneros analisados (artigos acadêmicos e trabalhos finais de disciplina). Esses gêneros foram escolhidos por duas razões principais: a) eles estão vinculados a uma mesma comunidade discursiva e possuem características em comum; b) um desses gêneros, o artigo acadêmico, é reconhecido pela comunidade acadêmica como um meio de comunicação importante, e constitui um dos modelos de escrita acadêmica. Os resultados demonstram que as escolhas lingüísticas que guiam a construção de citações são compartilhadas pelos membros expertos (autores dos artigos acadêmicos) e pelos novatos (alunos do curso de pós-graduação). Entretanto, a justificativa para o uso de citações pelos membros novatos revela desconhecimento das razões pelas quais a comunidade acadêmica faz uso de citações em gêneros do discurso acadêmico escrito. A análise também mostra que alguns dos alunos não reconhecem as citações quando estas se apresentam em forma de paráfrase. As conclusões deste estudo apontam para a necessidade de os cursos de graduação e pós-graduação adotarem uma abordagem de ensino/aprendizagem baseada em gêneros visando a desenvolver a consciência retórica dos alunos em relação à escrita acadêmica e, conseqüentemente, ao uso de citações em textos escritos, desenvolvendo, dessa forma, o posicionamento autoral dos alunos em sua comunidade discursiva.

Anexo 6: Orientações de Célia Magalhães até 2011

Titulo	Autor	Ano	Orientador	IFES	Tipo
14) O estilo de tradutores: apresentação do discurso no corpus paralelo Heart of Darkness/(No)Coração das trevas	Carolina Pereira Barcellos Nao se encontra no Banco de Teses da Capes	2011	Célia Maria Magalhães	UFMG	MA
15) A representação de atores sociais no discurso sobre o ensino de espanhol no Brasil em corpus jornalístico	Ariel Novodvorski	2008	Célia Maria Magalhães	UFMG	MA
16) Analisando significados de capas da Revista Raça Brasil: um estudo de caso à luz da semiótica social	Viviane Seabra Pinheiro	2007	Célia Maria Magalhães	UFMG	MA
17) Os significados composicionais e a formação de subjetividades na primeira página de jornais mineiros: um estudo de caso à luz da gramática do design visual	Flaviane Faria Carvalho	2007	Célia Maria Magalhães	UFMG	MA
18) Um caso de polícia: a representação dos discursos no noticiário policial de dois jornais impressos brasileiros, abordada à luz da Lingüística de Corpus e da Análise Crítica do Discurso	Izabella dos Santos Martins Mendes	2003			
19) Tradução como reescrita e retextualização: uma análise do (para)texto	Miriam de Oliveira	2003			

de Um amor de Swann, de Mário Quintana, e de Um amor de Swann, de Fernando Py, a partir das abordagens cultural e discursiva dos Estudos da Tradução	Rondas Nazareth				
20) O lugar do trabalhador e das relações de trabalho em propagandas publicadas em revista brasileira de informação geral: um estudo de caso em ACD	Nádia Dolores Fernandes Biavati	2001			
21) Aspectos híbridos do discurso da Igreja Universal do Reino de Deus: entre a religião e o marketing	Cláudio Márcio do Carmo.	2001			
22) O discurso relatado na imprensa brasileira - jogo de estratégias de apropriação de vozes e de construção de efeitos	Ana Rosa Vidigal Dolabella	1999			
23) Chapeuzinho Vermelho e Caperucita Roja: uma investigação de reescritas com base na representação (visua) de atores sociais e na representação da ação social	Letícia Taitson Bueno Nao se encontra no Banco de Teses da Capes	2011			PhD
24) A palavra-chave racismo e suas relações lexicais: uma análise crítica das relações raciais brasileiras em corpus de jornal impresso	Paulo Henrique Caetano	2007			
25) Relações lexicais, interdiscursividade e	Cláudio do Carmo	2005			

representação: o sincretismo e a questão racial em corpus de jornais e revistas brasileiras 1					
26) A prática sócio-institucional do licenciamento ambiental: a tensão entre os gêneros discursivos, discursos e vozes	Maria Carmen Aires Gomes	2003			

Resumos

Ariel Novodvorski. A Representação de Atores Sociais nos Discursos Sobre o Ensino de Espanhol no Brasil em Corpus Jornalístico.. 01/08/2008

1v. 278p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: FALE; BU

Email do autor:

Palavras - chave:

análise crítica do discurso; análise de corpus

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Cláudio Marcio do Carmo

Célia Maria Magalhães

Nina Célia Almeida de Barros

Linha(s) de pesquisa:

ESTUDOS DA TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

A pesquisa em representação de atores sociais tem se concentrado na investigação do modo como os participantes são incluídos ou excluídos nos discursos midiático, jurídico ou político, entre outros. No escopo do CORDIAL/LETRA/UFMG, os trabalhos têm focalizado a inclusão/exclusão de personagens em romances ou de identidades negras em corpus jornalístico. Em continuidade a essa tradição, mas usando a produção de textos multilíngües e aspectos de suas representações culturais e ideológicas relacionados a questões de mudança social como tema, este trabalho aborda a representação do aumento da demanda pelo ensino de espanhol no Brasil em corpus jornalístico. O estudo consiste numa análise interpretativa de uma prática social recontextualizada, mais especificamente do modo como os atores sociais que participam nos discursos sobre a importância do ensino de espanhol no Brasil foram representados pela mídia. O referencial teórico adotado é a Análise Crítica do Discurso, entendida como um campo de estudos interessado pela compreensão de momentos de mudança e de problemas sociais vigentes, e na qual confluem diversas abordagens com metodologias próprias. A perspectiva aqui assumida é a proposta de van Leeuwen (1996), de base hallidayana, para a análise do componente experiencial da metafunção ideacional, especificamente pela aplicação do inventário de categorias sócio-semânticas na representação de atores sociais. Recorre-se, ainda, aos pressupostos básicos da Linguística Sistemico-Funcional e de sua Gramática Sistemico-Funcional, tal como desenvolvida por Halliday e Matthiessen (2004), no intuito de observar o modo como as representações sócio-semânticas analisadas são realizadas linguisticamente. A Linguística de Corpus também foi utilizada como um suporte para o tratamento do corpus de análise tanto em termos quantitativos como qualitativos; nesse sentido, o corpus foi preparado e sistematizado para uma posterior manipulação com o programa WordSmith Tools®. O corpus de análise está formado por notícias jornalísticas on-line que reportam o assunto em foco, publicadas no Brasil, na Espanha e na Argentina, num período compreendido entre os anos 1998 e 2007. Os resultados da análise apontam uma escassa representação da América Latina e do Mercosul no corpus, apesar de estes serem indicados entre as principais causas pela necessidade do ensino de espanhol no Brasil, e uma elevada representação da Espanha no corpus, sugerindo uma manifestação do entusiasmo de um país que não forma parte nem do Mercosul nem da referida integração latino-americana em torno do tema. Por outro lado, a recorrência de atores sociais representados por meio de anacronismos e simbolismos evoca e parece manter latente um passado de conquista e colonização, indicando, ainda, que o momento atual de mudança discursiva no Brasil em torno do ensino de espanhol atende mais a interesses político-econômicos que culturais e de integração. Ao mesmo tempo, observa-se a recorrência de um discurso modificado que se distancia das questões sócio-culturais implicadas com a inclusão da língua espanhola no sistema de ensino brasileiro.

Viviane Seabra Pinheiro. Analisando Significados de Capas da Revista Raça Brasil: um estudo de caso à luz da semiótica social. 01/05/2007

1v. 130p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

semiótica social, preconceito, consciência negra

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Viviane M. Heberle

Linha(s) de pesquisa:

ESTUDOS DA LINGUAGEM, IDENTIDADE E

REPRESENTAÇÃO Esta linha reúne pesquisas com base em corpora de textos monolíngües, em formato eletrônico e não eletrônico, a partir da noção de linguagem como representação e de linguagem como semiótica social.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CAPES - PROF

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

No contexto brasileiro, tem havido um crescente interesse pelo uso do arcabouço descritivo da gramática do design visual (KRESS e VAN LEEUWEN, 1996) como método de análise eficaz para uma abordagem semiótica social, especialmente quando se trata de sua conjugação com o instrumental da gramática sistêmico-funcional (HALLIDAY e MATTHIESSEN, 2004). O presente trabalho busca contribuir para a interface entre essas gramáticas e para o desenvolvimento dessa abordagem no contexto brasileiro, a partir da análise da capa das edições de novembro de 1996, 1998, 2004 e 2006 da revista Raça Brasil, uma publicação que, presumivelmente, se dirige ao público negro. Utilizando ferramentas analíticas das referidas gramáticas, o principal objetivo desta pesquisa é analisar os significados representacionais, interativos e composicionais construídos pelo componente visual, bem como os significados ideacionais da chamada mais saliente e daquela referente à

imagem da capa e investigar a relação entre esses significados e a noção de 'consciência negra'. Realiza-se uma comparação entre os significados construídos por meio dos modos semióticos visual e verbal nas capas do corpus e investiga-se a existência de mudanças discursivas nas edições e/ou nos períodos analisados. A análise corrobora as afirmações de estudiosos a respeito da postura ambivalente de Raça, a qual se manifesta em vários sentidos. Verifica-se um movimento por parte da revista no sentido de promover/fortalecer a consciência negra dos leitores, por meio, principalmente, do estímulo à sua auto-estima, ao mesmo tempo em que se observa que as representações construídas muitas vezes evocam discursos preconceituosos, que são assimilados e reproduzidos. Além disso, é possível notar que, paralelamente à postura politizada da revista, caminha o interesse mercadológico, que parece determinar a construção do layout da capa e mesmo o conteúdo de determinadas chamadas. Por fim, as mudanças semióticas observadas suscitam as hipóteses de que, por um lado, a revista pode ter se consolidado como uma publicação dirigida aos negros e, por outro, pode ter expandido seu público-alvo, renunciando à segmentação deste quanto à raça.

