

Crop



REVISTA DA ÁREA DE
ESTUDOS LINGÜÍSTICOS E LITERÁRIOS EM INGLÊS
DEPARTAMENTO DE LETRAS MODERNAS DA UNIVERSIDADE DE SÃO PAULO

LINGÜÍSTICA DE CORPUS

GUEST EDITOR
STELLA E. O. TAGNIN



Universidade de São Paulo

Reitor Prof. Dr. Adolpho José Melfi

Vice-Reitor Prof. Dr. Hélio Nogueira da Cruz



Faculdade de Filosofia, Letras e Ciências Humanas

Diretor Prof. Dr. Sedi Hirano

Vice-Diretora Sandra Margarida Nitrini

Departamento de Letras Modernas

Chefe Prof. Dr. Mario Miguel González

Comissão Editorial

Anna Stegh Camati

Carlos Daghlían

Kanavillil Rajagopalan

Leila Darin

Regis Bonvicino

Sigrid Renaux

Solange Ribeiro de Oliveira

Victor Seidler

Editor Responsável

Stella E. O. Tagnin

Proibida a reprodução parcial ou integral desta obra por qualquer meio eletrônico, mecânico, inclusive por processo xerográfico, sem permissão expressa do editor (Lei n.º 9.610, de 19.02.98).

Todos os direitos desta edição reservados à:

FFLCH/USP

Rua do Lago, 717 – Cid. Universitária

05508-900 – São Paulo – SP – Brasil

Tel.: 3091-2920 / Telefax: 3091-4593

e-mail: editorahumanitas@usp.br

<http://www.flch.usp.br/humanitas>

Foi feito o depósito legal na Biblioteca Nacional (Lei n. 1.825, de 20/12/1907)

Impresso no Brasil / *Printed in Brazil*

Junho 2005

ISSN 1415-6253

Crop



REVISTA DA ÁREA DE LÍNGUA E
LITERATURA INGLESA E NORTE-AMERICANA
DEPARTAMENTO DE LETRAS MODERNAS

GUEST EDITOR
STELLA E. O. TAGNIN

Crop • n. 10 • p. 1-330 • São Paulo • 2004

UNIVERSIDADE DE SÃO PAULO • FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS

Copyright © 2004 dos autores

É proibida a reprodução parcial ou integral,
sem autorização prévia dos detentores do *copyright*

Serviço de Biblioteca e Documentação da FFLCH/USP
Ficha catalográfica: Márcia Elisa Garcia de Grandi CRB 3608

Crop: revista da área de língua e literatura inglesa e norte-americana do Departamento de Letras Modernas/Faculdade de Filosofia, Letras e Ciências Humanas. Universidade de São Paulo.– n. 1 (1994) –.– São Paulo: Humanitas /FFLCH/USP, 1994 –

Anual

ISSN 1415-6253

1. Língua inglesa. 2. Literatura inglesa 3. Literatura de expressão inglesa 4. Crítica literária 5. Tradução I. Universidade de São Paulo. Faculdade de Filosofia, Letras e Ciências Humanas.

CDD 420
820

Esta publicação é indexada por GeoDados: Indexador <<http://www.geodados.uem.br>>

HUMANITAS FFLCH/USP

Editor Responsável

Prof. Dr. Milton Meira do Nascimento

Coordenação Editorial, Projeto Gráfico e Capa

M^a Helena G. Rodrigues – MTb 28.840

Diagramação

Selma M. Consoli Jacintho MTb 28.839

Revisão

Stella E. O. Tagnin e Luciana C. F. Corrêa Pinto

Sumário

Apresentação	7
Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua	13
<i>Inacio Abdulkader</i>	
Idéias que Cruzam o Oceano	43
<i>Belinda Maia e Luís Sarmiento / Stella E. O. Tagnin e Sandra Maria Aluísio</i>	
Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais	65
<i>Marcia Epstein Fiker / Stela Foley</i>	
Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Lingüística de <i>Corpus</i>	113
<i>Helmara Febeliana Real de Moraes</i>	
Uma Análise Contrastiva Preliminar do Verbo <i>Limp</i> em Inglês e suas Traduções em Português do Brasil	139
<i>Kátia Hanna / Tereza Cristina Hilst</i>	
Verbos de Elocução – As Diferenças entre o Inglês e o Português	167
<i>Lais Loffredo / Deborah Grossman / Gladys Bitar / Janice Gonçalves</i>	
The Case of <i>laugh</i> : a Parallel Corpus-based Research	185
<i>Josimeire Cristina Martins</i>	

Crop, 10, 2004

Absolutamente e <i>Absolutely</i> – São ou Não São Absolutamente Intercambiáveis?	193
<i>Alvamar Helena de Campos Andrade Lamparelli</i>	
Linguística de Córpus e Lexicografia Bilíngüe: O Caso Experimental de <i>Como</i> e suas Traduções para o Francês	211
<i>Adriana Zavaglia</i>	
A construção do Sentido em Vocabulários Técnicos: o Uso de Corpora e Outros procedimentos	225
<i>Guilherme Fromm</i>	
Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos	241
<i>Diana Costa Fortier Silva / Renato Barros da Costa</i>	
Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Linguística de Corpus	261
<i>Benivaldo José de Araújo Júnior</i>	
A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil	279
<i>Catherine Anne Lonngren Sampaio</i>	
Linguística de Corpus e Análise Literária: uma Aplicação a <i>Dubliners</i> de Joyce	309
<i>Lourdes Bernardes Gonçalves</i>	
Normas de apresentação dos trabalhos enviados à CROP	329

Apresentação

Este número especial da revista *Crop* pretende dar um panorama das possibilidades de pesquisa baseadas e/ou direcionadas por corpora eletrônicos. A maioria dos artigos resultou de trabalhos elaborados pelos autores para alguma disciplina por mim ministrada na Universidade de São Paulo e na Universidade Estadual do Ceará, tendo por foco a Lingüística de Corpus.

A Lingüística de Corpus é uma abordagem que privilegia a observação de uma grande quantidade de dados autênticos com o intuito, entre outros, de identificar as ocorrências mais **prováveis** na produção de falantes nativos, em detrimento das ocorrências apenas **possíveis**. Em outras palavras, interessa à Lingüística de Corpus o que de fato o falante diz, não apenas o que ele poderia dizer, já que muitas dessas ocorrências possíveis jamais são realmente produzidas.

A observação de um grande número de dados a partir de corpora eletrônicos – coletâneas de textos reunidos com um objetivo específico de pesquisa e analisáveis com ferramentas computacionais próprias – também permite a visualização de padrões recorrentes, quer lexicais ou gramaticais, padrões que na maioria das vezes passariam despercebidos sem o uso desses recursos eletrônicos. Conhecer esses padrões recorrentes é essencial para o aprendizado de uma língua, pois vai fornecer ao aprendiz elementos para tornar sua produção mais natural, mais fluente, mais próxima da produção de um falante nativo.

Ainda nesse campo, os corpora de aprendizes – constituídos de redações não corrigidas de alunos – são extremamente úteis para identificar áreas problemáticas de aprendizagem, o que, por sua vez, pode servir de base para um ensino mais bem direcionado, assim como para a elaboração de material didático mais eficaz.

Uma das áreas que também se beneficia muito de estudos baseados em corpora é a da tradução. Corpora paralelos, isto é, textos originais juntamente com suas respectivas traduções, permitem ao pesquisador estudar as estratégias empregadas por tradutores profissionais, tanto num nível macrotextual de or-

Apresentação

ganização do texto, de acréscimos ou omissões norteadas ou não pela função da tradução no contexto de chegada, quanto num nível microtextual como, por exemplo, a tradução de nomes próprios, de termos culturalmente marcados, de termos técnicos ou mesmo do vocabulário de determinado campo semântico.

Outro tipo de *corpus* essencial para a identificação de vocabulário técnico bilíngüe é o *corpus* comparável – composto de textos originais nas duas línguas em estudo –, pois fornece ao pesquisador os termos empregados em contexto natural naquelas línguas, sem a possível ocorrência do que se convencionou denominar, em inglês, de *translationese*, ou, em português, de tradutês.

Todas essas áreas de pesquisa estão representadas neste número.

O artigo que abre o volume, *Carving Language at its Joints – Os Corpora como Ferramentas de (Des)Trinchar a Língua*, de Inacio Abdulkader, serve de introdução à Lingüística de Corpus, ou melhor, relata a primeira incursão do autor no mundo dos corpora ao fazer uma pesquisa sobre os verbos ingleses *have* e *go*. O autor chama atenção para a relevância da sintaxe da expressão de busca, isto é, a formulação da expressão que se pretende investigar. Sua pesquisa é feita no *corpus* COMPARA, um *corpus* paralelo bidirecional inglês-português de textos literários, que permite uma sintaxe de busca bastante elaborada. Abdulkader salienta também que o potencial dos corpora não se restringe a pesquisas lexicais, podendo oferecer aspectos de interesse para a Filosofia da Ciência e a Filosofia da Linguagem.

Segue-se *Idéias que Cruzam o Oceano*, uma apresentação de projetos similares e até complementares que estão sendo desenvolvidos nos dois lados do Oceano Atlântico: na Universidade do Porto, em Portugal, por Belinda Maia e Luís Sarmiento, e na Universidade de São Paulo, por Sandra Maria Aluísio e Stella E. O. Tagnin. Enquanto na Europa o enfoque é na criação de ferramental para a construção de corpora e glossários, como o *Corpógrafo*, ou para a avaliação da tradução automática, como o *TrAva*, no Brasil o objetivo é a própria construção de corpora para diversos fins, desde pesquisas lexicográficas até o desenvolvimento de corretores gramaticais e ortográficos, a saber, o portal Lácio-Web, que disponibiliza, entre outros, um *corpus* de português brasileiro com 10 milhões de palavras, e o Projeto COMET, composto de três corpora, um Corpus Técnico, um Corpus de Traduções e um Corpus de Aprendizizes.

Os dois artigos seguintes, *Os Corpora como Ferramentas para a Solução de Problemas de Tradução de Colocações Verbais*, de Márcia Epstein Fiker e Stela Foley, e *Traduzir Colocações Adverbiais. Tarefa fácil? Parece, mas não é*, de Helmara Febeliana Real de Moraes, abordam, como já se depreende dos títulos, a tradução de unidades de duas categorias convencionais, a das colocações verbais e a das colocações adverbiais. Enquanto no primeiro as autoras estudam detalhada-

mente as colocações com os verbos *make* e *perform*, e o substantivo *attention*, no segundo a autora detém-se em critérios quantitativos e qualitativos para identificar as informações que devem integrar os verbetes de um glossário de colocações verbais. Para tanto, a título de exemplo, apresenta o verbete para a colocação *deeply hurt*.

A análise do vocabulário de determinados campos semânticos é o tópico dos próximos três artigos.

Kátia Hanna e Teresa Cristina S. Hilst fazem uma minuciosa análise contrastiva do verbo *limp* e suas possíveis traduções para o português, com base em dados levantados do *corpus* COMPARA, um *corpus* paralelo bidirecional inglês-português de textos literários.

Os *Verbos de Elocução* são estudados por Laís Cardoso Loffredo, Deborah Grossman, Gladys Bitar e Janice Gonçalves, também de uma perspectiva contrastiva, conforme ocorrem no *corpus* COMPARA. As autoras observam que nos textos originais em língua inglesa há muita repetição do verbo *said*, enquanto nos em língua portuguesa há um empenho em empregar um vocabulário mais variado. Esse padrão, no entanto, não ocorre na mesma proporção nos textos traduzidos, gerando, por vezes, o fenômeno a que aludimos acima, o tradutês.

Nessa mesma linha, o artigo de Josimeire Martins, *The Case of laugh: a parallel-corpus-based research*, explora os verbos do campo semântico de *laugh*, ainda sob uma perspectiva contrastiva baseada em dados do *corpus* COMPARA.

Um *corpus* comparável é utilizado por Alvamar Lamparelli para estudar um problema de cognatos em *Absolutamente e Absolutely – são ou não são absolutamente intercambiáveis?*. Examinando as ocorrências desses vocábulos em corpora de textos jornalísticos originais, em inglês e em português, a autora conclui que, embora alguns itens possam parecer inicialmente equivalentes, diferenciam-se pelos seus padrões de co-seleção.