Flaviane Faria Carvalho. Os significados posicionais e a formação de subjetividades na primeira página de jornais mineiros: um estudo de caso à luz da gramática do design visual. 01/09/2007

1v. 124p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

textos multimodais, mídia impressa, linguística aplicada

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Célia Maria Magalhães

Josenia Antunes Vieira

Maria Carmen Aires Gomes

Linha(s) de pesquisa:

ESTUDOS DA LINGUAGEM, IDENTIDADE E REPRESENTAÇÃO Esta linha reúne pesquisas com base em corpora de textos monolíngües, em formato eletrônico e não eletrônico, a partir da noção de linguagem como representação e de linguagem como semiótica social.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

No contexto acadêmico nacional tem havido um interesse crescente por trabalhos integrando estudos sobre os textos multimodais e a mídia impressa. Desta perspectiva, a presente dissertação visa contribuir para os estudos do layout da primeira página dos jornais mineiros de maior circulação no estado: Estado de Minas, Diário da Tarde, Aqui, Hoje em Dia, O Tempo, e Super Notícia, edições coletadas durante o mês de fevereiro de 2007. O objetivo principal do estudo é verificar as tendências destes jornais quanto à representação de significados sociais e a formação de subjetividades no layout da primeira página de cada um deles, investigando também se o formato pode ser considerado um parâmetro para a representação de tais significados. O referencial teórico da pesquisa é o da semiótica social e teoria da multimodalidade. A metodologia do trabalho baseia-se na aplicação da rede de sistemas referente aos significados composicionais. Os resultados da análise indicam que os recursos representacionais configurados nos jornais Diário da Tarde, Aqui e Super Notícia, o primeiro de formato padrão e os dois últimos de formato tablóide, parecem não orientar os leitores para a possibilidade de reflexão e formação social, política e cultural de suas subjetividades, uma vez que os habitua a destinar atenção a informações vinculadas ao prazer, ao relaxamento, à distração, ao entretenimento, ao consumo e à violência, privilegiando o visual e a conseqüente apreensão imediata das informações. Em contraste, os elementos representacionais articulados nos jornais Estado de Minas, Hoje em Dia e O Tempo, todos de formato padrão, parecem disponibilizar, ainda que de forma rasa, uma estrutura semiótica orientada para a formação social, cultural e política da subjetividade dos leitores, ao dispor mais informações textuais relativas a variadas seções e cadernos, e também por configurar as chamadas de maneira ordenada e sistemática na página, possibilitando, assim, uma melhor compreensão dos fatos e acontecimentos que permeiam o seu mundo. Contudo, verifica-se que os jornais O Tempo e Diário da Tarde, ambos de formato padrão, sinalizam traços correspondentes aos jornais “populares”: o primeiro, de modo mais tênue, parece orientar os leitores para o consumo e para a violência; já o último, de maneira mais acentuada, parece habituar os leitores ao consumo, à violência, ao prazer e ao entretenimento. Por fim, é possível notar que todos os jornais analisados atribuem elevado teor de emoção e idealização ao futebol, assegurando-lhe o status de valor culturalmente construído pela sociedade brasileira.

Isabella dos Santos Martins Mendes. Um caso de polícia: a representação dos discursos no noticiário policial de dois jornais impressos brasileiros, abordada à luz da Linguística de Corpus e da Análise Crítica do Discurso. 01/07/2003

1v. 197p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG

Email do autor:

debella@ig.com.br

Palavras - chave:

noticiário policial, mídia

Área(s) do conhecimento:

Banca examinadora:

Adriana Silvina Pagano

Antônio Paulo Berber Sardinha

Célia Maria Magalhães

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO Abordagens variadas do discurso, refletindo a diversidade de fontes, de instrumentais teóricos, de metodologias, bem como a natureza diversa do objeto-discurso e os múltiplos interesses que nele são projetados.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CNPq

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Este trabalho tem como objetivo analisar a representação dos discursos das fontes jornalísticas no noticiário policial dos jornais "Folha de S. Paulo" e "O Dia". Foram analisados dois meses de notícias e reportagens policiais de cada jornal, cujos discursos foram classificados nas categorias básicas "discurso não institucional" e "discurso institucional", esta dividida em subcategorias. Como o discurso não institucional e o policial foram os mais numerosos em ambos os jornais, estas categorias foram analisadas mais detalhadamente. Foram usados os referenciais teórico-metodológicos da Linguística de Corpus e da Análise Crítica do Discurso. Para a pesquisa essencialmente quantitativa, foi usada a proposta teórica e metodológica da Linguística de Corpus, com a utilização das ferramentas computacionais "Concord" e "WordList" do programa "WordSmith Tools". Para um refinamento qualitativo, foram usados os referenciais teóricos e metodológicos da Linguística de Corpus e da Análise Crítica do Discurso. Para tanto, foram analisadas linhas de concordância cujo nóculo eram verbos de elocução, procurando contextualizar estes itens do corpus dentro

das ordens do discurso midiático. A análise dos resultados aponta para diferenças no modo de representação dos atores sociais, assim como para diferenças nos modos de construção da subjetividade dos leitores dos dois jornais. Conclui-se que o modo de representar os vários discursos presentes nas notícias e reportagens jornalísticas é indicativo de ideologias, e que o estudo das ordens do discurso aponta relações de poder e hegemonias que norteiam a representação dos fatos.

Miriam de Oliveira Rondas Nazareth. Tradução como reescrita e retextualização: uma análise do (para) texto de "Um amor de Swann", de Mário Quintana, e de "Um amor de Swann", de Fernando Py, a partir das abordagens cultural e discursiva dos Estudos da Tradução. 01/07/2003

1v. 179p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG

Email do autor:

Palavras - chave:

Estudos da Tradução, traduções brasileiras

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Solange Ribeiro de Oliveira

Linha(s) de pesquisa:

TRADUÇÃO Estudos da tradução baseados em corpora sob as perspectivas discursiva e cognitiva, examinando-se os elementos contextuais da tradução, a tradução intralingual, processos de hibridismo e a tradução de textos sensíveis (sagrados).

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CNPq

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

A presente pesquisa tem por objetivo contribuir para os Estudos da Tradução, especialmente para a interface entre as abordagens cultural e lingüístico-discursiva, abordagens tradicionalmente tidas como opostas nos Estudos da Tradução. Para tanto, analisa-se o paratexto e a textura das

traduções brasileiras de "Un amour de Swann", de Marcel Proust, realizadas por Mário Quintana e por Fernando Py em períodos históricos diferentes. Sob a perspectiva da abordagem cultural, enfatiza-se a tradução como reescrita, como processo que envolve fatores extra-textuais diversos, como deslocamento no tempo e no espaço, processo de leitura e de construção de texto realizados por um tradutor/leitor, novo público leitor, dentre outros. Elemento fundamental na reflexão sobre a inserção dos romances, enquanto gêneros literários, em um novo contexto, é o paratexto, cuja análise constitui uma importante fonte de informações sobre a recepção de um texto e, portanto, é foco de interesse nesta pesquisa. Sob a perspectiva da abordagem linguístico-discursiva, o aspecto privilegiado é o da coesão lexical, que se liga diretamente à construção do texto e às relações que se estabelecem dentro dele. As redes coesivas formadas a partir das escolhas lexicais do autor e dos tradutores permitem construir uma interpretação coerente de suas retextualizações, especialmente ao buscar relacioná-las aos elementos para e extra-textuais. Os resultados da pesquisa confirmam o caráter complementar das duas abordagens dos Estudos da Tradução contempladas nesta pesquisa, o qual permitiu uma análise mais abrangente de duas traduções diferentes de um mesmo texto de partida, à luz da(s) metodologia(s) de análise (para)textual.

Nádia Dolores Fernandes Biavati. O lugar do trabalhador e das relações de trabalho em propagandas publicadas em revista brasileira de informação geral: um estudo de caso em ACD. 01/11/2001

1v. 169p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: Biblioteca da FALE - UFMG

Email do autor:

Palavras - chave:

trabalho, propaganda, lingüística

Área(s) do conhecimento:

EDUCAÇÃO

Banca examinadora:

Célia Maria Magalhães

Deise Prina Dutra

Maria Izabel Santos Magalhães

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO As pesquisas examinam a produção, recepção e circulação do discurso, sob os enfoques enunciativo, organizacional e interacional. Estas pesquisas orientam-se pelas vertentes teóricas francesa e anglo-americana.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CNPq

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

O objetivo principal deste trabalho é a caracterização das representações da identidade do trabalhador e das relações de trabalho em propagandas de empresas publicadas em revista brasileira de informação geral, buscando entender, em última instância, o modo como a linguagem forma o social e como o social forma a linguagem no gênero discursivo pesquisado. Esta pesquisa baseia-se no referencial da Análise Crítica do Discurso, preferencialmente em Fairclough (1992a; 1994; 1995a; 1995b; 1999; 2000; 2001), nos estudos que privilegiam a linguagem como prática social. Para tanto, realizou-se um estudo de propagandas publicadas no biênio 1988-1989 e em 1999, buscando-se comparar as representações nos dois períodos. Uma das constatações desta pesquisa foi a de que a transformação caracterizada por teóricos sociais está expressa nas propagandas, as quais refletem e ajudam a construir, discursivamente, identidades que partem do trabalhador operário, passam pelo técnico e o executivo e chegam à opacidade da identidade do trabalhador.

Cláudio Márcio do Carmo. Aspectos híbridos do Discurso da Igreja Universal do Reino de Deus (IURD) na mídia televisiva: entre a religião e o marketing. 01/12/2001

1v. 235p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS
- ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: Biblioteca da FALE

Email do autor:

Palavras - chave:

Igreja, Religião, mídia, marketing

Área(s) do conhecimento:

EDUCAÇÃO

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Luiz Paulo da Moita Lopes

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO As pesquisas examinam a produção, recepção e circulação do discurso, sob os enfoques enunciativo,

organizacional e interacional. Estas pesquisas orientam-se pelas vertentes teóricas francesa e anglo-americana.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Este trabalho versa sobre os discursos do programa Pare de Sofrer, da Igreja Universal do Reino de Deus, veiculado pela Rede Record de Televisão. Calçados no referencial teórico da Análise Crítica do Discurso, analisamos o referido evento enquanto texto, prática discursiva e prática social, de forma a desvelar alguns aspectos argumentativos e persuasivos daqueles discursos e as possíveis razões para que uma instituição religiosa se insira na mídia. Procuramos verificar como são construídas as representações e identidades sociais veiculadas pelo programa, as quais estão intimamente ligadas a relações de poder que nem sempre estão claras, por serem frutos de ideologias e naturalizações. A análise do corpus mostra que o discurso da Igreja Universal do Reino de Deus se opõe ao discurso das outras religiões numa relação de exclusão, cumprindo principalmente uma função mercadológica, uma vez que o evento funciona como marketing para a instituição religiosa. O hibridismo do gênero analisado converge para a promoção ideológica da Igreja Universal do Reino de Deus, pois os diferentes discursos e gêneros que o compõem funcionam como um dos mecanismos constituintes e solidificadores da religião como indústria cultural no Brasil de hoje.

Cláudio Márcio do Carmo. Relações lexicais, interdiscursividade e representação: o sincretismo e a questão racial em corpus de jornais e revistas brasileiras. 01/09/2005

1v. 234p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: FALE e BU

Email do autor:

Palavras - chave:

questão racial, revistas brasileiras, discurso midiático

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Luiz Paulo da Moita Lopes

Nilma Lino Gomes

Orlando Vian Junior

Linha(s) de pesquisa:

ESTUDOS DA LINGUAGEM, IDENTIDADE E

REPRESENTAÇÃO Esta linha reúne pesquisas com base em corpora de textos monolíngües, em formato eletrônico e não eletrônico, a partir da noção de linguagem como representação e de linguagem como semiótica social. A fundamentação teórica adotada contempla a lingüística s

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Esta pesquisa versa sobre um fenômeno sociocultural tradicionalmente estudado dentro das Ciências Sociais, especialmente Antropologia, chamado sincretismo. As constantes leituras sobre o assunto revelaram uma carência de estudos sobre como esse fenômeno é construído e reproduzido através da linguagem. Ou seja, havia muitas pesquisas que se dedicaram ao estudo das relações raciais no Brasil e sobre o sincretismo, mas a linguagem, como um importante veículo de comunicação e representação, não foi contemplada. Buscando, então, contribuir com os estudos sobre relações raciais no Brasil, partindo do sincretismo como um de seus aspectos, procurei esclarecer se existe um padrão de ocorrência para a palavra-chave sincretismo em gêneros do discurso midiático, com quais palavras ela se associa e quais prosódias semânticas essas associações criam. Também tentou-se encontrar uma forma de vincular essas associações lexicais e suas prosódias semânticas ao contexto de situação e, por fim, ao contexto de cultura, associando-as à questão racial brasileira. Para tanto, foram constituídos um corpo teórico-metodológico interdisciplinar escolhido na interface Análise Crítica do Discurso, Lingüística Sistemico-funcional e Lingüística de Corpus e um corpus monolíngue para a investigação. A perspectiva intitulada pesquisa de empoderamento também foi usada como base para o estudo. O corpus foi coletado online em dois jornais (O Globo e Folha de São Paulo) e duas revistas de informação geral (Veja e Época) no período 1998-2003. A análise da padronização da palavra sincretismo no corpus mostra a existência de uma forma bastante prototípica de ocorrência para ela e uma prosódia semântica preponderantemente positiva, que pode ser interpretada como ideológica. Os campos evocados pela palavra sincretismo são muito variados, devido a seu significado ser predominantemente associado ao da palavra mistura. O fenômeno colocacional em si não ajuda a construir um

perfil semântico para o nóculo da pesquisa e seu valor pode ser percebido na tessitura dos textos, uma vez que gera coesão e confluência dos campos, redirecionando-os em torno do significado de sincretismo como mistura de culturas e religiões diferentes. O perfil semântico da palavra decorre das relações lexicais que sinalizam para o campo, imprimindo valorações oriundas de diferentes pontos de vista acerca do fenômeno e de suas possíveis relações com a questão racial brasileira. Os gêneros discursivos mais utilizados na veiculação de idéias acerca do sincretismo – a reportagem e a propaganda – predominantemente constroem e reproduzem discursos informativos e promocionais. O único gênero que apresenta essa questão com a profundidade e a discussão exigidas pelas inúmeras variáveis a ele relacionadas é o gênero artigo, numericamente menor em quase todos os veículos. As tensões discursivas mostram que há um mascaramento de um problema grave ligado ao desrespeito às diferenças e à intolerância. Ele também poder ser interpretado como uma tentativa de apaziguar conflitos, reprimindo o contexto de hibridismo étnico e cultural do Brasil, mantendo o país como uma construção monocultural da democracia racial e indicando que ainda existe racismo subjacente, que tenta impor a necessidade de clarear que impele os grupos de descendentes de africanos a uma situação de marginalidade.

Ana Rosa Vidigal Dolabella. O DISCURSO RELATADO NA IMPRENSA BRASILEIRA. 01/02/1999

1v. 378p. Mestrado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: FALE/UFMG

Email do autor:

Palavras - chave:

DISCURSO, IMPRENSA, JORNALISMO

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

ANA RACHEL MACHADO

Célia Maria Magalhães

Maria Sueli de Oliveira Pires

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO As pesquisas em análise do Discurso dividem-se em dois enfoques: o primeiro examina as condições da produção, a existência dos enunciados e os efeitos extralingüísticos que o uso da linguagem busca obter, o segundo, a organiz. da língua acima.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Este trabalho busca traçar um perfil da configuração gráfico-lingüístico-discursiva do discurso de outrem, designado pôr Discurso relatado (Dr.), na imprensa de referência brasileira. Para tanto, optamos pôr focalizar esse objeto de estudo no material noticioso de política de três jornais de grande circulação no país : Folha de São Paulo, O Estado de Minas e O Globo. Os dados considerados correspondem a observação de cinco parâmetros de análise, que abordam as configurações gráfica, lingüística e discursiva do Dr., elaborados a partir de contribuições teóricas de autores como Patrick Charaudeau, Oswald Ducrot e Laurent Perrin sobre o fenômeno da polifonia. A análise interpretativa dos resultados baseia-se na proposta teórica semiolingüística de Patrick Charaudeau. Os procedimento expostos acima nos permitem levantar hipóteses sobre os processos de des/construção do discurso do outro, contrapondo-os com a prática jornalística difundida pelos manuais de redação de cada jornal considerado. Procuramos mostrar que o Dr. exerce função específica dentro do objetivo maior de 'informar', como um jogo de estratégias de apropriação de vozes e de construção de efeitos de posicionamento e de distanciamento do jornalista diante do acontecimento (ou melhor, do dito) que reporta. Esse comportamento revela a superposição das finalidades de 'fazer saber' e 'fazer sentir', que regem o contrato de comunicação midiático do jornal de referência brasileiro.

Paulo Henrique Caetano. A palavra-chave racismo e suas relações lexicais: uma análise crítica dos discursos sobre relações raciais brasileiras em corpus de jornal impresso. 01/01/2007

1v. 237p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: FALE e BU

Email do autor:**Palavras - chave:**

racismo, identidades sociais, Lingüística Aplicada

Área(s) do conhecimento:

LINGÜÍSTICA APLICADA

Banca examinadora:

Adriana Silvina Pagano

Cláudio Márcio do Carmo

Célia Maria Magalhães

Nilma Lino Gomes

Paulo Henrique Aguiar Mendes

Linha(s) de pesquisa:

ESTUDOS DA LINGUAGEM, IDENTIDADE E REPRESENTAÇÃO Esta linha reúne pesquisas com base em corpora de textos monolíngües, em formato eletrônico e não eletrônico, a partir da noção de linguagem como representação e de linguagem como semiótica social.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

FAPEMIG

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Algumas escolas de Análise Crítica do Discurso baseadas na Lingüística Sistemática Funcional de Michael Halliday, têm se constituído importantes ramos de AD no vasto campo da Lingüística Aplicada. Com destaque e pioneirismo na ACD, a teoria social do discurso, de Norman Fairclough, busca um diálogo transdisciplinar com disciplinas das Ciências Sociais, propondo análises discursivas em diversos contextos sócio-culturais, e descrevendo criticamente o caráter discursivo das relações de poder e hegemonia na contemporaneidade. O discurso, aspecto do mundo social, construção social da realidade, é elemento constitutivo das relações sociais, e objeto da ACD, sendo perpassado por gêneros (que realizam ações), por discursos (que representam), e por estilos (que identificam). O discurso como articulado por veículos de comunicação de massa é um enfoque bastante explorado no contexto disciplinar brasileiro, com importantes núcleos de pesquisa na Universidade de Brasília, na Universidade Federal do Rio de Janeiro, e na Universidade Federal de Minas Gerais, por exemplo. Essa ênfase busca discernir aspectos das práticas discursivas da mídia que são constitutivas de práticas sociais. O projeto CORDIAL, da Faculdade de Letras da UFMG, além de outras orientações, reúne trabalhos centralizados na coesão lexical em corpora de mídia impressa, com atenção especial para as relações de gênero e relações raciais brasileiras. Este trabalho se insere em tal nicho, utilizando instrumentos da análise textual em favor da pesquisa social. Procurou-se associar a descrição de padrões da coesão lexical a discursos, gêneros e contextos nas relações raciais brasileiras, análogo ao enfoque de pesquisas anteriores do projeto, com ênfase nas mudanças discursivas e orientações para a diferença. O corpus da pesquisa é constituído por todos os textos contendo a palavra-chave (WILLIAMS, 1988) racismo publicados nos anos de 1995, 2001 e 2005, no jornal Folha de São Paulo. Tais momentos revelaram uma grande produtividade de discursos em competição, além de notável dinamismo em termos de conceitos centrais nas relações raciais

brasileiras. Esses conceitos foram descritos quanto a sua ocorrência no corpus, em contraste com seus usos e definições em dicionários e em publicações acadêmicas especializadas. Foi realizada uma análise das relações lexicais da palavra chave racismo e de itens lexicais e colocações semanticamente relacionadas, organizadas em campos semânticos. Esses foram nomeados campo semântico das relações raciais naturalizadas, campo semântico das identidades sociais, com ênfase em identidades sócio-raciais, e campo semântico das relações raciais em construção. Uma série de determinações contextuais foi evidenciada no léxico das relações raciais, e algumas tendências a mudanças discursivas, como a democratização e tecnologiação do discurso, foram apontadas. Também foi observada a sobrevivência de noções anacrônicas e conceitos já refutados na literatura especializada, mas que exercem grande pressão sobre os discursos que a mídia impressa veicula. E, sendo a mídia impressa ainda um espaço de grande influência na formação ética e política de seus/suas leitores(as), exercendo o seu poder simbólico, há que se expandir as formas de análise e de interpretação de seu funcionamento como forma de empoderamento. Isso contribuiria para uma maior inserção social da Lingüística Aplicada em temas críticos para o desenvolvimento ético e humanitário da sociedade brasileira, tarefa para a qual o trabalho procurou contribuir.