Adriana Zavaglia, em *Linguística de Corpus e Lexicografia Bilingüe: o caso experimental de como e suas traduções para o francês*, amplia o escopo das pesquisas aqui relatadas. Em primeiro lugar, seu trabalho aborda a língua francesa, em segundo, entra no campo da lexicografia, sugerindo, inclusive, um verbete bilingüe experimental, que privilegia a definição e a contextualização do marcador *como*.

Da lexicografia passamos à terminologia com o trabalho de Guilherme Fromm, *A Construção do Sentido em Vocabulários Técnicos: o uso de corpora e outros procedimentos*. O autor, ao propor um modelo de elaboração de vocabu-

Apresentação

lários técnicos, discute diversas abordagens de construção do sentido à luz da Semântica e da Lingüística de Corpus.

Ainda dentro dos estudos lexicais, Diana Costa Fortier Silva e Renato Barros da Costa oferecem-nos um estudo dos *Anglicismos no Português do Brasil: uma perspectiva de análise de usos baseada em corpora eletrônicos*. Os autores selecionam quinze anglicismos e examinam seus usos e suas acepções na língua inglesa e as eventuais modificações que sofrem ao passarem para a língua portuguesa.

Uma questão sintática é o tópico de estudo de Benivaldo José de Araújo Júnior, em *Estruturas Passivas no Espanhol Escrito: um estudo de caso utilizando a Lingüística de Corpus*. A partir da observação de que os livros didáticos e manuais de espanhol como língua estrangeira pouca ou nenhuma relevância dão às estruturas passivas – mencionando apenas a voz passiva –, o autor lança mão da Lingüística de Corpus para melhor estudar o fenômeno, partindo da hipótese de que a análise de uma grande quantidade de ocorrências pode revelar dados inesperados que permitirão compreendê-lo melhor.

O artigo seguinte, *A Investigação de Mudança de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*, é um exemplo claro de como a construção de um *corpus* é ditada pelos objetivos da pesquisa a que se destina. Para estudar a alternância de código (*code switching*) de duas crianças bilingües aprendendo inglês e português simultaneamente, a autora, Catherine Anne Lönngren Sampaio, construiu um *corpus*, gravando vários tipos de interações dos dois sujeitos, ao longo de dois anos. Para transcrever o material, seguiu as normas do projeto CHILDES (Child Language Data Exchange System), mas, para atender às peculiaridades de sua pesquisa, inseriu outros códigos que lhe permitissem identificar os momentos em que ocorria a mudança de código. Trata-se de um dos raros trabalhos – senão o único – que se vale de um *corpus* eletrônico de língua falada para estudar os estágios iniciais do bilingüismo infantil.

Mas a metodologia da Lingüística de Corpus também se aplica aos estudos literários. É o que demonstra o trabalho de Lourdes Bernardes Gonçalves, *Lingüística de Corpus e Análise Literária – uma aplicação em Dubliners de Joyce*. A partir de um *corpus* composto pelos quinze contos irlandeses que compõem a coletânea *Dubliners*, James Joyce, a autora faz um levantamento eletrônico do vocabulário mais recorrente nos contos e compara-o ao vocabulário de outros contos da mesma época, porém escritos na Inglaterra, o que lhe permite identificar o léxico peculiar empregado pelo autor. A análise desse léxico – só identificável por meio dos procedimentos oferecidos pela Lingüística de Corpus – revela o papel da música na construção dos personagens joyceanos.

Acreditamos que a seleção dos artigos que compõem este número especial da *Crop* deixe clara a relevância e abrangência da Lingüística de Corpus nas

Crop, 10, 2004

várias áreas dos estudos lingüísticos, como a lexicologia, a terminologia, a tradução, o ensino e a aprendizagem de língua estrangeira e a sintaxe, entre outros, assim como nos estudos literários, em que sua aplicação ainda é incipiente, porém promissora.

Gostaria, ainda, de deixar registrado meu agradecimento a Luciana Carvalho Fonseca Corrêa Pinto, que coordenou a coleta e o acompanhamento dos artigos que integram este volume. Sem a ajuda dela, este trabalho não teria sido publicado no prazo estipulado.



Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua

*Inacio Abdulkader**

Resumo: *O presente artigo apresenta os resultados de uma pesquisa de colocações verbais de have e go de interesse para o tradutor, que foi realizada utilizando-se corpora eletrônicos (COMPARA, Cobuild e WebCorp). Foram dois os tipos de buscas realizadas: corpus driven, e corpus based, e através delas foi possível o levantamento de um número de colocações que é bastante significativo frente ao que a lexicografia oferece para dois verbos tão freqüentes e de tal “dimensão gramatical” na língua. Ademais, procura-se chamar a atenção para a importância do poder expressivo da sintaxe dos strings de busca dos corpora que deve a nosso ver merecer uma maior atenção do pesquisador em corpora. No caso do COMPARA p. ex., essa sintaxe propiciou a construção de um “dicionário” que, muito provavelmente, incluiu a totalidade das ocorrências dessas colocações verbais no corpus. Além disso, o potencial dos corpora para pesquisas que vão além do aspecto lexical, envolvendo, p. ex., aspectos de interesse da Filosofia da Ciência e a da Linguagem é discutido e exemplificado.*

Palavras-chave: *corpora eletrônicos; colocações verbais; colocações.*

Abstract: *This paper presents the results of a study on verbal collocations with “have” and “go”, which are of interest to translators working into Portuguese. Electronic corpora such as COMPARA, Cobuild and WebCorp were the main tools used to obtain the results*

* O autor é engenheiro e mestrando do programa de Estudos Lingüísticos e Literários em Inglês da FFLCH-USP (kader@allnet.com.br).

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

shown. Two types of searches were carried out, one “corpus driven” and one “corpus based”, which produced a significant number of collocations when compared to what is offered by the lexicographic aids for two verbs of such “grammatical dimension” within the language. We also discuss the expressive power of the syntax of the search string in the corpora, which proved to have a strength that, we believe, deserves closer attention. In the case of COMPARA, for instance, it permitted the construction of a “dictionary” that, probably, includes all occurrences of verbal collocations with “have” and “go” in that corpus at the time of search. We also point out, and try to exemplify, the potential of electronic corpora for research beyond the lexical level, including, for instance, aspects which are of interest to the Philosophy of Science and the Philosophy of Language.

Keywords: *electronic corpora; verbal collocations; collocations.*

1. Introdução

O presente artigo apresenta os resultados de uma pesquisa sobre colocações verbais de *have* e *go* com ênfase em aspectos de interesse do tradutor, e que utilizou como ferramenta básica os corpora eletrônicos. Essa pesquisa propiciou ademais ao autor seu primeiro contacto com os corpora, cuja existência até então desconhecia.

Assim, um objetivo adicional do artigo será o de descrever um pouco da experiência desse primeiro contacto com a pesquisa em corpus e deixar registradas algumas impressões sobre a ferramenta recém manuseada. Dessa forma, no que se segue haverá algum relato subjetivo e alguma digressão, além da usual apresentação objetiva de materiais, métodos e resultados.

O contexto é o de um aluno do primeiro semestre do CETRAD, o Curso de Especialização em Tradução com duração de quatro semestres que o Departamento de Línguas Modernas da FFLCH-USP oferece a graduados de todas as áreas que possuam o domínio requerido da língua de escolha.

Os resultados aqui apresentados são parte de um trabalho mais geral¹ enfocando colocações de vários outros verbos, elaborado por um grupo de qua-

¹ Souza, Beatriz, Inacio Abdulkader, Marcia Fiker e Stela Foley. “Colocações Verbais: os corpora como ferramenta para solução de problemas de tradução”, monografia submetida à disciplina de Lingüística Contrastiva II do CETRAD, DLM, FFLCH-USP, 2o semestre 2002.

Crop, 10, 2004

tro alunos para a avaliação de conclusão da disciplina “Linguística Contrastiva II” ministrada no segundo semestre de 2002. As três outras componentes do grupo cursavam àquela altura o último semestre e já tinham portanto tido algum contato com a pesquisa em *corpus*.

Deve-se ademais registrar a gama das formações profissionais no grupo – uma física, uma psicóloga e uma bacharel em letras, além deste engenheiro – que é típica do ambiente multiprofissional que o CETRAD propicia, e que se revelou enriquecedora, dada a diversidade dos enfoques e a troca de experiências que proporcionou.

2. Corpus e Colocações

Cabe de início definir, talvez com um simplismo próprio de iniciante, o que é um *corpus* e o que são colocações.

Corpora eletrônicos são bancos de dados de textos. Ou seja, coleções de textos que são reunidos e catalogados de acordo com algum critério, armazenados de forma eletrônica, e que podem ser acessados ou apresentados de forma organizada ao usuário, conforme critérios que o usuário define.

“Critério” e “organização” são os conceitos-chave. Num *corpus* que contivesse todas as peças de Shakespeare, o usuário poderia, por exemplo, solicitar que lhe fossem apresentadas todas as ocorrências da expressão “*whether it's nobler*” e mais, para uma primeira análise, as quinze palavras que a antecedem e lhe seguem em cada uma dessas ocorrências. Ou então solicitar da mesma forma as ocorrências de “*to be*”, caso em que o *corpus* apresentaria um número máximo de ocorrências, digamos quinhentas, estocasticamente colhidas dentre o total de instâncias de “*to be*” nas peças de Shakespeare. O texto mais amplo que inclui qualquer uma dessas ocorrências pode também ser exibido caso haja interesse.

Há portanto corpora de todos os tipos, organizados por autor, por assunto, por gênero, por época, corpora de registros da oralidade, corpora bilíngües e multilíngües, etc. A forma de apresentação pode também caracterizar o *corpus*. O COMPARA², por exemplo, é um *corpus* literário, bilíngüe e paralelo. Inclui originais em português ou inglês e as respectivas traduções para a outra língua (várias traduções da mesma obra em alguns casos). Nele uma ocorrência é sempre apresentada de forma sinóptica com sua respectiva tradução (daí a denominação “corpus paralelo”).

² <http://www.linguateca.pt>

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

Um outro tipo de informação é a que se pode obter pela função *collocate sampler*, disponível em alguns corpora. Através dela as palavras que mais co-ocorrem com um determinado termo no corpus são listadas por ordem de frequência. Assim, um corpus que se componha de todas as edições de algum jornal brasileiro de grande circulação a partir de janeiro de 2003, com certeza indicará como *collocates* de alta frequência da palavra “Presidente” as palavras “Lula” e “Bush” (*hopefully* nessa ordem). E da palavra “Ministro”, as palavras “José” e “Dirceu”. Como se vê, pode-se extrair bem mais do que meras informações lingüísticas de um corpus.

Para o presente trabalho foram realizadas buscas no COMPARA e no Cobuild³, que é um corpus de textos os mais variados em inglês, já “fechado” – ou seja ao qual não se acrescentam novos textos – e que inclui cerca de quarenta milhões de palavras. Consultou-se também o WebCorp⁴ um *quasi-corpus* que na realidade usa toda a internet como seu banco de textos, permitindo, se necessário, algum controle da amplitude da busca (via buscador especificando-se, p. ex., google ou yahoo, via endereço, p. ex. PubMed, ou via domínio, p. ex., .br ou .uk).

Colocações, por outro lado, têm a ver com a convencionalidade da língua, com o “jeito como se fala”. Com a tendência de certas palavras de co-ocorrerem – aparecerem juntas com frequência na língua – e com o fato de que outras expressões, embora semanticamente corretas, são simplesmente descabidas na língua em uso. No que concerne ao tradutor, contrastes tais como “**have** breakfast” e “**tomar** café [da manhã]”, “**go** nuts” e “**ficar** doido”, mas não “ir doido” ou “ir pras nozes”.

Trata-se ademais de poder reconhecer se um dado padrão lexical tem ou não um conteúdo semântico próprio e que difere, naquele caso específico, da “soma dos significados” dos componentes. Situações como a de, por exemplo, “*have us over a barrel*”, que pode ser bem complicada, e não só para o tradutor.

Apresentado o contexto do pesquisador iniciante, e definidos o problema e a ferramenta, passemos a uma apresentação da metodologia e dos resultados.

3. Dificuldades iniciais e definição da metodologia

A busca de colocações verbais de *have* e *go* desenvolveu-se em duas frentes. A primeira, uma busca aberta, *corpus driven* – ou seja, direcionada em grande parte pelos próprios dados extraídos dos corpora conforme iam sendo obtidos –, visou extrair do COMPARA o maior número possível de colocações verbais de *have* e *go* e suas respectivas traduções. A segunda, *corpus based* – ou seja, baseada na

³ <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

⁴ <http://www.webcorp.org.uk>

busca de colocações verbais específicas – realizou a busca de determinadas colocações nos vários corpora visando a análise de aspectos semânticos particulares ou gerais que interessam ao tradutor, como os que mencionamos na seção 2 acima.

Circunstâncias específicas envolveram a procura e a definição da metodologia:

- 1) A tarefa foi inicialmente entendida, por quem estava tendo seu primeiro contacto com a pesquisa em corpora, como sendo a de “introduzir esses verbos nos corpora para obter suas colocações”.
- 2) A pesquisa se concentrou em dois dos verbos de maior frequência na língua inglesa, sendo um deles, aquele pelo qual se iniciou a pesquisa, um verbo auxiliar.
- 3) A pobreza da lexicografia, tão decantada pelos tradutores, é nesse caso ainda mais aguda. Embora existam dicionários de colocações (o mais recente o da Oxford de 2002) nos quais para cada verbete se listam as palavras que com ele mais co-ocorrem na língua, *have*, por ser verbo auxiliar, nem consta como verbete. Aparece citado ao longo de todo o dicionário nos verbetes cabíveis (p. ex., nos verbetes de *bath* e *drink*). E o verbete de *go* se restringe às mais óbvias colocações com preposições ou advérbios (*go well*, *go with*). Algo tão corriqueiro quanto *go home*, p. ex., aparece apenas no verbete *home*.
- 4) A pesquisa iniciou-se francamente *corpus driven* (*‘avant la lettre’* para quem a realizava) sendo que as palavras de busca inicialmente usadas nos corpora (Webcorp, Cobuild e COMPARA), foram as formas básicas de “*to have*”. Assim, as primeiras buscas geraram uma quantidade não administrável de soluções, a imensa maioria delas com *have* na função de verbo auxiliar.

Essa situação gerou duas iniciativas que permitiram a definição de uma metodologia que conduziu aos resultados almejados:

- a) O suporte inestimável do pessoal de apoio do COMPARA possibilitou a construção de *strings* de busca que “filtraram” os resultados a favor de instâncias mais “léxicas” que propiciassem a ocorrência de colocações verbais. Com isso conseguiu-se extrair do COMPARA um considerável número de colocações verbais de *have* e *go*, algumas delas com várias ocorrências.
- b) Foi rapidamente reconhecida a necessidade de complementar-se as buscas do tipo acima com buscas de colocações específicas nos outros corpora, de forma a confirmar que alguns casos tratavam-se efetivamente de uma “colocação”. Além disso, essa busca de colocações específicas foi realizada sempre que havia interesse de se estudar mais a

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

fundo determinadas colocações. Assim, procurou-se reconhecer alguns pontos de natureza semântica importantes para o trabalho do tradutor.

4. Resultados das buscas *corpus driven*

Foram dois os tipos de *strings* que o suporte do COMPARA nos proveu, de forma a contornarmos a “explosão” do *have* auxiliar.

O primeiro tipo pode ser exemplificado pelo *string* que transcrevemos a seguir, e que é o que consta da primeira linha do item b1 na Tabela 1, onde é denominado de filtro 1:

“have|had”[word!=”.*ed|been|made|brought|done”]{4,4}

TABELA 1
PESQUISA DE COLOCAÇÕES VERBAIS DE “HAVE” NOS CORPORA

a) **Webcorp** (sem filtros*): “have”, “have - *ed”, “had”: 1902 *concordances*.

11 colocações verbais
(6 coincidentes com obtidas no COMPARA)

0,57%

b) **COMPARA** (com filtros): 1967 *solutions*,

32 colocações verbais

1,62%

b1) filtro 1 (dois casos):

“have|had”[word!=”.*ed|been|made|brought|done”]{4,4},
1787 *solutions*, 500 *shown*

“having|has” [word!=”.*ed|been|made|brought|done”]{4,4},
467 *solutions*

b2) filtro 1 a:

“have|had”[word!=”.*ed|been|made|brought|done|thought|seen|br
oken|left|bought|sent|had|given|to|set”]{4,4}, 1200 *solutions*, 500
shown

b3) filtro 2:

“have|had” “a|the|my|some|to”, 608 *solutions*, 500 *shown*

c) **Cobuild**(sem filtros): 80 *concordances*, 1 colocação verbal (coincidente com obtida no Webcorp)

d) **Cobuild collocater sampler**: 1 colocação verbal (não coincidente com as acima)

TOTAL: 38 COLOCAÇÕES VERBAIS DE “HAVE”

Ele é “traduzido” como: “busque as formas “*have*” ou “*had*” seguidas por (exatamente) quatro ocorrências de formas que não terminam em “*ed*”, e que não são as formas “*been*”, “*made*”, “*brought*”, “*done*.”

O *string* na segunda linha do item b1 na **Tabela 1** é igual ao acima com a diferença de que é aplicado para as formas “*having*” e “*has*”. E o **filtro 1 a**), na mesma tabela, é também um *string* deste primeiro tipo, apenas mais restritivo, pois “veta” que a “*have*” e “*had*” se siga um maior número de participios irregulares.

O **segundo tipo** de *string* tem a sintaxe do **filtro 2** que consta no item b3 da **Tabela 1**:

“**have | had**” “**a | the | my | some | to**”

Essa sintaxe busca os casos em que é provável que um participio não se siga à expressão buscada, no caso, “*have a*”, “*have some*”, “*have the*”.

Esses “filtros” tornaram a busca no COMPARA manejável e permitiram obter desse corpus 32 tipos de colocações verbais de *have* (**Tabela 1**), num total de 56 ocorrências. A **Tabela 1** mostra também os números bem mais reduzidos de colocações obtidos nas buscas (sem filtros) de *have* no corpus do Cobuild, e também no Webcorp, sendo que esta última busca tornou-se manejável apenas na medida em que os resultados do COMPARA nos alentaram a enfrentá-la. Realizamos também uma busca com o *collocater sampler* do Cobuild que nos propiciou obter apenas um novo tipo de colocação verbal.

Ao fim, totalizamos 38 diferentes colocações verbais de *have*, listadas na **Tabela 2**. Esse número é significativo frente ao que os dicionários e correlatos oferecem. Das colocações levantadas através dessa literatura, que são quantificadas na **Tabela 3**, nem todas são verbais. Normalmente aparecem sob outros títulos (p. ex. “*Have*” em Swan: 2002), visando ilustrar outros aspectos e não o fato de se tratarem de colocações verbais. E são, geralmente, colocações bem conhecidas.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

TABELA 2
Colocações de Verbais "Have" Obtidas

Have a baby	Have a termination
Have a bath**	Have an affair
Have a bite	Have an argument
Have a child	Have an operation
Have a degree	Have breakfast
Have a go	Have eyes for
Have a good laugh	Have food
Have a haircut	Have fun
Have a head start*	Have in mind
Have a look	Have priority
Have a nice time*	Have qualms***
Have a nightmare	Have second thoughts
Have a part*	Have sex
Have a role*	Have some sleep
Have a row	Have the curse
Have a say	Have the heart
Have a scotch	Have the nerve
Have a seat	Have therapy
Have a temperature	Have trouble

(*) só no Webcorp

(**) no Webcorp e no Cobuild

(***) do collocata sampler do Cobuild

todas as demais colocações são do COMPARA

TABELA 3
COLOCAÇÕES DE "HAVE" EM MANUAIS, DICIONÁRIOS E GRAMÁTICAS

a) Oxford Collocations Dictionary (2002): nenhuma colocação (i.e., não tem verbete para "have").

b) Cambridge Grammar of the English Language (2002): 6 colocações verbais, 4 colocações "não-verbais". ("Collocations" não consta do índice remissivo de conceitos).

c) Bloomsbury Good Word Guide (1990): nenhuma colocação

d) Swan, Practical English Usage (1995): 14 "tipos" de colocações verbais de "have"— p. ex. "have lunch" e "have dinner" são consideradas do mesmo "tipo". (Citadas no parágrafo sobre "have: actions" e não sob "collocations").

e) Tagnin, Expressões Idiomáticas e Convencionais (1989): 4 colocações verbais, 3 "não-verbais".

f) Collins Dictionary of Phrasal Verbs and their Idioms (1974): nenhuma colocação verbal, 6 "phrasal verbs".

Crop, 10, 2004

A busca de colocações verbais de *go* no COMPARA seguiu procedimentos muito semelhantes aos adotados para a busca dos de *have*. Não sendo *go* um verbo auxiliar, não ocorreu aqui nenhum fenômeno “explosivo” como no caso de *have*. Em nenhuma situação o total de ocorrências ultrapassou as quinhentas que são o número máximo de exibições do COMPARA. Como se vê na **Tabela 4**, para as formas “*go*” e “*goes*” não foi necessário utilizar nenhum filtro.

TABELA 4
PESQUISA DE COLOCAÇÕES VERBAIS DE “GO” NOS CORPORA

a) COMPARA total de:
22 colocações 3,26%
“ <i>go</i> ” e “ <i>goes</i> ” (sem filtros): 310+58 <i>solutions</i>
“ <i>went gone</i> ” e “ <i>going</i> ” (com os filtros abaixo): 202 + 104 <i>solutions</i>
"went gone" [word!=".*ed to away somewhere anywhere nowhere on off into"]{4,4}
e
"going" [word!=".*ed to away somewhere anywhere nowhere on off into"]{4,4}

b) Cobuild (sem filtros): 80 <i>concordances</i> , 2 colocações não coincidentes com o Compara.

c) Cobuild <i>collocate sampler</i> : 6 colocações (não coincidentes com as acima).

d) Webcorp (busca específica): 1 colocação .
--

TOTAL: 31 COLOCAÇÕES DE “GO”

O filtro utilizado para “*went|gone*” e “*going*”, transcrito abaixo para o caso de *went|gone* (item a da **Tabela 4**), é um *string* do primeiro tipo acima descrito. “*went|gone*”[word!=".*ed|to|away|somewhere|anywhere|nowhere|on|off|into"]{4,4}

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

Sua aplicação não foi ditada por qualquer “explosão” de resultados, que, como vimos, não ocorreu nas buscas de *go*. Ela visou apenas facilitar o *sorting*, evitando alguns advérbios de lugar e movimento, *ing-words* (via “*on*”, “*off*” e “*into*”), e participios (via “**ed*”).

Do total das 31 colocações verbais de *go* levantadas (vide **Tabela 5**), 22 são do COMPARA, totalizando 40 ocorrências de colocações verbais de *go* nesse corpus. Não sendo *go* um verbo auxiliar – portanto mais “léxico” ou menos “gramatical” que *have* – o *collocate sampler* do Cobuild forneceu 6 interessantes colocações contra apenas 1 de *have*. Da **Tabela 6**, vê-se que o número de colocações verbais por nós obtidas é, no caso de *go*, ainda mais significativo, frente ao (quase nada) que se consegue extrair da literatura.