Maria Carmen Aires Gomes. A Prática Sócio-Institucional do Licenciamento Ambiental: a tensão entre os gêneros discursivos, discursos e vozes. 01/07/2003

1v. 250p. Doutorado. UNIVERSIDADE FEDERAL DE MINAS GERAIS - ESTUDOS LINGÜÍSTICOS

Orientador(es): Célia Maria Magalhães

Biblioteca Depositária: Biblioteca da FALE e Biblioteca Universitária da UFMG

Email do autor:

Palavras - chave:

gêneros discursivos, Teoria Social do Discurso

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Adriana Silvina Pagano

Célia Maria Magalhães

Dylia Lysardo Dias

Maria Izabel Santos Magalhães

Sônia Maria de Oliveira Pimenta

Linha(s) de pesquisa:

ANÁLISE DO DISCURSO Abordagens variadas do discurso, refletindo a diversidade de fontes, de instrumentais teóricos, de metodologias, bem

como a natureza diversa do objeto-discurso e os múltiplos interesses que nele são projetados.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

Idioma(s):

Português

Dependência administrativa

Federal

Resumo tese/dissertação:

Neste trabalho, investigam-se a constituição e o entrecruzamento de gêneros discursivos, discursos e vozes na prática sócio-institucional de licenciamento ambiental para construção de barragens como lugar e suporte de lutas pela transformação do social, a partir das hibridizações e ambivalências, questões que emergem nas narrativas da modernidade tardia. Foram usados os pressupostos da Teoria Social do Discurso, tal como proposta por Norman Fairclough, pelo fato de esta investigar os problemas emergentes da sociedade contemporânea através dos diálogos produzidos na esfera pública, por meio da análise intertextual das ordens do discurso (gêneros e discursos) que se projetam e se articulam no curso das lutas hegemônicas pelo poder. À trajetória desta proposta crítica, duas abordagens teóricas foram integradas: os estudos sobre a Análise de Gêneros, sob a perspectiva de John Swales e a Semiótica Discursiva, de Gunther Kress e Theo van Leeuwen. O corpus se constitui de cinco gêneros discursivos bastante distintos: dois gêneros discursivos oficiais, que fazem parte legalmente da prática sócio-institucional em questão _ Relatório de impacto ambiental e Audiência Pública _ e três gêneros discursivos ditos "não-oficiais": cartilha dirigida às comunidades atingidas, concurso de poesias dos atingidos por barragens e jornal Atingidos. Verificou-se que há um hibridismo entre os gêneros discursivos que sugere uma aproximação entre os domínios públicos e privados, religioso e técnico-científico, entre objetividade/tecnicismo e subjetividade/tradição a qual segue as tendências discursivas impostas pela nova ordem mundial. Esse hibridismo sugere, ainda, que há um reenchaie de ordens discursivas jurídico-ambientais em ordens sociais do consumo, privilegiando a função promocional do discurso, fazendo emergir "incongruentes" discursos, gêneros do discurso e vozes. Sob esta perspectiva de contenção, a prática ora analisada não se realiza para empoderar a comunidade atingida, mas para legitimar a imagem da empresa que irá construir a barragem, produzindo um discurso promocional, mostrando as vantagens e desvantagens que a comunidade atingida e o meio ambiente terão com a construção da barragem.

Izabella dos Santos Martins Mendes. Construção e Representação de Realidades no Discurso de Falantes com Esquizofrenia: Uma Abordagem Sistêmico-Funcional. 01/08/2008

1v. 200p. Doutorado. PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO - LINGÜÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

Orientador(es): Leila Barbara

Biblioteca Depositária: Biblioteca da PUC/SP

Email do autor:

Palavras - chave:

Discurso, Falantes com Esquizofrenia, Corpus

Área(s) do conhecimento:

LINGÜÍSTICA

Banca examinadora:

Antonio Paulo Berber Sardinha

Célia Maria Macêdo de Macêdo

Célia Maria Magalhães

Heloisa Collins

Leila Barbara

Linha(s) de pesquisa:

LINGUAGEM E TRABALHO Os objetivos desta linha são: ampliar a atuação profissional do lingüística a diferentes setores da sociedade, entre eles, empresa, escola, serviço público e hospital; proceder à análise e à descrição de gêneros que circulam em diferentes situações.

Agência(s) financiadora(s) do discente ou autor tese/dissertação:

CNPq

Idioma(s):

Português

Dependência administrativa

Particular

Resumo tese/dissertação:

Esta tese, que se insere na área de estudos sobre linguagem e esquizofrenia, objetiva analisar a representação da realidade no discurso de pacientes diagnosticados como portadores de esquizofrenia. Para tanto, é utilizado o referencial teórico-metodológico da Lingüística Sistêmico-Funcional e da Lingüística de Corpus. O corpus da pesquisa é constituído pela transcrição de 12 consultas psiquiátricas realizadas em clínicas particulares e em unidades públicas de saúde mental, com 12 pacientes portadores de esquizofrenia. A pergunta geral da pesquisa é: quais os padrões de representação da realidade expressos no discurso dos

pacientes? Para respondê-la, analiso a metafunção ideacional (Halliday, 1994), tocando nos processos escolhidos pelos falantes com esquizofrenia para a representação dos acontecimentos - tanto do mundo exterior quanto os do mundo interior. Assim, a pesquisa aborda o discurso dos pacientes em termos de padrões léxico-gramaticais e de padrões semântico-discursivos. Os resultados evidenciam uma tendência de uso muito freqüente de operadores de polaridade negativa e elementos modalizadores, bem como de atribuição de características negativas tanto a si próprios quanto às entidades, lugares, objetos e pessoas em geral. O relato de acontecimentos negativos também foi constante. A projeção de uma identidade investida de poder também pôde ser percebida em alguns momentos de seu discurso, que puderam ser mais claramente identificáveis como instâncias delirantes

Anexo 7: Emails trocados com Ariana Zanella

Oi Ariana,

(...)

Enfim, estou agora no meu doutorado fazendo analise de todas as pesquisas que trabalharam com descricao de traducao em interface com a Linguistica Sistematica-Funcional e a Linguistica de Corpus. Sua pesquisa, claro, se encontra entre elas. Achei muito interessante na tua analise que voce eh a unica que utiliza o software Syscoder. Voce inclusive codifica seu corpus nesse software. Como nao conheco e nao sei como consegui-lo (pois barato tambem nao deve ser), tenho umas perguntinhas pra te fazer, se voce nao se importa (e se vc se lembrar, porque acho que o tempo passou igualmente a nos duas, ne? :).

1. Voce sabe como conseguir esse software?
2. Voce se lembra se esse programa eh capaz de enumerar a quantidade de ocorrencias de Processos Materiais, por exemplo?
3. Honestamente, vc achou facil trabalhar como esse programa, ou seria mais facil ter feito tudo manualmente??

Desculpa a intromissao no seu email, espero ter acertado o endereco electronico. Agradeço qualquer ajuda que puder dar :)

Um grande abraço pra ti,



Ariana Zanella <[REDACTED].br>

28/05/1

2

para mim

Oi Lilian

Claro que lembro de voce. (...).

Fico orgulhosa em poder ajuda-la. O syscoder é fácil de baixar e manipular. Você precisa saber bem fazer a matriz sistêmico-funcional, é como na linguística de corpus, depois tagging o seu corpora. Manualmente seria impossível. O melhor que ele te dará um percentual de confiança utilizando student T test e estatísticas tratadas para cada processo do sistema de transitividade. Não tenho minha pesquisa neste computador, mas posso procurar o cd que gravei toda a minha pesquisa e te repassar.

(...)

bjs

Ariana Zanella

De: lilian fleuri <[REDACTED].com>

Para: Ariana Zanella <[REDACTED].br>

Enviadas: Quinta-feira, 24 de Maio de 2012 2:27

Assunto: perguntas sobre sua dissertacao



lilian fleuri <[REDACTED]@[REDACTED].com>

01/06/1

2

para Ariana

Oi Ariana,

Muito obrigada pela resposta! (...) Eu encontrei a versao mais nova do Syscoder e ja estou tentando usar. Nao entendo ainda o conceito de "matriz sistematico-funcional" (cada pesquisador cria sua propria ou ha alguma sugestao no proprio programa?). Acho que vou descobrir mais quando entrar na fase de "futucar" o programa. Voce tem como alinhar os textos codificados?

(...)

Acho que com a resposta que vc me deu ja eh suficiente pra eu dar continuidade ao meu trabalho.

Eh bom saber que seu email continua o mesmo, se eu tiver mais duvidas, e vc nao se importar, vou voltar a te escrever.

Beijo bem grande!

Lilian

Anexo 8: Perguntas no Fórum MS Office

MS EXCEL: (MACRO) JUNTAR CÉLULA DE BAIXO COM CÉLULA DE CIMA

[answers.microsoft.com/en-us/office/forum/office_2010-excel/edit-text-in-a-cell-with-text-from-another-cell/a8dc5ef3-f9b7-47e9-988d-80be73619389](#)
[competing - Dicion...](#) [YouTube - Ila2727...](#) [Doutorado](#) [Aulas](#) [Universidades](#) [TextBook](#) [Dicionarios](#) [Meus Sites](#) [Imported From Fref...](#) [Learn More](#)

By using this site you agree to the use of cookies for analytics, personalized content and ads.

[Microsoft](#)
Community

[Home](#) [Categories](#) [Getting Started](#) [FAQ](#) [Additional Support](#)

Office
 Applies To: [Office](#) | [Office 2010](#) | [Microsoft Office Excel](#)


Question
 victorcarreto asked on June 29, 2012

edit text in a cell with text from another cell

Hello can any one help me I want to use excel to align two texts. Anyways, I am looking for a way to take the text on cell A2 to the end of the text on cell A1, and delete A2, shifting the other cells up. Is it possible to do it by using a shortcut key? How can I do this?

I really appreciate any help.
 Regards
 Vic-

[Reply](#) | [Reply with quote](#) | [Report abuse](#) | [Subscribe to Updates](#)

Search the Community Post to the Community

Sign in

1 Had this question Me Too

Related Content
 Is there a formula for when the text in one cell changes, excel...
 Search Text and place findings in another cell
 Matching cell from input sheet to a list in another worksheet
 Remove trailing spaces in a cell with a number
 If when referencing a cell from one worksheet to another in the same...







Answer

1 Found this helpful Me Too

Jeeped replied on June 29, 2012 ▾

MVP Wiki Author Community Moderator ★ MCC: Content Creator

... I am looking for a way to take the text on cell A2 to the end of the text on cell A1, and delete A2, shifting the other cells up. Is it possible to do it by using a shortcut key?

Tap **Alt+F11** and when the VBE opens, *immediately* use the pull-down menus to *Insert, Module*. Paste the following into teh new pane titled something like *Book1 - Module1 (Code)*,

```
Sub mcrAppend2Above()
  If ActiveCell.Row > 1 Then
    ActiveCell.Offset(-1, 0).Value = ActiveCell.Offset(-1, 0).Text & ActiveCell.Text
    ActiveCell.Delete
  End If
End Sub
```

Tap **Alt+Q** to return to your worksheet. Tap **Alt+F8** to open the *Macros* dialog and use the *Options* button to assign a keystroke combination hotkey to the new macro.

Note that I have used the *.Text* property instead of the *.Value* property to concatenate the cells' contents. This has the effect of transferring any additional formatting. For example, *6/27/2012* formatted as *27-Jun-12* will append *27-Jun-12* to the cell above, not *6/27/2012*. If this is undesirable, change the *.Text* properties to *.Value*.

SP2 for Office 2010 has been released. Update your Office 2010 installation soon.



Answer

1 Found this helpful Me Too

Jeeped replied on July 9, 2012 ▾

MVP Wiki Author Community Moderator ★ MCC: Content Creator

[↶ In reply to victorcarreto post on July 8, 2012](#)

... but how can I do to make the text from the cell bellow, when joining the text in the cell above, to have a space between them ...

Change the third line to add a space in between the top value and the added text like this, `ActiveCell.Offset(-1, 0).Value = ActiveCell.Offset(-1, 0).Text & " " & ActiveCell.Text`

... how can I do to not deactivate **Ctrl+Z** function? (because when I use this macro, I can use **Ctrl+Z**...

If you use the *Options* button from the *Macros* dialog to assign **Ctrl+Z** to the macro, then **Ctrl+Z** will not be *Undo* anymore, it will be this macro. Personally, I typically use **Ctrl+Q** or **Ctrl+W** to assign to a macro because they are not assigned to something already.

SP2 for Office 2010 has been released. Update your Office 2010 installation soon.

[Reply](#) | [Reply with quote](#) | [Report abuse](#) ▾



victorcarreto replied on July 8, 2012 ▾

0

Found this helpful
Me Too

[↶ In reply to Jeeped post on June 29, 2012](#)

Thank you! It worked, but how can I do to make the text from the cell below, when joining the text in the cell above, to have a space between them, i.e. the last word of cell A2 and the first word of cell A3. What can I change in the Macro to make it happen?

Also, how can I do to not deactivate Ctrl+Z function? (because when I use this macro, I can use Ctrl+Z... Thank you very much!

[Reply](#) | [Reply with quote](#) | [Report abuse](#) ▾



victorcarreto replied on July 9, 2012 ▾

0

Found this helpful
Me Too

[↶ In reply to Jeeped post on July 9, 2012](#)

Hi,

Thank you again! It worked!

Regarding my second question...

If you use the *Options* button from the *Macros* dialog to assign *Ctrl+Z* to the macro, then *Ctrl+Z* will not be *Undo* anymore, it will be this macro. Personally, I typically use *Ctrl+Q* or *Ctrl+W* to assign to a macro because they are not assigned to something already.

Actually, I am not assigning *Ctrl+Z* to the macro and even though this function can not be performed once I start to run the macro created. I heard that the *undo* function stops working once a macro is created. Do you think that is there a way to add the *undo* macro again? I mean, recording or VBA a macro that does the same as the function *Ctrl+z* usually does (*undo*)? See how the "undo" button is not activated? How could I reactivate it?

[Reply](#) | [Reply with quote](#) | [Report abuse](#) ▾



Jeeped replied on July 9, 2012 ▾

MVP Wiki Author Community Moderator ★ MCC: Content Creator

1 Found this helpful Me Too

[↶ In reply to victorcarreto post on July 9, 2012](#)

... Do you think that is there a way to add the undo macro again? I mean, recording or VBA a macro that does the same as the function Ctrl+z usually does (undo)?

I believe I understand what you were referring to now. The commands executed by running a macro cannot be undone using the *Undo* command (e.g. *Ctrl+Z*) and any *Undo* history is wiped clean upon running a macro. Save your work just prior to running the macro, and if you are not happy with the results, close without saving again and reopen to the last saved point. Conventional typing and editing commands can be *undone* up to a maximum number (set in *Options*) stored as history. A macro could conceivably execute anywhere from a single command to thousands of commands including reiterative loops and dynamic variables. There is no internal mechanism to store all of the interaction within VBA code macros and as such, no *Undo* command is available. This is something you will just have to live with and learn to guard against data mishaps. If you are ever not absolutely sure, just run the VBA macro on a copy of the original, retaining both copied at least until you are sure that the changes are to be made permanent.

"With great power comes great responsibility."

SP2 for Office 2010 has been released. Update your Office 2010 installation soon.

[Reply](#) | [Reply with quote](#) | [Report abuse](#) ▾

MS EXCEL: (MACRO) SEPARA UMA CÉLULA EM DUAS

Question

1

Had this question
Me Too

victorcarreto asked on July 8, 2012 ▾

macro to separate lines in one cell

1) Hello I have some text cells that are separated in lines (alt+enter) spread throughout a column, I've tried to record a macro but it is not working it is too specific and I am looking for a macro that could work on cells formatted like this. any ideas?

kinkind regards

vicvic

[Reply](#) | [Reply with quote](#) | [Report abuse](#) ▾ | [Subscribe to Updates](#)



Answer

1

Found this helpful
Me Too

Tom Ogilvy replied on July 9, 2012 ▾

MVP

If you have something like
The quick
brown fox
jumped over
the lazy
Dog's back
in a single cell and you want that to be on separate cells in a single column, then you can try the below code

select the cell and run it. It inserts the necessary number of additional cells.

```
Sub ABC()
Dim s As String, v As Variant, l As Long
s = ActiveCell.Value
v = Split(s, vbLf)
l = UBound(v) - LBound(v) + 1
If l > 1 Then
ActiveCell.Offset(1, 0).Resize(l - 1, 1).Insert Shift:=xlShiftDown
ActiveCell.Resize(l, 1).Value = Application.Transpose(v)
End If
End Sub
```

victorcarreto replied on July 9, 2012 ▾

0

Found this helpful
Me Too

[In reply to Tom Ogilvy post on July 9, 2012](#)

Hi Tom,

Thanks for your answer.

If I want to add the command **Alt+Enter** to this macro, how can I do it?

What I am trying to do is to take the pointer to a part of the text in, lets say, cell A2, click a shortcut key, and break this text into other cells. So what I did was to first separate the text into 2 lines within one cell (took the pointer and clicked Alt+ Enter), and then run the macro you made. It works perfectly, but it would be excelent if I could do it directly, without first breaking the text in 2 lines. Any idea how to do it?

Thanks,

if you don't want to insert cells, then you can remove that line and use

```
Sub ABC()
Dim s As String, v As Variant, l As Long
s = ActiveCell.Value
v = Split(s, vbLf)
l = UBound(v) - LBound(v) + 1
If l > 1 Then
    ActiveCell.Resize(l, 1).Value = Application.Transpose(v)
End If
End Sub
```

but that could overwrite existing data so you are responsible for that

--

Regards,
Tom Ogilvy

Bob Peeples replied on August 20, 2013 ▾

0 Found this helpful
Me Too

[↶ In reply to Tom Ogilvy post on July 9, 2012](#)

This is sort of on the topic, I hope. I am citing cells by reference with 3 lines (name, birth, death) into a new family member's tree by using the existing grandparent's first few columns as the last few columns of the baby's worksheet.

The top line is bold and the entire cell comes over bold. Is there a way to clear the bold format on the entire cell and start over with some sort of highlight of the first character through the first vbLf character and the bold it. I only have to do it to each cell once, then I'm going to lock them so that they have to be changed from the grandparent worksheet.

Having two identical cells to update in two sheets of the same worksheet is not very elegant programming. I have attached the workbook and have just started placing the references in Avery's sheet.

Thanks for any help you can give,
Bob Peeples, PE
http://www.bopeep.com/FamilyTree_AveryMariePeeples.xls



HansV MVP replied on July 8, 2012 ▾

MVP ★ MCC: Content Creator

0 Found this helpful
Me Too

Try this. Make sure that there are enough blank columns to the right of the selection before you run it, to avoid the prompt whether you want to overwrite existing data.

```
Sub SplitLines()
    Selection.TextToColumns Other:=True, OtherChar:=vbLf
End Sub
```

.....
Kind regards, HansV
www.eileenslounge.com

Rick Rothstein, MVP - Excel replied on July 9, 2012 ▾

MVP ★ MCC: Content Creator

0 Found this helpful
Me Too

[↶ In reply to HansV MVP post on July 8, 2012](#)

Try this. Make sure that there are enough blank columns to the right of the selection before you run it, to avoid the prompt whether you want to overwrite existing data.

```
Sub SplitLines()
    Selection.TextToColumns Other:=True, OtherChar:=vbLf
End Sub
```

On my system, I had to do it this way....

```
Sub SplitLines()
    Selection.TextToColumns DataType:=xlDelimited, Other:=True, OtherChar:=vbLf
End Sub
```

otherwise, things broke apart incorrectly if there were spaces in any of first n-1 individual lines (it worked correctly if the only spaces were on the last line).

*** PLEASE READ ***

If this response answers your question, please mark it as the "Answer".

MS EXCEL: (MACRO) BUSCA POR CÉLULAS PARALELAS DESALINHADAS

Question

1

Had this question
Me Too

victorcarreto asked on July 15, 2012 ▾

finding pair of cells with different numbers of words

Hello, I am dealing with texts, not numbers. Column A has a text in English, and column B a text in Portuguese. I need to find pairs of cells A1-B1 or A1234-B1234 whose number of words have a difference of 9 words minimum. For that I have this formula which I wrote on column C, to identify such case by writing the word "MISMATCH":

```
=IF(ABS((LEN(A1)-LEN(SUBSTITUTE(A1," ",""))+1)-(LEN(B1)-LEN(SUBSTITUTE(B1,"",""))+1))>=9,"MISMATCH",0)
```

The thing is that I don't want to use a column, like I used column C, to add such a formula, because I would be inserting and deleting cells in column A and B, and the formula loses its reference every time a modification like this is done.

So, what I need is a macro that, with just a shortcut key ,finds next or takes me directly to the next pair of cells in the condition of this formula (I mean, a pair A-B in which A or B has more than 9 words in relation to its pair).

Any idea of how it could be done??

Thank you!

Answer

1

Found this helpful
Me Too

HansV MVP replied on July 15, 2012 ▾

MVP ★ MCC: Content Creator

Try this macro. It will start searching at the row below the active cell, so you can use it repeatedly. You can assign the macro to a keyboard shortcut by selecting it in the Macros dialog (Alt+F8 from Excel) and clicking Options...

```
Sub FindMismatch()
    Dim r As Long
    Dim m As Long
    m = Range("A:B").Find(What:="", SearchOrder:=xlByRows, _
        SearchDirection:=xlPrevious).Row
    For r = ActiveCell.Row + 1 To m
        If Abs(UBound(Split(Range("A" & r).Value)) - _
            UBound(Split(Range("B" & r).Value))) >= 9 Then
            Application.Goto Range("A" & r), True
            Exit For
        End If
    Next r
End Sub
```

Kind regards, HansV
www.eileenslounge.com

victorcarreto replied on July 16, 2012 ▾

0

Found this helpful
Me Too

[↩ In reply to HansV MVP post on July 15, 2012](#)

Excellent! It worked perfectly! Thank you very much!!

MS EXCEL: (MACRO) BUSCA POR CÉLULAS PARALELAS DESALINHADAS.