TABELA 5
Colocações Verbais de “Go” Obtidas

Go amiss*	Go in late	Go public***
Go astray*	Go into detail	Go quiet
Go awry*	Go mad	Go to any trouble
Go bald	Go nuts	Go to bed
Go bankrupt	Go off beat**	Go to perdition
Go berserk*	Go off the rails	Go to prison
Go blank	Go on stage	Go to sea
Go crazy	Go on the run***	Go wide
Go haywire*	Go out of (their) way	Go wild
Go his own way	Go overboard*	Go wrong
		Go years back

* do *collocate sampler* do Cobuild
** busca específica no Webcorp
*** do Cobuild
todas as demais colocações são do COMPARA

TABELA 6
COLOCAÇÕES DE “GO” EM MANUAIS, DICIONÁRIOS E GRAMÁTICAS

-
- a) **Oxford Collocations Dictionary (2002):** nenhuma colocação verbal, 15 colocações adverbiais e phrasal verbs .
- b) **Cambridge Grammar of the English Language (2002):** nenhuma colocação verbal, 2 colocações “não-verbais” (“Collocations” não consta do índice remissivo de conceitos).
- c) **Bloomsbury Good Word Guide (1990):** nenhuma colocação.
- d) **Swan, Practical English Use (1995):** nenhuma colocação.
- e) **Tagnin, Expressões Idiomáticas e Convencionais (1989):** 4 colocações verbais, 1 “não-verbal”.
- f) **Collins Dictionary of Phrasal Verbs and their Idioms (1974):** nenhuma colocação verbal, 37 phrasal verbal.
-

Ademais, é bem provável que tenhamos levantado **todos** os tipos de colocações verbais de *go* existentes no COMPARA à data das buscas, e a **quase totalidade** dos de *have*. Isso demonstra o considerável poder expressivo da sintaxe de busca do COMPARA, que merece ser comentado com mais detalhe.

4.1 O potencial dos “filtros”.

À parte os “filtros” terem sido fundamentais para a viabilização de nossas buscas por colocações verbais de *have*, sua utilização indicou que a sintaxe dos *search strings* do COMPARA é suficientemente forte (ou seja, consegue exprimir suficientes conteúdos semânticos) para merecer uma maior atenção do tradutor, de forma a que o potencial de extração de dados desse corpus possa ser devidamente explorado. No presente caso ao menos, e que certamente é típico de uma série de outras buscas⁵, tudo indica que os “filtros” compensaram muito satisfatoriamente o fato de o COMPARA não ser um

⁵ Em geral as expressões cuja sintaxe por um lado deve incluir palavras ou formas de alta freqüência na língua, e por outro deve excluir outras palavras ou formas também muito freqüentes na língua. Uma solução que conseguiu “filtrar” as formas em que um verbo funciona como auxiliar, servirá para uma série de outras situações que requerem um menor *screening*.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

corpus etiquetado⁶. E que podem chegar a compensá-lo plenamente, em termos de propiciar a extração da totalidade das colocações existentes nesse corpus.

Recorde-se de início que é através de um *sorting* das soluções exibidas pelos corpora – i. e. do exame de cada solução – que se localizam as colocações existentes no corpus. E que para todos os três corpora apenas uma pequena porcentagem das soluções exibidas (entre 0,5 e 3,3%, vide Tabelas 1 e 4) são colocações. Graças à utilização dos filtros, as porcentagens mais altas são as do COMPARA.

No que respeita especificamente ao COMPARA e seus filtros, note-se que em apenas três de nossas buscas (que são, todas as três, das formas “*have | had*”) o número de soluções ultrapassou as quinhentas, que é o número máximo de soluções que podem ser apresentadas numa exibição. Isso determinou para esses casos uma escolha randômica pelo sistema das soluções que foram exibidas. Em todas as outras buscas (que incluem a totalidade das buscas de *go*) o número de soluções foi inferior a 500, o que indica que teremos levantado nesses casos, salvo falta de atenção ou erro de julgamento, todas as colocações verbais das respectivas *wordforms* no COMPARA.

Os três casos com mais de 500 soluções foram: uma busca com o filtro 1 que obteve 1787 soluções, outra com o filtro 1a) que obteve 1200, e a terceira com o filtro 2 obtendo 608 soluções.

Por um lado, isso indica que devem (ou podem) existir no COMPARA colocações de *have* que não nos foram exibidas. Por outro lado, um cálculo de probabilidade bem simples (σ^3 de V para curva normal) indica que mais uma reiteração de exibição da busca do filtro 2, duas mais de exibições do filtro 1 a) e três do filtro 1, propiciarão a exibição da totalidade das colocações verbais de *have* no COMPARA com mais de 99,9% de probabilidade.

Os casos que parecem ser mais críticos são os dos filtros 1 a) e 1, em que o número de soluções exibidas é bem inferior (respectivamente, cerca de duas e três vezes) ao de soluções encontradas. Mas como o filtro 1 a) gera subconjuntos das soluções do filtro 1, as exibições do filtro 1 a) constituem uma espécie de reiteração parcial das exibições do filtro 1. Isso reduz para duas por filtro o número de reiterações requeridas para que se consiga a exibição – estatística – do total de colocações constantes nas soluções de 1 a) e 1. Já no caso do filtro 2,

⁶ Corpora etiquetados são aqueles cuja base de dados contém etiquetas morfossintáticas indicativas da categoria gramatical das palavras. Ideal para o caso do presente estudo seria um corpus que etiquetasse todas as colocações verbais que ocorrem em seus textos.

Crop, 10, 2004

o número de soluções por exibição é bem próximo ao de soluções totais (500 de 608), de forma que uma reiteração, como vimos, já é suficiente.

Há ainda um recurso de outra natureza: a melhora da **eficiência** das exibições pelo aumento da porcentagem de soluções relevantes em cada exibição de 500 soluções. Isso pode ser obtido pelo aumento da lista de palavras restritivas nos filtros de primeiro tipo. Para tanto, serve como orientação o resultado (ou melhor, as soluções descartáveis) de exibições anteriores, menos restritivas.

Assim, através de recursos como os acima descritos e de outros que se venha a desenvolver, pode-se evitar a explosão do número de soluções irrelevantes que tanto desalenta (como desalentou o iniciante no seu primeiro input, *naive*, no COMPARA).

E esse refinamento do *string* de busca tornar-se-á cada vez mais importante com o crescimento da base de textos do COMPARA. Ele será fundamental, por exemplo, para que no futuro o trabalho envolvido no *sorting* de soluções exaustivas do corpus mantenha-se, como é hoje, dentro do que é razoavelmente manejável.

A lição é clara mesmo para quem apenas se inicia na busca em corpus: a sintaxe de busca dos corpora, em especial para o tradutor a do COMPARA, tem um poder de expressão semântica dos requisitos de busca que merece ser mais bem explorado.

É importante que se aprofunde o estudo da sintaxe de busca dos corpora, para que o *string* de busca vá além de uma passiva introdução da palavra, morfema, ou expressão no campo de *search*, na esperança de que as exibições surjam, numerosas e relevantes.

Com isso o potencial da ferramenta corpus passa a ser explorado de forma mais plena e eficaz.

4.2 Alguns resultados do COMPARA

Abaixo, nos restringimos a apresentar cinco casos de colocações verbais de *have* e *go* menos usuais, obtidas no COMPARA através das buscas acima descritas. Algumas colocações “não verbais”, ou seja *phrasal verbs*, expressões idiomáticas ou fixas obtidas nessas buscas e que consideramos ser de interesse para o tradutor, são apresentadas no **Apêndice 1**.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

EBDL1T1 (1660):	<i>At that moment I was surprised to hear Lewis Parker saying from the set, «Well, if she is pregnant she'll have to have a termination,» and Debbie replying, «I suppose you think that will solve everything.»</i>	Nesse momento surpreendeu-me ouvir as palavras do Lewis Parker vindas do cenário: «Bem, se ela está grávida, tem de fazer um aborto.» E a Debbie responder-lhe: «Deves achar que isso resolve tudo.»
EBDL1T2 (1945):	<i>I had a go at The Concept of Dread this evening - thought I'd start with the title that seemed most obviously relevant to me - but it was a great disappointment.</i>	Comecei a ler <i>O conceito de angústia</i> hoje à noite - achei que seria bom começar pelo título que mais combinava comigo --, mas foi uma grande decepção.
EBIB1(949):	<i>Avoir ses ours? , to have one's bears, means 'to have the curse' (presumably because at such times the woman is supposed to behave like a bear with a sore head) .</i>	<i>Avoir ses ours</i> significa estar com o período (possivelmente porque se pensa que nessas alturas a mulher tem o comportamento de um urso com mau perder).
EBDL1T2 (1973):	<i>I would say that learning to know dread is an adventure which ever man has to affront if he would not go to perdition either by not having known dread or by sinking under it.</i>	Eu posso dizer que aprender a conhecer a angústia é uma aventura com que cada homem tem de confrontar-se se não quiser cair na perdição ou por não conhecer a angústia ou por se afundar nela.
EBJT1(776):	<i>Some hope, he thought now, putting cheese between the bread slices, some hope, with costs going up all the time, Gareth on £13,000 plus his house, cows needing several hundred pounds ' worth of bought-in feed a year each, interest rates... and now this cold wet spring would mean the maize would go in late.</i>	Alguma esperança, pensava ele pondo queijo entre as fatias de pão, alguma esperança, com os preços a subirem constantemente, com Gareth a custar 13 000 libras mais a casa, com as vacas a precisarem todos os anos de rações em que despendia várias centenas de libras, as taxas de juro... e agora esta Primavera fria e chuvosa, o que significava que o trigo se iria atrasar .

5. Resultados das buscas corpus-based

Além da acima descrita busca aberta no COMPARA, foram realizadas, como já mencionado, buscas específicas de algumas colocações verbais de *have* e *go* no Cobuild e no Web Corp. Essas colocações (inclusive as do COMPARA) estão listadas nas **Tabelas 7 e 7A**, junto com o número de ocorrências obtidas para cada uma delas em cada corpus.

Nos casos das colocações menos comuns a busca visou confirmar que a expressão encontrada no COMPARA tratava-se mesmo de uma colocação, com o uso e conteúdo semântico nele (COMPARA) apontados sendo efetivamente empregados em outros exemplos.

No caso de colocações mais frequentes, o objetivo foi o de aprofundar o estudo da colocação em si e da natureza das colocações verbais em geral, procurando reconhecer exemplos em que o(s) significado(s) e as diversas formas de emprego fossem evidenciados. A seguir apresentamos dois exemplos mais signi-

Tabela 7

Busca Específica de colocações verbais de “Have” nos Corpora

Have a bath (132)
Have a drink (89+3)
Have a look (60+8)
Have a nice time (97)
Have a role (108)
Have an argument (133)
Have fun (42)
Have the heart (132+1)
Have a go(62)
Have a temperature (20)
Have a termination (118)
Have the curse (2+1)
Webcorp + COMPARA

TABELA 7 A

Busca Específica de colocações verbais de “Go” nos Corpora

Go amiss 10**
Go awry 9**
Go bald 183*
Go berserk 157*
Go in late 40*
Go mad 72*
Go off beat 17*
Go off the rails 106*
Go on the run 118* + 1**
Go out of their way 134*
Go overboard 14**
Go public 119* + 1**
Go to perdition 126*
Go to sea 244*
Go years back 34*
*Webcorp **Cobuild

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

ficativos com as observações de interesse para o tradutor que julgamos pertinentes para cada caso.

(1) http://www.download2me.com/jokes/joke_content/dart_board.htm

The Model Lodger

Doris and Fred had started their retirement years and decided to raise some extra cash by advertising for a lodger in their terrace house. After a few days, a young attractive woman applied for the room [...]

*“There’s just one problem, “explained the model. “Because of my job, I have to **have a bath** every night, and I notice you don’t **have a bath.**”*

Neste primeiro exemplo temos, na mesma sentença, duas ocorrências da expressão “have a bath”. Os conteúdos semânticos são, no entanto, totalmente distintos. O primeiro é o da colocação verbal “tomar banho” (que denominaremos doravante “sentido **colocativo**” correspondendo a “*collocate sense*” em inglês). No segundo, claramente, cada palavra retém um conteúdo semântico próprio e usual: “têm (= dispõem de uma) banheira”. É o que denominaremos “sentido **livre**” ou “**comum**” da expressão – correspondendo a seu “*plain sense*”, em inglês. É interessante notar que, se não estiver especificamente “em busca” de ocorrências da expressão, o leitor/tradutor (como o próprio autor da sentença) nem “repara” que está lendo/traduzindo (/escrevendo) duas vezes a mesma frase sintática.