Question

1

Had this question
Me Too

victorcarreto asked on July 14, 2012 ▾

Refreshing formula in a column

1) I've a Macro to split a cell and shift cells down, as the following:

```
Sub SplitCell()
Dim s As String, v As Variant, l As Long
s = ActiveCell.Value
v = Split(s, vbLf)
l = UBound(v) - LBound(v) + 1
If l > 1 Then
ActiveCell.Offset(1, 0).Resize(l - 1, 1).Insert Shift:=xlShiftDown
ActiveCell.Resize(l, 1).Value = Application.Transpose(v)
End If
End Sub
```

And, another macro to join cell below with the one above shifting cells up, as the following:

```
Sub JoinCellsMoveup()
If ActiveCell.Row > 1 Then
ActiveCell.Offset(-1, 0).Value = ActiveCell.Offset(-1, 0).Text & " " & ActiveCell.Text
ActiveCell.Delete
End If
End Sub
```

rich007a replied on July 15, 2012 ▾

★ MCC: Content Creator

0

Found this helpful
Me Too

I'm also looking for a macro that would do this function itself (update/refresh the formula in column C).

To do this, right-click on the sheet's tab and View Code. Paste the following:

```
Private Sub Worksheet_Change(ByVal Target As Range)
If Not Intersect(Target, Range("A:A")) Is Nothing Then
Dim lnLastRow As Long
lnLastRow = Cells(Cells.Rows.Count, "A").End(xlUp).Row
Range("C1").Formula = "=Len(A1)-Len(B1)"
Range("C1").AutoFill Destination:=Range("C1:C" & lnLastRow)
End If
End Sub
```

That will run automatically whenever column A is changed. It will reapply the formula to C1, then fill down as far as the row of the last non-empty cell in column A.

Cheers
Rich

rich007a replied on July 15, 2012 ▾

★ MCC: Content Creator

0

Found this helpful
Me Too

Two ways to do this, either change your formula in col C to:
=LEN(INDIRECT("A"&ROW()))-LEN(INDIRECT("B"&ROW()))

or, in your macros, use the line:

```
Range("C1").AutoFill Destination:=Range("C1:C100")
```

to re-fill down the formula from C1. Note 100 should be the last cell in your range. To find that dynamically (since it will probably be changing as you add or delete rows) you could use:

```
Dim InLastRow As Long
```

```
InLastRow = Cells(Cells.Rows.Count, "A").End(xlUp).Row
```

```
Range("C1").AutoFill Destination:=Range("C1:C" & InLastRow)
```

Hope that helps.

Cheers

Rich

When I use these macros, the formula in column C (=Len(A1)-Len(B1)) loses its reference (from the row I used the macro to below). Is there a way to add an update/refresh of the formula in column C, maybe adding some code to my macros that copy the formula in C1 to the rest of the cells in column C?

I'm also looking for a macro that would do this function itself (update/refresh the formula in column C).

Thank you!

victorcarreto replied on July 15, 2012 ▾

0

Found this helpful
Me Too

[↶](#) In reply to rich007a post on July 15, 2012

Thank you Rich,

It worked with the formula =Len(A1)-Len(B1). But when I changed the formula in column C and replaced it in the Macro code, it showed me an error. The new formula in C is:

```
=IF(ABS((LEN(A2)-LEN(SUBSTITUTE(A2," ",""))+1)-(LEN(B2)-LEN(SUBSTITUTE(B2," ",""))+1))>=9,"MISMATCH",0)
```

I changed the macro then for:

```
Private Sub Worksheet_Change(ByVal Target As Range)
    If Not Intersect(Target, Range("A:A")) Is Nothing Then
        Dim InLastRow As Long
        InLastRow = Cells(Cells.Rows.Count, "A").End(xlUp).Row
        Range("C1").Formula = "=IF(ABS((LEN(A2)-LEN(SUBSTITUTE(A2," & "" & "")))+1)-(LEN(B2)-LEN(SUBSTITUTE(B2," & "" & "")))+1))>=9,"MISMATCH",0)"
        Range("C1").AutoFill Destination:=Range("C1:C" & InLastRow)
    End If
End Sub
```

I am pretty new in Excel and I can't see what caused the error. Any idea?

Thank you!!

Tom Ogilvy replied on July 15, 2012 ▾

MVP

0

Found this helpful
Me Too

[↶ In reply to victorcarreto post on July 15, 2012](#)

Victor Carreto,

when you have double quotes in a string you need to replace each double quote within the string to two double quotes (a pair of double quotes). Try this:

```
Private Sub Worksheet_Change(ByVal Target As Range)
    If Not Intersect(Target, Range("A:A")) Is Nothing Then
        Dim InLastRow As Long
        InLastRow = Cells(Cells.Rows.Count, "A").End(xlUp).Row
        Range("C1").Formula = "=IF(ABS((LEN(A2)-LEN(SUBSTITUTE(A2,"" "" "" "" ""))+1)-(LEN(B2)-LEN(SUBSTITUTE(B2,"" "" "" "" ""))+1))>=9,""MISMATCH",0)"
        Range("C1").AutoFill Destination:=Range("C1:C" & InLastRow)
    End If
End Sub
```

--
Regards,
Tom Ogilvy

rich007a replied on July 15, 2012 ▾

★ MCC: Content Creator

0

Found this helpful
Me Too

[↶ In reply to victorcarreto post on July 15, 2012](#)

Yep, what Tom said.

Cheers

Rich

PS. you might want to change the last line to:

```
    If InLastRow > 1 Then Range("C1").AutoFill Destination:=Range("C1:C" & InLastRow)
```

since autofilling to 1 row causes an error (e.g. if you only have data in row 1 in col A).

victorcarreto replied on July 15, 2012 ▾

[↶ In reply to Tom Ogilvy post on July 15, 2012](#)

Thanks Tom,

It almost worked. When tried to join cell B32 with B31 using the macro I mentioned above:

```
Sub JoinCellsMoveUp()
    If ActiveCell.Row > 1 Then
        ActiveCell.Offset(-1, 0).Value = ActiveCell.Offset(-1, 0).Text & " " & ActiveCell.Text
        ActiveCell.Delete
    End If
End Sub
```

... the cell C32 presented a formula referring to A33 and B33:

```
=IF(ABS((LEN(A33)-LEN(SUBSTITUTE(A33," "" "" "" ""))+1)-(LEN(B33)-LEN(SUBSTITUTE(B33," "" "" "" ""))+1))>=9,"MISMATCH",0)
```

Any idea why it's happening?
Many thanks!

MS EXCEL:**ATIVAR O CTRL+Z (UNDO) SOBRE OS MACROS CRIADOS**

Question

1

Had this question
Me Too

victorcarreto asked on July 8, 2012 ▾

how to undo after creating a macro in Excel

Always when I create a Macro in Excel, I am not able to "undo" or to use the command Ctr+Z. What am I doing wrong? What can I do to not desactivate the "undo" command?

Thank you!

Answer

1

Found this helpful
Me Too

HansV MVP replied on July 8, 2012 ▾

MVP ★ MCC: Content Creator

Clearing the undo stack is an unfortunate side effect of running VBA code. You can't prevent it, but there are rather cumbersome ways to provide undo capability - see <http://www.jkp-ads.com/Articles/UndoWithVBA00.asp>

Kind regards, HansV
www.eileenslounge.com

Answer

1

Found this helpful
Me Too

OssieMac replied on July 8, 2012 ▾

★ MCC: Content Creator

You cannot Undo changes that are made with code.

When testing code you should frequently save the Workbook with version numbers. Ensure you save before running the code and then if necessary you can close the workbook without saving and re-open the saved version.

Added with Edit: Also I always try to code so that existing (or source data) is not changed. This is done by having Source data on one worksheet and Output data on another worksheet. This may not always be practical but wherever possible it should be done this way.

Regards,

OssieMac

Answer

1

Found this helpful
Me Too

HansV MVP replied on July 8, 2012 ▾

MVP ★ MCC: Content Creator

Clearing the undo stack is an unfortunate side effect of running VBA code. You can't prevent it, but there are rather cumbersome ways to provide undo capability - see <http://www.jkp-ads.com/Articles/UndoWithVBA00.asp>

Kind regards, HansV
www.eileenslounge.com

Answer

1

Found this helpful
Me Too

OssieMac replied on July 8, 2012 ▾

★ MCC: Content Creator

You cannot Undo changes that are made with code.

When testing code you should frequently save the Workbook with version numbers. Ensure you save before running the code and then if necessary you can close the workbook without saving and re-open the saved version.

Added with Edit: Also I always try to code so that existing (or source data) is not changed. This is done by having Source data on one worksheet and Output data on another worksheet. This may not always be practical but wherever possible it should be done this way.

Regards,

OssieMac

MS WORD: (MACRO) SUBSTITUIR TODOS OS COMANDOS CRIADOS NO FIND & REPLACE.

Question

1 Had this question Me Too

victorcarreto asked on July 8, 2012

Macro to find and replace in Word for line break

I have many MSWord documents that must have their lines broken into sentences. In order to not breaking abbreviations as they were sentences I have to go through ten steps in "Find and Replace" using "wildcards". I wonder if it would have a way to transform these tens steps into one Macro, or each "replacement" into a macro. Can anyone help me with that?

PS.: consider "space" (when existing) between a formula not at the beginning or the end of the formula. At the beginning or the end a space is represented as {type space}. These are the ten steps I have to go through:

- Find ([\?\\]) Replace \1
- Find { @}[2,4]Replace {type space}
- Find ([\?\\])^([A-Z])Replace \2\1 \3
- Find ([A-Z][bcdfghijklmnpqrstvwz]{1,2}). ([A-Z]) Replace \1 \2 ;
- Find (<[a-z]{1}>). ([A-Z]) Replace \1 \2 ;
- Find ([\?\\]\>^)([A-Z]) ([A-Z]) Replace \1\2^\3
- Find ([0-9]{3,4})([A-Z]) ([A-Z]) Replace \1\2^\3
- Find ([!][a-z])([A-Z]) ([A-Z]) Replace \1\2^\3
- Find ([!][A-Z]{2})([A-Z]) ([A-Z]) Replace \1\2^\3
- Find ^t Replace (nothing)

Thank you!!

Answer

0 Found this helpful Me Too

HansV MVP replied on July 8, 2012

MVP ★ MCC: Content Creator

I don't understand a) - it doesn't seem to do anything.
Did you really intend the semi-colon at the end of e) ?
^| (vertical bar) in f), g), h) and i) is not valid. Did you mean ^l (lower case L)?