(2) <http://www.mtholyoke.edu/~ebarnes/python/argument-clinic.htm>

Monty Python: Argument Clinic

*Man: Ab. I’d like to **have an argument**, please.*

Receptionist: Certainly sir. Have you been here before?

[...]

Man: Oh look, this isn’t an argument.

Mr Vibrating: Yes it is.

Man: No it isn’t. It’s just contradiction.

Mr Vibrating: No it isn’t.

Man: It is! [...]: An argument isn’t just contradiction

Mr Vibrating: It can be.

Man: No it can’t. An argument is a connected series of statements intended to establish a proposition. [...]

Aqui um *sketch* humorístico, a situação que tanta dificuldade traz para o tradutor. Como traduzir “*have an argument*”? “Comprar uma briga” é a tradução óbvia de início, mas o tradutor logo se deparará com o “intraduzível” ao chegar na “briga” sobre “*what is an argument*”. O importante neste caso é notar que, ao con-

trário do que acontece no exemplo anterior, a existência de dois sentidos para a mesma frase sintática – neste caso **concomitantes** na **única** instância da frase que surge no texto acima – não só se faz notar mas é inclusive ressaltada intencionalmente. O autor joga o sentido **colocativo** contra o sentido **livre** para criar a situação de humor básica para o *sketch*.

Exemplos poderiam suceder-se: “*go off beat*” claramente significando “sair do ritmo” num caso; claramente significando “mudar de assunto” em outro; ou ficando no “meio” dos dois, num site musical. “*Go out of their way*” significando “fazem de tudo”; ou então criando uma situação entre o **livre** “desviar do caminho” e o **colocativo** “fazer de tudo”, que soa de péssimo gosto num site que busca conscientizar os motoristas a respeitarem os ciclistas. A frase de um site de doação de órgãos: “*Do you have the heart to be an organ donor...*”, cuja tradução, é claro, teria que seguir o sentido **colocativo** e perder a alusão: “Você tem coragem de ser um doador de órgãos”. E que no entanto “clama” por uma solução de tradução (não encontrada), para que a alusão não se perca.

A lição aqui é a de que sentido **livre** e sentido **colocativo** são os dois extremos de um contínuo e não um “zero ou um”. Assim, o requisito para a correta tradução de uma colocação verbal muitas vezes vai além de saber-se que ela existe e conhecer o seu sentido **colocativo**.

Em certas situações, o autor joga com ambos os sentidos, **colocativo** e **livre**, e aí se encara o intraduzível. Noutras, é necessária uma análise para que se decida qual dos dois sentidos adotar, em qual “ponto do contínuo” se situa aquela determinada acepção. Essa decisão nem sempre é simples, como se vê no exemplo 7 do Apêndice 2 e no caso exposto em 5.1 abaixo, onde algo mais do que a mera tradução está em jogo.

Ainda no que respeita à “definição do ponto do contínuo”, foi possível – e interessante – reconhecer para o caso de pelo menos uma colocação (*have the heart*) a existência de algo como um “índice sintático da posição no contínuo”, que pode orientar o tradutor.

A **preposição** que se segue a “*have the heart*” parece indicar essa posição no contínuo de acepções. Seguida de “*to*” (ou quando situada no fim de uma sentença, caso em que o “*to*” é elíptico), a colocação claramente traduz-se por “coragem [a despeito de tudo]”, como nos exemplos 1 e 3 do Apêndice 2. Quando seguida de “*for*” como no ex. 4, a tradução ainda é “coragem” mas a coragem vem acompanhada de algo como “a predisposição, o talento” que abranda a “ousadia da coragem”. Esse é o caso de, por exemplo, “*I have the heart for the job*”.

Finalmente, quando seguida de *of* a acepção já se situa bem mais próxima do outro extremo do contínuo e a tradução será, normalmente, “ter o coração” (na acepção de “ter o espírito”, vide exemplo 2 do Apêndice 2).

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

5.1 Referência e tradução

Isso posto, a busca em corpus, ao que parece, pode às vezes render mais do que se espera. A seguir examinamos uma outra ocorrência de *have a heart* no WebCorp, que merece uma reflexão mais atenta.

Consideremos o seguinte texto, manchete e trecho de uma notícia do *The Independent*:

(3) <http://www.balkanpeace.org/hed/archive/july00/hed402.shtml>

The Independent (UK), July 30 2000

By Phil Rees in Podgorica

We have the heart for battle, says Montenegrin trained by SAS [...]

*The Special Police has a fierce reputation in Montenegro its gung-ho approach seemingly unsettling the SAS. "They thought we were crazy. When two of us banged into a house and started shooting into walls, bullets were flying around and they said 'Oh, it's a real gun, real bullets? You're crazy guys, you don't have protection'. But **we have a heart, we don't have protection but we have a heart. A big heart.**"*

Nesse texto jornalístico não há a princípio maior dificuldade para a tradução da colocação na manchete. Ela diz: "Temos coragem para a batalha", ou, numa tradução menos literal que talvez soe melhor em português, "Somos destemidos no combate". Nem há dificuldade na frase do combatente montenegrino que consta do corpo da notícia. Sua tradução seria algo como: "...temos coração, não temos proteção mas temos coração. Um grande coração".

A pergunta (de quem traduz para um redator que vai decidir a manchete em português) é se a manchete em inglês reflete de fato a frase do entrevistado, frase que o jornalista Phil Rees toma como o aspecto mais marcante da notícia a ponto de utilizá-la em sua manchete. E a utiliza na manchete como quem citasse *ipsis verbis* o entrevistado, o que, já se vê, não é o caso. Por um lado, o entrevistado não usa a palavra "*battle*" – mas, é claro, o contexto autoriza o jornalista a considerá-la em "elipse" na frase. Por outro lado, o entrevistado não diz "*we have the heart*", mas "*we have a heart*". Estaria ele se referindo (tão somente) à coragem ou a "algo maior"?

Posta em outros termos, a questão é a de se estabelecer se a informação que foi captada pela grande maioria⁷ de leitores do *The Independent* – aqueles que apenas passaram os olhos pela manchete e não leram o corpo da notícia – corresponde ou não à realidade do que expressou o entrevistado.

⁷ Vide Lonardoni (1999, p.112), citando o *Novo Manual de Redação* da Folha de São Paulo. Em seu artigo Lonardoni apresenta também um modelo de T. Van Dijk para a superestrutura do discurso da notícia na imprensa escrita. Esse modelo aborda, entre outros temas, o que de comum deve haver (ou o leitor espera que haja) entre

A resposta a essa questão requer uma digressão.

* * *

W. V. Quine (1908-2000), um dos mais importantes filósofos norte-americanos do pós-guerra, exerceu grande influência na corrente analítica do pensamento recente em áreas como a lógica, a filosofia da linguagem, a epistemologia e a metafísica.

Sua doutrina mais famosa, a da “indeterminação da tradução radical”, foi formulada no âmbito da epistemologia – e não no da teoria da tradução⁸ – como refutação a certas posições do empiricismo⁹.

Com essa doutrina Quine visava estabelecer o “naturalismo”, o ponto de vista de que a Filosofia deveria ser desenvolvida num “contínuo” com as Ciências sem qualquer status radicalmente distinto para a Matemática ou a Lógica, e sem o apelo a certas “verdades analíticas”, ou seja, verdades que são verificáveis em virtude do “significado” das respectivas proposições.

Não há “significados” que tenham tal poder ou tal “transcendência”, diz Quine, na medida em que não existem referentes que possam ser absolutamente determinados.

O que cabe aqui notar é que para apresentar seu argumento acerca da indeterminação dos **referentes** de uma forma “menos abstrata e mais realística” (QUINE: 1960, 27), Quine viu-se levado a lançar mão da **tradução** como **exemplo**. E a descrever uma situação hipotética de “tradução radical, i.e., a tradução

os conteúdos da manchete e do corpo da notícia (em especial o conteúdo da “lide”, o trecho inicial do corpo da notícia). Algo com implicações para a tradução de textos jornalísticos e mesmo para a própria ética do jornalismo.

⁸ Note-se, no entanto, que o argumento de Quine, basicamente um alerta contra a sinonímia, influenciou teóricos da tradução como I. A. Richards, Anton Popovic e mais recentemente Anne Mette Hjort. Dada a estatura de Quine, o nome adotado para a doutrina e a forma de apresentação de seu argumento, esse tipo de influência seria de se esperar. Veja-se Gentzler: 1993, pp. 15, 81, 197.

Cabe ainda notar que Quine lecionou no Brasil, na FFCL-USP, por três anos no início da década de 1940. Publicou em português um texto com conceitos que se tornaram seminiais no desenvolvimento da Lógica, *O Sentido da Nova Lógica* (Editora Martins, 1944).

⁹ Posições que valorizavam o significado das proposições como algo que define o tipo de experiências (científicas) que poderiam vir ou não a contar como evidência, contra ou a favor, do valor de verdade dessas proposições.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

da língua de um povo até então intocado” (id., 28) com o qual um lingüista pela primeira vez depara, no momento exato em que um coelho passa chispando na frente de um nativo, o qual, ato contínuo, grita “gavagai!”.

À moda dos filósofos, Quine passa boa parte de “Translation and Meaning”, o hoje famoso segundo capítulo de sua obra maior, *Word and Object*, a destruir qualquer esperança de que o lingüista possa concluir que a tradução de “gavagai” é “coelho” e não outra coisa qualquer como por exemplo “meros estágios, ou breves segmentos temporais, de coelhos” (ibid., 51). É impossível, argumenta Quine, que se decida algo de definitivo acerca de qual é, na língua do “povo intocado”, o referente de “gavagai”.

É preciso ressaltar novamente que Quine não está argumentando contra a possibilidade da tradução – que na realidade do dia-a-dia é “não-radical” e totalmente factível e corriqueira (ibid., 28) – mas contra uma certa, digamos, reificação ou “transcendência” do significado. Para isso, Quine busca demonstrar a indeterminação de **todo tipo** de referência. Nada contra a tradução em si.

Por outro lado, é significativo que uma situação extrema da tradução – ainda que idealizada – tenha sido a escolhida para a explicitação do argumento, e que isso o tenha marcado a ponto de ter-se adotado para a doutrina uma denominação algo imprecisa: “indeterminação da **tradução radical**”¹⁰. Melhor teria sido talvez – inclusive para o bem da Tradução – a denominação mais precisa “doutrina da indeterminação da **referência**”.

* * *

Pois bem, a matéria do *The Independent* que aflorou no WebCorp traz uma espécie de *gavagai* real. A dúvida do tradutor é “quineana”: não se trata (só) de encontrar a tradução, mas de se determinar qual é o referente, que **uso** o entrevistado estará fazendo da expressão “*have a heart*”.

- 1) O inglês do entrevistado será apenas sofrível, ele tem conhecimento de uma certa colocação com *have e heart* que significa “coragem”, e ao utilizá-la confundiu o artigo?
- 2) Ou “*a heart*” aqui é “mais que coragem”. É “um coração”, “uma razão maior”, “energia, entusiasmo, ardor”, dos quais a “coragem ousada” é apenas uma das conseqüências? (Note-se que *heart*, isoladamente, pode ter a

¹⁰ Ademais, o fato de ser mais freqüentemente adotada uma denominação abreviada, “indeterminação da tradução”, que omite tratar-se da “tradução radical” tal como definida por Quine, facilita a interpretação apressada e superficial de que “**toda** tradução é, a rigor, impossível”.

Crop, 10, 2004

acepção específica de “coragem”, como por exemplo em “*I plucked whatever heart I could and said...*” (NSOD), mas **não** quando precedido pelo artigo indefinido “**a**” como na frase do entrevistado. Por outro lado o “**for**” que se segue à colocação na manchete não ajuda. Como vimos acima ele “abrandou a coragem”, que no caso é, no mínimo, “ousada”).

No primeiro caso a manchete estaria “correta” (i.e., refletiria a realidade do que disse o entrevistado). No segundo caso, o *plain sense* (com *heart* num sentido figurado) terá sido tomado, equivocadamente pelo jornalista, pelo *collocate sense*.

E há ainda a possibilidade de tanto a acepção de “*have a heart*” para o entrevistado, quanto a de *have the heart* para o jornalista (não obstante o *for* que se segue à colocação), ficarem neste caso “no meio do caminho”, entre a “coragem” e a “razão maior”, entre o *collocate sense* e o *plain sense* (figurado) de “*have the heart*”.

Enfim, difícil decidir. E assim, em respeito a Quine, deixamos a questão indecisa. Já o tradutor no mundo real, premido é claro pelo *deadline*, terá tomado uma decisão e produzido a tradução.

Seja como for, eis aqui um exemplo de que os corpora como ferramenta não se prestam apenas ao estudo da língua em uso. Deles se pode extrair também aspectos referentes à epistemologia e às filosofias da ciência e da linguagem.

Mas ainda há mais com respeito aos corpora.

6. Carving Language

Ao estudar a estrutura e função das teorias científicas e de seus conceitos, Carl Hempel (1905-1997), um dos líderes do “empiricismo lógico” na filosofia da ciência, define como “relevância sistemática (*systematic import*) de um conceito” o fato de que ele se preste à formulação de princípios teóricos que reflitam as uniformidades do objeto de estudo (HEMPEL:1965, 146). Esse requisito, diz-nos Hempel, aplica-se também no caso das classificações científicas e está por trás da distinção que intuitivamente se faz entre **classificações “naturais” e “artificiais”**.

Por exemplo, os dois conjuntos de características sexuais primárias que determinam a divisão dos seres humanos entre homens e mulheres estão, cada um desses conjuntos, associados através de leis ou de correlações estatísticas a uma larga variedade de outras características de várias ordens e que lhes são concomitantes: características físicas, fisiológicas, psicológicas, etc. Isso contrasta com uma classificação que, por exemplo, dividisse os seres humanos em “até

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

1,70 m de estatura” e “acima de 1,70m” (classificação esta que pode até ser útil em casos específicos).

Nesse contexto Hempel recupera uma expressão que remonta a Platão e que se tornaria lapidar para designar os conceitos que fazem avançar as ciências. Diz ele do primeiro exemplo de classificação acima: “*It is understandable that a classification of this sort should be viewed as somehow having objective existence in nature, as ‘carving nature at its joints’*”.

Essa expressão vem à mente quando se reflete de forma mais ampla acerca do que ocorre na pesquisa em corpora. Foi fascinante constatar o quanto essa pesquisa é um “mergulho” na língua viva. Através dela uma série de conceitos afloram, são inferidos de maneira natural, automática. E a forma como esses conceitos se articulam transparece.

Pode-se de certa forma dizer que, ao utilizar os corpora para prospectar a língua, vivencia-se a realidade de alguns conceitos como “*somehow having objective existence in **language**, as ‘carving **language** at its joints’*”.

Por exemplo, a dimensão léxico-gramatical das palavras e o fato de ela constituir um contínuo, revelou-se através da grande efetividade do *collocate sampler* quando aplicado a *go* para buscar colocações, em comparação com sua baixa efetividade para *have*. E no fato de para *have* (mas não para *go*) terem sido necessários *strings* na forma de “filtros” para que a busca de colocações se tornasse manejável. Assim, essa gradação das palavras de “gramaticais” a “léxicas”, com toda uma gama intermediária de palavras que, como *have* e *go*, exercem ambas funções em diferentes graus, “salta aos olhos” via os corpora. Isso demonstra que os conceitos de “dimensão léxica” e “dimensão gramatical” têm *systematic import*, ou seja, indica que “*they carve language at a joint*”.

O exame de colocações específicas nos corpora revelou um outro contínuo, desta feita o da “acepção” da expressão, cujos extremos denominamos sentido “livre” ou “comum” (*plain sense*) e sentido “colocativo” (*collocate sense*). A necessidade de se avaliar o “ponto” exato em que se “está” nesse contínuo, e a possibilidade de que exista alusão de um *sense* ao outro constitui uma realidade que precisa ser levada em conta pelo tradutor. Ou seja, o fato de que os sentidos livre e colocativo não são um “zero ou um” é um outro *joint of language carved via corpora*. Essa constatação corresponde – para o caso específico das colocações que estudamos – ao que, em termos mais gerais, Baker (1992, pp. 205-206) denomina *instantial meaning* de uma palavra ou expressão num texto, que é específico de cada texto, e que o próprio ambiente textual determina.

Nesse sentido, um aspecto que discutimos para o caso de *have the heart* na seção 5 acima é interessante. Como se recorda foi possível utilizar a preposição que se segue a essa colocação – *to, for* ou *of* – como uma espécie de “índice da

posição daquela ocorrência no contínuo de acepções da colocação”. Isso significa que, graças aos corpora, identificou-se de uma forma bastante objetiva “um tijolo” – por mínimo que seja – da construção dos vários *instantial meanings* de “*have the heart*”. Tem-se aí um pouco de *instantial meaning* “*in the making*”. Um pouco da semântica emergindo e se **articulando**, a partir da sintaxe.

Por outro lado, os corpora podem indicar que certas classificações tradicionais nas gramáticas não têm base na forma como a língua se articula. Isso parece ser o caso da distinção entre colocações verbais e não verbais, colocações verbais e *phrasal verbs*, e mesmo entre colocações, expressões fixas e expressões idiomáticas.

Aparentemente trata-se sempre do mesmo fenômeno básico: a co-ocorrência freqüente de determinadas palavras na língua em detrimento de outras que raramente ou nunca co-ocorrem, embora tenham o mesmo conteúdo semântico. E mais a aquisição de um “significado” por parte dessas expressões consagradas, que pode ser desde próximo até totalmente distinto da “soma dos significados das palavras” que as compõem.

Não foi possível reconhecer no *output* que os corpora produziram – nem na natureza das dificuldades de tradução ligadas a essas expressões – qualquer tipo de diferenciação que justificasse a distinção em pauta. Nossa busca *corpus driven* gerou indistintamente soluções com colocações verbais ou com *phrasal verbs*, expressões idiomáticas e expressões fixas (vide Apêndice 1).

Isso não significa que essa distinção não seja justificada e útil em certas aplicações. Porém ela não se origina, ao que tudo indica, em algo intrínseco à língua. *It does not carve language at a joint.*

7. Considerações Finais

Assim, os corpora emergem do presente trabalho como uma versátil e multifacetada ferramenta para a pesquisa de colocações verbais, especialmente para o tradutor, tolhido pela rigidez e escassez da lexicografia a respeito do assunto.

As diversas abordagens que adotamos permitiram a execução de uma série de tarefas: a busca e localização de colocações no corpus paralelo, obtendo assim uma ou várias opções de tradução; a confirmação de que se está de fato diante de uma colocação verbal através da constatação de seu uso, freqüência e “identidade semântica” nos corpora; a confirmação da acepção corrente de uma colocação; a constatação de uma segunda acepção até então desconhecida; a busca a partir de indicações dos *collocate samplers*, revelando colocações insuspeitadas, incluindo palavras que nos eram até aqui desconhecidas; e um caso de “índice

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

sintático”: a preposição que se segue à colocação *have the heart* orientando o tradutor na definição da acepção exata.

Ademais, a busca *corpus driven* revelou a possibilidade de construir-se um “dicionário vivo e permanentemente atualizável”, que incluía todas as colocações de um determinado verbo de alta frequência na língua constantes no COMPARA. Até mesmo as de um verbo auxiliar, como *have*. E isso graças ao fato de que a sintaxe de busca do COMPARA demonstrou um surpreendente poder de expressão semântica, que permitiu que se restringisse as soluções a um *have* bastante “despido de sua dimensão gramatical”. O poder expressivo dessa sintaxe de busca merece ser mais bem estudado pelo tradutor. Ficou-nos mesmo a impressão de se estar diante de um potencial ainda não devidamente explorado.

Enfim, o uso de corpora na pesquisa das colocações verbais com vistas à tradução tornou viável e fascinante o que talvez sem eles fosse quase inviável ou extremamente árduo.

Mas não só para a execução desse tipo de tarefas serviram os corpora. Foi possível surpreender neles – ou surpreender-se com – situações que instanciam aspectos de interesse da epistemologia, da filosofia da linguagem, da filosofia da ciência. E, mais ainda, usá-los como uma espécie de “campo de prova” da relevância de conceitos empregados no estudo da língua. Como ferramentas através das quais conceitos que são intrínsecos à natureza da linguagem afloram, “*carving language at its joints*”, revelando a forma como a língua se articula.

To carve tem duas acepções. Por um lado, a de “esculpir, cinzelar, gravar, entalhar”. Por outro, a de “trinchar”: “cortar com certa elegância, em especial a carne que se servirá à mesa”. É essa a acepção que aqui se aplica. Quem “*carves at the joints*”, “trincha nas juntas”. Mas quem trincha – para servir à mesa – me informa que **sempre** se trincha nas juntas (quando há juntas). Portanto, a tradução da expressão que emprestamos de Hempel é, simplesmente, “trinchar a língua”.

E assim os corpora se revelam “ferramentas de trinchar a língua”. Trinchar para traduzir. Trinchar para destrinchar.

Agradecimentos

A Diana Santos e ao pessoal do COMPARA, pelo suporte crucial no momento exato.

A Beatriz Souza, Marcia Fiker e Stela Foley, pela acolhida carinhosa e pelas discussões tão profícuas.

A Stella Tagnin, pelo entusiasmo e pelo apoio.

Crop, 10, 2004

Bibliografia

- ALSTON, W. P. "Quine on Meaning", in Hahn, L.P. e Schilipp, P. A. eds., *The Philosophy of W.V. Quine*. La Salle, Open Court. 1988.
- BAKER, M. *In Other Words*. New York, Routledge. 1992.
- GENTZLER, E. *Contemporary Translation Theories*. New York, Routledge. 1993.
- HEMPEL, C. G. "Fundamentals of Taxonomy", in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York, The Free Press. 1968.
- HUDDLESTON, R. D. & PULLUM, G. K., eds., *The Cambridge Grammar of the English Language*. Cambridge, Cambridge University Press. 2002.
- LONARDONI, M. "Aconteceu virou manchete – Um estudo dos vetores de manchetes jornalísticas", in VASCONCELOS, S. I. *Os discursos jornalísticos: reportagem, manchete, classificados & artigo*. Itajaí: Edit. da Univali / Maringá: Eudem, 1999.
- MANSER, M. H., ed., *Good Word Guide*. Londres, Bloomsbury. 1990.
- QUINE, W. V. "Translation and Meaning", segundo capítulo de *Word and Object*. Cambridge, The MIT Press. 1960.
- SWAN, M. *Practical English Usage*. Oxford, Oxford University Press. 2002.
- TAGNIN, Stella O. *Expressões Idiomáticas e Convencionais*. São Paulo, Ed. Ática. 1989.

Dicionários:

- DICTIONARY of English Phrasal Verbs and Their Idioms*. London, Collins. 1990.
- NOVO DICIONÁRIO da Língua Portuguesa*. Coord. Aurélio Buarque de Holanda Ferreira, Rio de Janeiro, 1ª. edição, Editora Nova Fronteira, s/d.
- OXFORD Collocations Dictionary for Students of English*. Oxford, Oxford University Press. 2002.
- THE CONCISE Oxford Dictionary*. Oxford, Clarendon Press. 1990.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

APÊNDICE 1

COLOCAÇÕES “NÃO VERBAIS” DE “HAVE” E “GO” OBTIDAS NO COMPARA (CONSIDERADAS DE INTERESSE PARA O TRADUTOR)

COLOCAÇÕES DE “HAVE”

EBDLIT 1(1520):	<i>She had it off with young Ritchie at the New Year's Eve.»</i>	Foi para a cama com o Ritchie mais novo na noite de fim de ano.
EBDLIT 2(1520):	<i>She had it off with young Ritchie at the New Year's Eve.»</i>	Ela transou com o garoto Ritchie na festa de Ano Novo.
EBDLIT 1(332):	<i>Doesn't waste time poking around in your unconscious, asking you about potty training, or whether you saw your parents having it off together, that sort of thing.»</i>	Não perde tempo a vasculhar o inconsciente, a fazer perguntas sobre a indiferença com que fomos criados, ou se alguma vez vimos os nossos pais na cama e esse tipo de coisas.»
EBDLIT 2(2092):	<i>He has better grounds for complaint than me (should that be «I»?) - indeed, he could probably have me for assault.</i>	Ele tinha mais razões para reclamar do que eu (ou deveria dizer «para reclamar de mim-?) – na verdade, poderia me mandar prender por agressão à sua pessoa.
EBDLIT1 (1764):	<i>«To be frank, you and Jake have us over a barrel on this one.»</i>	– Para falar com toda a franqueza, estamos à mercê de ti e do Jake nesta questão.
EBDLIT2 (1764):	<i>«To be frank, you and Jake have us over a barrel on this one.»</i>	– Para ser honesto, você e Jake nos têm pelo dedo mindinho nessa aqui.
EBDLIT 2(146):	<i>«He's only having you on,» said the nurse.</i>	Ele está tirando sarro de você – disse a enfermeira.