In the macro below I have assumed that you didn't intend the ; and that vertical bar should be lower case L.

```
Sub BreakLines()
    Application.ScreenUpdating = False
    With ActiveDocument.Content.Find
        .ClearFormatting
        .Replacement.ClearFormatting
        .MatchWildcards = True
        .Execute FindText="([\?\\])", ReplaceWith="\1", Replace=wdReplaceAll
        .Execute FindText="{ @}[2,4]", ReplaceWith=" ", Replace=wdReplaceAll
        .Execute FindText="([\?\\])^([A-Z])", ReplaceWith="\2\1 \3", Replace=wdReplaceAll
        .Execute FindText="([A-Z][bcdfghijklmnpqrstvwz]{1,2}). ([A-Z])", ReplaceWith="\1 \2", _
            Replace=wdReplaceAll
        .Execute FindText="(<[a-z]{1}>). ([A-Z])", ReplaceWith="\1 \2", Replace=wdReplaceAll
        .Execute FindText="([\?\\]\>^)([A-Z]) ([A-Z])", ReplaceWith="\1\2^\3", _
            Replace=wdReplaceAll
        .Execute FindText="([0-9]{3,4})([A-Z]) ([A-Z])", ReplaceWith="\1\2^\3", _
            Replace=wdReplaceAll
        .Execute FindText="([!][a-z])([A-Z])", ReplaceWith="\1\2^\3", _
            Replace=wdReplaceAll
        .Execute FindText="([!][A-Z]{2})([A-Z]) ([A-Z])", ReplaceWith="\1\2^\3", _
            Replace=wdReplaceAll
        .Execute FindText="^t", ReplaceWith="", Replace=wdReplaceAll
    End With
    Application.ScreenUpdating = True
End Sub
```

Answer

0

Found this helpful
Me Too

Graham Mayor replied on July 8, 2012 ▾

MVP ★ MCC: Content Creator

I think you need to check your syntax.
The first example appears to find a punctuation character then replaces it with the found character? I am not sure what that is meant to achieve.
The second should probably be Find []{1,}
'd' could be abbreviated to ([A-Z][b-z]{1,2}). ([A-Z])
I didn't bother checking the rest - see http://www.gmayor.com/replace_using_wildcards.htm

For the macro you would need something along the lines of

```
Sub ReplaceList()
Dim vFindText As Variant
Dim vReplaceText As Variant
Dim oRng As Range
Dim i As Long
vFindText = Array("Find1", "Find2", "Find3") 'your find strings
vReplaceText = Array("Replace1", "Replace2", "Replace3") 'the corresponding replacement strings
For i = 0 To UBound(vFindText)
    Set oRng = ActiveDocument.Range
    With oRng.Find
        .ClearFormatting
        .Replacement.ClearFormatting
    Do While .Execute(FindText:=vFindText(i), _
        MatchWildcards:=True, _
        Forward:=True, _
        Wrap:=wdFindStop) = True
        oRng.Text = vReplaceText(i)
        oRng.Collapse wdCollapseEnd
    Loop
    End With
Next
End Sub
```

but you claim Word Starter edition which does not support macros.

MS WORD:**(FIND & REPLACE) SEPARAÇÃO DE PARÁGRAFOS EM LINHA.**

Question

1

Had this question
Me Too

victorcarreto asked on July 3, 2012 ▾

Separate a paragraph in lines

Hello I am trying to separate a paragraph into individual sentences.

I've been using wild cards to find a point followed by an space and then an upper case. And replace that with a brake line. But abbreviations like "Dr. John" are being separated too. And, as I keep trying different things I find more and more special and particular cases.

Does and body have any ideas on how to solve this problem? Is there any good and easy to understand macro tutorial for word?

Regards.

Vic.-

Anexo 9: Macros, fórmulas e shortcuts *Template Corpus Paralelo*

Shortcuts e Macros no Template Corpus Paralelo

O comando **Ctrl+Q**, para a coluna A, representa o seguinte Macro:

```
Sub FormatarColunaA()
```

```
,
```

```
' FormatarColunaA Macro
```

```
,
```

```
' Keyboard Shortcut: Ctrl+q
```

```
,
```

```
Selection.SpecialCells(xlCellTypeBlanks).Select
```

```
ActiveWindow.SmallScroll Down:=39
```

```
Selection.Delete Shift:=xlUp
```

```
Union(Range(_
```

```
"A494,A498,A502,A545,A599,A605,A607,A609,A622,A734,A736,A803,A805,A823,A902,A904,A974,A976,A1003,A1070,A1072,A1076,A1078:A1079,A1081,A1083,A1085,A1089,A1092,A1098,A1104,A1108,A1111" _  
), Range(_
```

```
"A1115,A1118,A1120,A1126,A1129,A1131,A1135,A1137,A1140:A1141,A1144,A1148,A1154,A1157,A1161,A1164,A1172,A1177,A1179,A1191,A1195,A1219,A1221,A1301,A1303,A1375,A1377,A1457,A1459,A1536,A1538,A1597,A1599" _  
), Range(_
```

```
"A1679,A1681,A1685,A1688,A1690,A1695,A1697,A1702,A1704,A1706:A1707,A1710,A1771,A1773,A1840,A1842,A1913,A1915,A1972,A1974,A2050,A2052,A2136:A2137,A2139:A2141,A2143,A2149,A2151,A2153,A2158,A2160,A2162,A2166,A2168" _  
), Range(_
```

```
"A2171,A2175,A2178,A2181,A2183,A2187,A2191,A2194,A2196,A2199,A2201,A2204,A2206,A2209,A2252,A2254,A2261,A2263:A2264,A2267,A2269,A2286,A2288,A2341,A2343,A2452,A2454,A2527,A2529,A2561,A2563,A2590,A2592" _  
), Range(_
```

```
"A2632,A2634,A2638,A2643,A2647,A2650,A2654,A2657,A2661,A2663,A2665,A2668,A2670,A2673,A2675,A2677:A2679,A2682,A2711,A2713,A2801,A2803,A2822,A2897,A2911,A3015,A3017,A3057,A3059,A3091,A3093,A3131:A3132,A3134:A3135" _  
), Range(_
```

```
"A3137,A3139,A3141,A3144,A3146,A3150,A3152,A3154,A3156,A3160,A3162,A
```

```
3164,A3166,A3169,A3172,A3174,A3176,A3178,A3182,A3184,A3189,A3193:A3194,A3197,A3228,A3230,A3340,A3342,A3392,A3394,A3456,A3458,A3496" _
), Range( _
```

```
"A3498,A3527,A3566,A3568,A3631,A3633,A3668,A3670,A3798:A3799,A3801,A3803,A3805,A3807,A3809,A3811,A3813,A3816,A3818,A3825,A3827,A3829:A3831,A3834,A3845,A3847,A3902,A3904:A3905,A3979,A3981,A4049,A4051:A4052,A4054,A4056" _
), Range( _
```

```
"A4058,A4060,A4062,A4064,A4066,A4068,A4070:A4072,A4075,A4110,A4112,A4209,A4211,A4252,A4281,A4306,A4308,A4310,A4389,A4391,A4500,A4506,A4514,A4522,A4524,A4550,A4568,A4586,A4591,A4593:A4594,A4596,A4599,A4603" _
), Range( _
```

```
"A4607,A4609,A4613,A4620,A4622,A4625,A4627,A4630,A4633,A4636,A4639,A4643,A4648,A4654,A4656,A4660,A4665,A4668,A4671,A:A,A60,A62,A64,A111,A119,A122:A126,A129,A156,A158,A197,A199,A266" _
), Range( _
```

```
"A268,A333:A334,A336,A347,A349,A408,A410,A414,A418,A421,A423,A425,A429,A432,A436,A439,A459,A461,A489" _
)),Select
```

```
Range("A2").Activate
With Selection
.WrapText = True
.Orientation = 0
.AddIndent = False
.IndentLevel = -1
.ShrinkToFit = False
.ReadingOrder = xlContext
.MergeCells = False
```

```
End With
```

```
Selection.Borders(xlDiagonalDown).LineStyle = xlNone
```

```
Selection.Borders(xlDiagonalUp).LineStyle = xlNone
```

```
With Selection.Borders(xlEdgeLeft)
```

```
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
```

```
End With
```

```
With Selection.Borders(xlEdgeTop)
```

```
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
```

```

End With
With Selection.Borders(xlEdgeBottom)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlEdgeRight)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlInsideVertical)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlInsideHorizontal)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
End Sub

```

O comando Ctrl+W , para a coluna B, representa o seguinte Macro:

```

Sub FormatarColunaB()
'
' FormatarColunaB Macro
'
' Keyboard Shortcut: Ctrl+w
'
Selection.SpecialCells(xlCellTypeBlanks).Select
ActiveWindow.SmallScroll Down:=57
Selection.Delete Shift:=xlUp
Union(Range( _
"B467,B681,B735,B737,B883,B885,B961,B963,B965,B967,B969,B973,B976,B981
,B986,B990,B998,B1000,B1006,B1009,B1011,B1014,B1016,B1019,B1022,B1026,
B1030,B1032,B1036:B1037,B1039,B1046,B1051" _
), Range( _
"B1056,B1058,B1069,B1073,B1096,B1098,B1115,B1146,B1169,B1171,B1241,B1
243,B1317,B1319,B1388,B1390,B1451,B1453,B1535,B1537,B1539,B1542:B1543,

```

B1545,B1548:B1549,B1554,B1556:B1557,B1562,B1564,B1566,B1568,B1627,B1629" _

), Range(_

"B1662,B1682,B1684,B1753,B1755,B1775,B1808,B1810,B1879,B1881,B1898,B1928,B1959,B1961:B1962,B1964,B1968,B1970,B1972,B1977,B1980,B1982,B1986,B1988,B1991,B1994,B1997,B2000,B2002,B2005,B2009,B2013,B2015" _

), Range(_

"B2018,B2022,B2025,B2027:B2028,B2030,B2071,B2073,B2080,B2082,B2084,B2087,B2104,B2106,B2152,B2154,B2196,B2198,B2254,B2256,B2321,B2323,B2348,B2350,B2376,B2378,B2409,B2412,B2414,B2417,B2422,B2426,B2429" _

), Range(_

"B2431,B2434,B2438,B2443,B2445,B2449,B2451,B2453,B2455:B2457,B2459,B2489:B2490,B2492,B2578,B2580,B2773,B2775,B2814,B2816,B2836:B2839,B2850,B2852,B2889,B2891,B2893,B2895,B2899,B2901,B2904,B2906,B2908,B2910,B2912" _

), Range(_

"B2914,B2916,B2918,B2920,B2923,B2928,B2930,B2932,B2934,B2938,B2940,B2945,B2949,B2951,B2983,B2985,B2993,B3087,B3139,B3141,B3201,B3203,B3237,B3239,B3266,B3268,B3306,B3308,B3379,B3381,B3414,B3416" _

), Range(_

"B3529,B3531,B3533,B3551:B3552,B3554,B3564,B3566:B3567,B3611,B3613,B3676,B3678,B3743,B3745,B3747,B3755:B3757,B3759,B3793,B3795,B3886,B3888,B3941,B3973,B3975,B4047,B4049,B4113,B4117,B4169,B4171,B4215,B4242,B4244" _