COLOCAÇÕES DE “GO”

EBDLIT 1(748):	<i>No go.</i>	Nem pensar.
EBIT1 (222):	<i>He was the only one of the Merediths to go in for cattle, too.</i>	Também era o único dos Merediths que criava gado .
ESNG1 (577):	<i>The subdued monosyllable was pronounced with such certainty; the habit of each other had made them even less demonstrative than they had been at the beginning of their marriage, but he was moved to go over to her.</i>	O monossílabo em voz baixa foi pronunciado com tal certeza; o hábito de cada um fizera-os menos afeitos a demonstrações que no começo de seu casamento, mas ele sentiu vontade de abraçá-la .
EBDLIT 1(600):	<i>If she were to go with anyone else, something new in her behaviour, some unfamiliar adjustment of her limbs, some variation in her caresses, would tell me, I'm certain.</i>	Se ela andasse com alguém , tenho a certeza de que haveria algo de diferente no seu comportamento, uma inovação qualquer na sua forma de ajustar os membros, uma variação nas suas carícias, que me levaria a descobrir.
EBDLIT 2(600):	<i>If she were to go with anyone else, something new in her behaviour, some unfamiliar adjustment of her limbs, some variation in her caresses, would tell me, I'm certain.</i>	Se ela tivesse dado para outro , alguma coisa nova no seu comportamento, alguma disposição nova de seus braços ou pernas, alguma variação no seu jeito de fazer carinho me diriam, disso estou certo.

APÊNDICE 2

Busca específica de colocações de “*have*” e “*go*” nos corpora: alguns exemplos adicionais

Adicionalmente aos exemplos constantes do corpo do artigo, apresentamos abaixo algumas das ocorrências obtidas nas buscas específicas por colocações nos corpora. Alguns dos casos, inclusive, não correspondem a colocações verbais, mas são apresentados por ilustrarem alguns pontos ressaltados no texto principal, ou por seu interesse para o trabalho de tradução. Cada grupo de exemplos é seguido de um breve comentário.

A) Exemplos 1, 2, 3, e 4

Exemplo 1

<http://www.patternlanguage.com/apl/aplsample/apl76/apl76.htm>

*Many small households, not large enough to have a full fledged nursery, not rich enough to have a nanny, find themselves swamped by the children. The children naturally want to be where the adults are; their parents don't **have the heart**, or the energy, to keep them out of special areas; so finally the whole house has the character of a children's room – children's clothes, drawings, boots and shoes, tricycles, toy trucks, and disarray*

Exemplo 2

http://1stholistic.com/Reading/liv_inspiration-the-heart-of-the-beginner.htm

*The Bible's very clear about it. While God does want us to grow in knowledge (Hosea 4:6), His desire is that we constantly **have the heart** of the Beginner. The heart of one who loves Him with all their heart, mind, strength and soul (Lk. 10:27). Oblivious to human praise. “The eyes of the Lord roam over the whole earth, to encourage those who are devoted to him wholeheartedly.” (2 Chr. 16:9)*

Exemplo 3

<http://bnewheart.homestead.com/>

*Do you **have the heart** to be an organ donor?*

As of July 30, 2002 in the US there is a total of 80,076 persons waiting for organs.

Broken down as follows: 52,686 waiting for a kidney, 1,313 waiting for a pancreas,...

Exemplo 4

http://www.charlotte.com/mld/charlotte/sports/columnists/ron_green_sr/3709833.htm

Posted on Mon, Jul. 22, 2002 (Charlotte Observer)

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

COMMENTARY

French **have the heart**, not the execution, for dares

RON GREEN SR.

Ah, those Frenchmen. They are so, how you say, so ... French.

When Jean Van de Velde came to the last hole at Carnoustie in 1999, he needed only to make a double-bogey to become the first Frenchman since 1907 to win the British Open. Logic dictated he hit something like a 5-iron off the tee, then lay up short of the water that fronts the green with a wedge, hit another wedge onto the green and leave himself three putts to win.

O exemplo 1 é claramente de um *collocate sense*. O exemplo 2 talvez não seja totalmente *plain*: “to have the heart of the Beginner” parece requerer algo de coragem (“strength”), mas a tradução é, certamente, “ter o coração...”. No exemplo 3, ao qual nos referimos de passagem no corpo do trabalho, há a alusão entre os *senses*, irrecuperável na tradução. O exemplo 4 é bem interessante: *you have one thing required “for dares” (“the heart”), but you miss the other (“the execution” = a técnica, o talento)*. Portanto, a estrutura sintática da frase puxa para o *plain* (ter “uma coisa” mas não ter “outra”), mas “heart” precisa ser traduzido como “coragem”. Um caso, diríamos, em que o “*in-between*” se situa não só no nível semântico, mas também “entre os níveis” sintático e semântico. Vide no corpo do artigo (seção 5) nossas observações sobre as diferentes preposições – *to, for* ou *of* – que se seguem a esta colocação servirem como um “índice da acepção”.

B) Exemplos 5, 6, 7 e 8

Exemplo 5

<http://www.fnal.gov/pub/inquiring/questions/danny.html>

Questions About Physics

How does an electron on a 2p orbital cross the node of nucleus?

To Fermilab,

*I have a question that needs answering for my chemistry class. In an atom, how does an electron on a 2p (bell shaped) orbital cross the node of the nucleus (a region of zero probability)? Does it **have something to do with** the wave properties of an electron? I cannot find the answer on the internet or in my chemistry book.*

Thank you very much.

Danny

Exemplo 6

<http://www.med.umich.edu/1libr/subabuse/tobacc12.htm>

Tobacco

Crop, 10, 2004

Cigarette Substitutes – What to Take Along

Just as you pack for any trip, you can also pack for your journey to freedom from smoking! Remember all the reasons you used to reach for cigarettes? These will be the reasons you reach for your cigarette substitutes.

*To **have something to do with** your hands.*

To have something to do while waiting in traffic jams or movie lines or for slow service at a restaurant.

As a “prop” to relax.

As a reward.

Exemplo 7

http://www.gerismith.com/music/music_chain_01.html

Something to Do with Laurie

Laurie and I drove into the city last night

Mostly we talked about the loves in our lives

*And I don't know why but **I have something to do with Laurie***

We have shared the same secrets, we have loved the same men

Our paths keep crossing again and again

*I don't understand why **I have something to do with Laurie***

This chain wraps around

And we can't break these links as hard as we try

This train knows where we're bound

*So I guess we should **JUST SIT BACK AND ENJOY THE RIDE***

Laurie and I cannot seem to forget

Our dances with partners that don't know the steps

*And I don't know what yet, but I **HAVE SOMETHING TO LEARN** from Laurie*

Cause I woke up this morning and everything had changed

My vantage points had been rearranged

*And I know it sounds strange, but it **bad***

Something to do with Laurie

Exemplo 8

<http://www.qwerty.com/hkr.htm>

*To me, the most challenging thing I can think of is coming up with software that is really interesting and helpful. And that's a lot more difficult than just breaking into a system and causing a mess. Although, to be completely honest, I did **have something to do with** a long-ago Christmas day when all the*

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

*computers in a very large international network suddenly started playing **Jingle Bells** in unison on their printers. But that's another story...*

Have something to do with não é colocação verbal, porém optamos por fazer constar estes exemplos neste apêndice já que neste caso, é claro, também existe um *plain* e um *collocate sense* da expressão, que muitas vezes é um problema para o tradutor em termos da decisão sobre qual dos dois *senses* ocorre num caso específico. O exemplo 5 é claramente *collocate* (“ter algo a ver com”, “ter alguma relação”), o texto técnico usa a expressão para indagar sobre a possibilidade de uma relação causal. Igualmente o exemplo 8 que confessa ter havido “alguma” responsabilidade, com uma ironia que indica que a responsabilidade foi “total”. A ocorrência no exemplo 6 é claramente *plain* (“ter algo para fazer com”), mas permite-nos inferir que “*smoking has something to do with* (*collocate*) *having something to do with* (*plain*) *one's hands*”. No caso do exemplo 7, como traduzir o título da música? A ocorrência da expressão na última estrofe é claramente *collocate* dado o *it* que a precede. Mas e as duas ocorrências da primeira estrofe como ficam? Um empático “ter a ver” ou algo mais físico “a fazer”? Dado o “sentido geral” da letra, intencionalmente (?) dúbio (há ou não uma alusão? vide as duas frases em caixa negrito), ficam essas duas ocorrências, como o título, “indeterminadas”. Cada leitura é uma leitura, e aqui prevalecerá a leitura do tradutor.

Idéias que Cruzam o Oceano¹

*Belinda Maia e Luís Sarmento**
*Stella E. O. Tagnin e Sandra Maria Aluísio***

Resumo: *O objetivo deste artigo é apresentar projetos similares e, por vezes, complementares implementados ou por implementar nos dois lados do Oceano Atlântico: na Universidade do Porto, em Portugal e na Universidade de São Paulo, no Brasil. Para tanto, subdivide-se em duas partes: na primeira, os autores portugueses (Maia e Sarmento) detalham e discutem ferramentas desenvolvidas para um público de tradutores e terminólogos. O Corpógrafo é um conjunto de ferramentas para a construção e análise de corpora e bases de dados terminológicos, enquanto o TrAva tem por objetivo avaliar tradutores automáticos. No início da segunda parte, Aluísio descreve o portal Lácio-Web, sua construção, os corpús de que se compõe, bem como as ferramentas disponíveis para sua exploração e montagem de sub-corpús de pesquisa. Na parte final, Tagnin descreve o projeto COMET, sua construção, seus vários subcorpús (Cortec, Corpús de Traduções e Corpús de Aprendizagem) e os tipos de pesquisa a que se destinam.*

Palavras-chave: *corpús; tradução; terminologia; ferramentas eletrônicas; Corpógrafo; TrAva; Lácio-Web; COMET.*

¹ Este trabalho foi originalmente apresentado no Workshop Luso-Brasileiro no Congresso da European Society for Translation Studies (EST), realizado em Lisboa, de 26 a 29 de setembro de 2004.

* Universidade do Porto.

** Universidade de São Paulo – USP.

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

Abstract: *This article presents similar and sometimes complementary projects developed and under development across the Atlantic: at the Porto University in Portugal and at the University of São Paulo, in Brazil. It is divided into two parts: in the first one the Portuguese authors (Maia and Sarmento) discuss the computational tools that have been developed for translators and terminologists: the Corpógrafo is a set of tools for building and analyzing corpora and terminological databases, while TrAva aims at evaluating automatic translation software. The second part starts with Aluísio's description of the Lácio-Web portal: its construction, the corpora that constitute it as well as the tools available for investigating and creating personal sub-corpora for research. In the last part Tagnin describes the COMET Project, its construction, its various subcorpora and the types of research these are meant for.*

Keywords: *corpus; translation; terminology; computational tools; Corpógrafo; TrAva, Lácio-Web; COMET.*

Parte 1 – Universidade do Porto, Portugal

0 Introdução

Num congresso que favorece o estudo de tradução do ponto de vista académico/literário/cultural, é possível que a proposta de falar em pesquisa académica apoiada em ferramentas não seja muito popular. No entanto, o estudo das línguas em geral, e da tradução em particular, têm muito a ganhar com o tipo de observação possibilitado pela digitalização de qualquer tipo de texto e pelo respectivo exame cuidadoso com as ferramentas produzidas pela lingüística computacional e pelo processamento da linguagem natural.