), Range(_

"B4246,B:B,B3,B61,B66,B68,B123,B126:B127,B129,B131,B153,B155,B188,B190,B249,B251,B314,B316,B370,B381,B384,B386,B389,B391,B393,B395,B397,B400,B403:B404,B406,B425,B427" _

), Range("B462")).Select

With Selection

.WrapText = True

.Orientation = 0

.AddIndent = False

.IndentLevel = -1

.ShrinkToFit = False

.ReadingOrder = xlContext

.MergeCells = False

End With

Selection.Borders(xlDiagonalDown).LineStyle = xlNone

Selection.Borders(xlDiagonalUp).LineStyle = xlNone

With Selection.Borders(xlEdgeLeft)

```

.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlEdgeTop)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlEdgeBottom)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlEdgeRight)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlInsideVertical)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
With Selection.Borders(xlInsideHorizontal)
.LineStyle = xlContinuous
.ColorIndex = 0
.TintAndShade = 0
.Weight = xlThin
End With
End Sub

```

Macro VBA gravado com o shortcut Ctrl+J
--

```

Sub JoinCellsMoveup()
If ActiveCell.Row > 1 Then
ActiveCell.Offset(-1, 0).Value = ActiveCell.Offset(-1, 0).Text & " " &
ActiveCell.Text
ActiveCell.Delete
End If
End Sub

```

Macro VBA gravado com o shortcut Ctrl+S
--

```

Sub SplitCell()
Dim s As String, v As Variant, l As Long
s = ActiveCell.Value
v = Split(s, vbLf)
l = UBound(v) - LBound(v) + 1
If l > 1 Then
ActiveCell.Offset(1, 0).Resize(l - 1, 1).Insert Shift:=xlShiftDown
ActiveCell.Resize(l, 1).Value = Application.Transpose(v)
End If
End Sub

```

Macro VBA gravado com o shortcut Ctrl+M,

```

Sub FindMismatch()
Dim r As Long
Dim m As Long
m = Range("A:B").Find(What:="*", SearchOrder:=xlByRows, _
SearchDirection:=xlPrevious).Row
For r = ActiveCell.Row + 1 To m
If Abs(UBound(Split(Range("A" & r).Value)) - _
UBound(Split(Range("B" & r).Value))) >= 9 Then
Application.Goto Range("A" & r), True
Exit For
End If
Next r
End Sub

```

Fórmulas criadas no *Excel*

Pasta de 'Alinhamento'

Função	Fórmula	Local
Contar palavras de uma determinada célula da coluna A.	=LEN(A1)-LEN(SUBSTITUTE(A1,"",""))+1	Fórmula utilizada na Fórmula seguinte.
Buscar pares de linhas A-B que apresentem uma diferença de 9 palavras absolutas (-/+9)	=IF(ABS((LEN(A1)-LEN(SUBSTITUTE(A1,"",""))+1)-(LEN(B1)-LEN(SUBSTITUTE(B1,"",""))+1))>=9,"MISMATCH",0)	Fórmula alocada na primeira célula da coluna C.

Pastas 'Rotulação TF' e 'Rotulação TA'

Fórmula	Significado
=IF(C3=1,\$C\$2,IF(D3=1,\$D\$2,IF(E3=1,\$E\$2,IF(F3=1,\$F\$2,IF(G3=1,\$G\$2,IF(H3=1,\$H\$2,IF(I3=1,\$I\$2,IF(J3=1,\$J\$2,IF(K3=1,\$K\$2,IF(L3=1,\$L\$2,IF(M3=1,\$M\$2,IF(N3=1,\$N\$2,IF(O3=1,\$O\$2,IF(P3=1,\$P\$2,IF(B3=1,\$B\$2,0))))))))))))))	Leva o nome da etiqueta à coluna R, na linha correspondente em que foi anotada (com o número 1).

Pasta 'Lista e Gráficos'

Fórmula	Significado
=COUNTIF(ROTULACAO TF!A3:A4000,"*?ranslator?*")	'ROTULACAO TF' corresponde à área de trabalho, ou seja, segunda pasta; A3:A4000 corresponde à coluna, * e ? são <i>Wildcards</i> que significam: * é qualquer ou nenhum caractere e ? é um qualquer caractere. Então * corresponde à todo o restante do texto da célula e ? a diferentes letras ou formatos de letras (ex. T maiúsculo ou t minúsculo, ou a presença ou ausência de um "s" indicador de plural) que podem ocupar essa posição.

Fórmula	Significado
----------------	--------------------

=COUNTIF(ROTULACAO TA!A3:A4000,"*?radutor?*").	<i>'ROTULACAO TA'</i> corresponde à área de trabalho, ou seja, segunda pasta; <i>A3:A4000</i> corresponde à coluna, * e '?' são <i>Wildcards</i> que significam: * é qualquer ou nenhum caractere e '?' é um qualquer caractere. Então * corresponde à todo o restante do texto da célula e '?' a diferentes letras ou formatos de letras (ex. T maiúsculo ou t minúsculo, ou a presença ou ausência de um "s" indicador de plural) que podem ocupar essa posição.
---	--

Fórmula	Significado
=COUNTIF(ROTULACAO TF!Q4:Q4001,"?*")	Número de elementos analisados no TF
=COUNTIF(ROTULACAO TA!Q4:Q4001,"?*")	Número de elementos analisados no TA

Anexo 10: Macros e Shortcuts Template MS Work

Shortcuts e Macros do Template Editor :

1) Alt+0: destaca os pontos

```
Sub Pontos()
'
' Pontos Macro
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
Selection.Find.Replacement.Highlight = True
With Selection.Find
.Text = "(.[\?!])"
.Replacement.Text = "\1"
.Forward = True
.Wrap = wdFindContinue
.Format = True
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

2) Alt+1 : elimina espaços duplos

```
Sub espacosduplos()
'
' espacosduplos Macro
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "[ @]{2,4}"
.Replacement.Text = " "
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
```

```

End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub

```

3) Alt + 2: invertem as aspas localizadas depois do ponto final

```

Sub pontoaspas()
'
' pontoaspas Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "([\.\?!\!]"") ([A-Z])"
.Replacement.Text = "\2\1 \3"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub

```

4) Alt + 3: elimina o ponto das abreviações de duas letras

```

Sub AbbrevDuasLetras()
'
' AbbrevDuasLetras Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "([A-Z][bcdfghjklmnpqrstvxwz]{1,2}). ([A-Z])"
.Replacement.Text = "\1 \2"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With

```

5) **Alt + 4: elimina pontos de abreviações com mais de três letras (casos de Rev. e de Prof.)**

```
Sub AbreviTresLetras()
'
' AbreviTresLetras Macro
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "( [PRM][evrofs]{1,3}). ([A-Z])"
.Replacement.Text = "\1 \2"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

6) **Alt + 5: elimina pontos de itens com letras minúsculas**

```
Sub MinuscPonto()
'
' MinuscPonto Macro
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "(<[a-z]{1}>). ([A-Z])"
.Replacement.Text = "\1 \2"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

7) **Alt + 6: quebra linhas em frases terminadas por parênteses (ou símbolos semelhantes) e pontos.**

```
Sub QuebraLinhaParenteses()
'
' QuebraLinhaParenteses Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "(\\|\\}|\\>\"")(\\.\\?\\!|) ([A-Z\""])"
.Replacement.Text = "\\1|2^1|3"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

8) **Alt + 7: quebra linhas após números (de três e quatro dígitos) mais pontos.**

```
Sub QuebraLinhaNumeros()
'
' QuebraLinhaNumeros Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "[0-9]{3,4}(\\.\\?\\!|) ([A-Z\""])"
.Replacement.Text = "\\1|2^1|3"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
```

End Sub

9) **Alt + 8: quebra linhas em geral, em frases terminadas por palavras em letras minúsculas e pontos.**

```
Sub QuebraLinhaGeral()
'
' QuebraLinhaGeral Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "(?!)[a-z](\?!) ([A-Z""])"
.Replacement.Text = "\1^2^1\3"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

10) **Alt + 9: quebra linhas em frases terminadas por palavras em letras maiúsculas e pontos.**

```
Sub QuebraLinhaMaiuscula()
'
' QuebraLinhaMaiuscula Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "(?!)[A-Z]{2}(\?!) ([A-Z""])"
.Replacement.Text = "\1^2^1\3"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
```

```
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

11) Alt+B: elimina os símbolos de Tab

```
Sub DeleteTabs()
'
' DeleteTabs Macro
'
'
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "([!][A-Z]{2})([.?!]) ([A-Z]*)"
.Replacement.Text = "\1\2^1\3"
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchAllWordForms = False
.MatchSoundsLike = False
.MatchWildcards = True
End With
Selection.Find.ClearFormatting
Selection.Find.Replacement.ClearFormatting
With Selection.Find
.Text = "^t"
.Replacement.Text = ""
.Forward = True
.Wrap = wdFindContinue
.Format = False
.MatchCase = False
.MatchWholeWord = False
.MatchWildcards = False
.MatchSoundsLike = False
.MatchAllWordForms = False
End With
Selection.Find.Execute Replace:=wdReplaceAll
End Sub
```

12) Alt+T: realiza todas as substituições mencionadas na ordem descrita.

```
Sub QuebraTextoSentencasTotal()
'
' QuebraTextoSentencasTotal Macro
'
'
Application.Run MacroName:="Normal.NewMacros.Pontos"
```

```
Application.Run MacroName:="Normal.NewMacros.espacosduplos"  
Application.Run MacroName:="Normal.NewMacros.pontoaspas"  
Application.Run MacroName:="Normal.NewMacros.AbbrevDuasLetras"  
Application.Run MacroName:="Normal.NewMacros.AbreviTresLetras"  
Application.Run MacroName:="Normal.NewMacros.MinuscPonto"  
Application.Run  
MacroName:="Normal.NewMacros.QuebraLinhaParenteses"  
Application.Run MacroName:="Normal.NewMacros.QuebraLinhaNumeros"  
Application.Run MacroName:="Normal.NewMacros.QuebraLinhaGeral"  
Application.Run  
MacroName:="Normal.NewMacros.QuebraLinhaMaiuscula"  
Application.Run MacroName:="Normal.NewMacros.DeleteTabs"  
Selection.ParagraphFormat.Alignment = wdAlignParagraphLeft  
Selection.WholeStory  
Selection.ParagraphFormat.Alignment = wdAlignParagraphLeft  
Selection.ParagraphFormat.Alignment = wdAlignParagraphLeft  
With Selection.ParagraphFormat  
    .SpaceBeforeAuto = False  
    .SpaceAfterAuto = False  
    .FirstLineIndent = CentimetersToPoints(0)  
End With  
End Sub
```