Para quem desconhece essa área de pesquisa, lembramos que é costume chamar a um conjunto de textos digitalizados um ‘corpus’, sendo o plural ‘corpora’. Em certos casos, é possível enriquecer um determinado *corpus* com informação adicional recorrendo a etiquetadores que inserem informação no texto. Essa informação é normalmente de natureza lingüística, sendo também possível incluir etiquetas morfossintácticas, utilizando ferramentas como o etiquetador “Palavras” de Eckhard Bick – ver <http://visl.hum.sdu.dk/visl/pt/> – utilizado pelo projecto Linguateca, ao qual pertence o PóloCLUP, da Universidade do Porto. O tipo de etiquetagem efectuada depende do objectivo do projecto sendo possível a criação de um novo esquema de etiquetagem, conforme a utilização prevista para o *corpus*.

No entanto, enquanto que os etiquetadores morfossintácticos já funcionam no nível automático ou semi-automático, qualquer etiquetagem de nível semântico ou textual terá ainda de ser feita manualmente.

Uma vez construído o *corpus*, etiquetado ou em 'cru', há ferramentas para observar e analisar o uso da língua nesse *corpus*. Essas ferramentas podem oferecer informação de natureza estatística, que poderá posteriormente ser analisada para objectivos específicos. A ferramenta mais conhecida, entretanto, é a que permite pesquisar concordâncias, i.e., a observação de fenómenos lingüísticos no nível da frase ou do contexto. Essa pesquisa pode ser feita com uma ou várias palavras, ou, no caso de um *corpus* etiquetado, utilizando as etiquetas, e permite ao pesquisador confirmar os seus palpites com exemplos concretos e quantificáveis.

Os grandes corpora monolingües oferecem a possibilidade de estudar a língua no nível lexical e sintáctico. O British National Corpus (BNC), que é composto por uma grande variedade de textos, escritos ou transcritos de textos orais, contendo 100 milhões de palavras, todas etiquetadas no nível de 'parts-of-speech', está disponível para uso simples (até 50 concordâncias) em linha via <http://sara.natcorp.ox.ac.uk/lookup.html> ao público em geral. A Linguateca – <http://www.linguateca.pt> – dispõe de corpora monolingües de mais de 250 milhões de palavras com quase 4 milhões anotados, livremente acessível em linha.

Há corpora bi- e multilingües que servem para o estudo comparativo das línguas em todos os níveis, desde o léxico ao texto. Esses corpora podem ser paralelos – os originais e as suas traduções – alinhados normalmente no nível da frase, ou comparáveis – textos originais em duas ou mais línguas, considerados representativos de determinados géneros, registos ou domínios. A Linguateca oferece um *corpus* paralelo em português e inglês de textos literários com aproximadamente 1 milhão de palavras em cada língua.

1. Corpora e Tradução

Actualmente o uso de corpora electrónicos para a pesquisa e ensino de tradução vai muito além de uma simples consulta lexicográfica. As ferramentas disponíveis, aliadas a uma preparação mais sofisticada do utilizador, permitem a análise sintáctica do texto e do discurso. É evidente que a informática obriga o tradutor a observar uma certa linearidade na tradução, o que vai contra as teorias que advogam uma interpretação do texto original adequado ao contexto do texto de chegada. Isto, em si, é um assunto para investigação. Mas também é pertinente considerar se essas teorias mais 'abrangentes' de facto se aplicam mais ao texto literário ou ao texto não literário. A criação de corpora comparáveis permite a observação de textos originais em duas ou mais línguas, dando uma

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

maior possibilidade de estudar as diferenças entre as convenções textuais em todos os níveis lingüísticos e culturais, e torna-se evidente que as diferenças existem em todo o tipo de texto (Maia, 2003a).

O uso comercial de corpora paralelos, como memórias de tradução, criou um novo paradigma nos estudos de tradução. No nível comercial é preciso melhorar, uniformizar e re-utilizar textos de uso diário. É costume falar de manuais de instruções nesse contexto, porque é óbvio que esses textos precisam de ser bem formulados, seguindo regras estudadas e formuladas pela disciplina de 'Escrita Técnica', ou 'Technical Writing', que as traduções têm de obedecer a níveis de qualidade comparáveis, e que é de interesse económico aproveitar tanto os textos como as traduções já feitas quando se tratam de ligeiras modificações ou aumentos dos textos para máquinas mais desenvolvidas. Entretanto, as memórias de tradução têm grande utilidade para a cuidadosa preparação de textos e traduções em qualquer área que imponha um alto nível de exigência estilística e terminológica. Também têm implicações para a tradução automática

Enquanto a lingüística computacional e a lingüística de corpora evoluíram bastante durante os anos 90, profissionais ligados a essas áreas, mas com responsabilidades na educação de tradutores, começaram a compreender o alcance da aplicação dessas tecnologias ao ensino da lingüística contrastiva e da tradução em geral. Com o advento da Internet e o acesso a uma grande quantidade de informação em todos os domínios, também surgiram novas possibilidades para melhorar o ensino da tradução especializada. As actas dos congressos de TALC – Teaching Applications and Language Corpora (1994, 1996, 1998, 2000, 2002 & 2004), PALC – Practical Applications of Language Corpora (1997, 1999, 2001, & 2003), e CULT – Corpus Use and Learning to Translate (1997, 2000, & 2004) demonstram vários aspectos dessas tendências.

É natural que o mundo da língua portuguesa tenha sido influenciado por essas correntes. Este trabalho vai apresentar e discutir projectos complementares em andamento nos dois lados do oceano, ou seja, nas Universidades do Porto (Portugal) e de São Paulo (Brasil). São projectos que se inserem nos Estudos de Corpora e têm por objectivo desenvolver corpora para diversos fins, assim como ferramentas de exploração desse material.

2. Corpora e tradução – ensino e pesquisa no PoloCLUP e LINGUATECA

Há vários anos, o projecto Linguateca se preocupa com a função de corpora no ensino e pesquisa de lingüística e tradução. Além de uma vasta selecção de textos em português e um forte leque de ferramentas de pesquisa lingüística, o seu *corpus* paralelo COMPARA está disponível em linha. Esse *corpus* foi iniciado em 1999 e, devido essencialmente aos esforços das duas autoras, continua a ser o maior

corpus paralelo revisto, com textos em português e inglês e as suas traduções, e com um elevado número de utilizadores em escala mundial (Frankenberg-Garcia & Santos, 2002 & 2003). É disponibilizado através do DISPARA (Santos, 2002), que oferece capacidades inovadoras de procura e tem uma interface rigorosamente paralela nas duas línguas.

Na Universidade do Porto, o interesse em corpora, lingüística contrastiva e terminologia já resultou em teses de doutoramento e mestrado e artigos de vária ordem desde 1994. Antes de pertencer ao projecto Linguateca, já tínhamos experimentado trabalhar com corpora paralelos e comparáveis e, especialmente, com '*do-it-yourself corpora*' (Maia, 1997) ou '*disposable corpora*' (Varantola, 2002). Quando a Internet começou a oferecer a possibilidade de adquirir informação sobre uma grande variedade de domínios especializados e, ao mesmo tempo, facilitou a utilização de textos em formato digital, os já interessados em corpora não tardaram em compreender as implicações pedagógicas dessas fontes de informação. Rapidamente desenvolveu-se uma metodologia para estudar os textos normalmente associados com essas áreas para o estudo de todos os níveis lingüísticos, desde o lexema ao discurso. O nosso interesse em domínios especializados levou a um estudo mais aprofundado da terminologia (Maia, 2003: vários) e à criação de um Mestrado em Terminologia e Tradução na Faculdade de Letras da Universidade do Porto (FLUP), que funciona desde 2000.

3. O Corpógrafo

Em outubro de 2002, o PoloCLUP da Linguateca foi inaugurado e, entre outras coisas, tem desenvolvido pesquisa no uso de corpora especializados comparáveis para o estudo e a extração de terminologia. Criámos, para esse efeito, o Corpógrafo, um conjunto de ferramentas disponível 'online' para quem estiver interessado em pesquisar autonomamente. O Corpógrafo permite coleccionar textos em vários formatos, formar e analisar corpora, extrair terminologia e criar bases de dados terminológicas com vários campos, inclusive com a possibilidade de criar relações semânticas e ontologias.

O Corpógrafo (Sarmiento & Maia, 2003; Sarmiento, Maia & Santos, 2004) é uma plataforma de pesquisa sobre corpora que surgiu da necessidade de integrar no mesmo ambiente todo um conjunto de operações e de processos que normalmente são realizados utilizando várias ferramentas ou sistemas, cujo acesso é muitas vezes restrito ou difícil. O Corpógrafo oferece ao utilizador, através de uma simples interface Web, a possibilidade de compilar e pesquisar os seus próprios corpora (a partir de documentos em formato PDF, Ms-Word, PostScript, RTF ou HTML) sem que para isso seja necessário ter conhecimentos especiais

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

de informática. De certa forma, o Corpógrafo complementa a oferta de corpora genéricos já oferecidos pela Linguateca (projectos AC/DC, CETEMPúblico e CETENFolha), possibilitando a construção e pesquisa em corpora pessoais e específicos, por utilizadores com interesses na área da Lingüística, Tradução ou Engenharia do Conhecimento.

Do ponto de vista mais específico dos estudos na área da Lingüística, o Corpógrafo possibilita pesquisas de concordância e colocações, e a realização de estudos de frequências de combinações de duas ou mais palavras sobre os corpora pessoais. Para tarefas associadas à Tradução e à Engenharia do Conhecimento, o Corpógrafo possui também funcionalidades de pesquisa terminológica. A pesquisa terminológica encontra-se directamente integrada num sistema de base de dados para uma fácil organização dos termos extraídos. As capacidades de pesquisa terminológica (fundamentalmente em português e inglês, mas também em espanhol, francês, italiano e alemão) são complementadas com funcionalidades que permitem a identificação de definições dos termos extraídos e de possíveis relações semânticas (actualmente meronímia e hiponímia) entre os conceitos. Actualmente o Corpógrafo foi experimentado por cerca de 200 pessoas e é utilizado regularmente por 40, localizadas maioritariamente em Portugal e no Brasil, sendo também utilizado por investigadores dispersos por vários países da Europa.

Na FLUP, há várias teses de doutoramento e mestrado e projectos de terminologia em curso utilizando o Corpógrafo para pesquisar áreas de Engenharia Mecânica, Engenharia Electrónica, Geografia da População, Geografia – Riscos Naturais, Genética, Neuroanatomia, GPS – Geographical Positioning System e outros. Podemos prever que, com a entrada numa nova fase de funcionamento do Corpógrafo e a continuação do seu desenvolvimento com a pesquisa em curso, poderemos dar um salto qualitativo e quantitativo no nível de trabalho a ser efectuado.

4. Avaliação de Tradução Automática – ensino e pesquisa

Uma outra experiência no âmbito do PoloCLUP e do Mestrado em Terminologia e Tradução surgiu com a avaliação de tradução automática (TA). Essa experiência começou mais no nível pedagógico, embora a tecnologia criada para esse efeito tenha vindo a ser a aproveitada por uma grande variedade de utilizadores.

A primeira ferramenta criada foi o METRA, que permite pedir traduções Português > Inglês e Inglês > Português a sete motores de tradução automática livremente disponíveis na web. A segunda ferramenta foi o BOOMERANG, que permite que cada tradução automática seja re-introduzida e re-traduzida automa-

ticamente até chegar a um ponto de paragem. Embora o resultado tenha um aspecto quase cómico, a idéia subjacente é teoricamente interessante. A terceira ferramenta chamou-se EVAL e resultou no TrAva, uma ferramenta que permite a observação e análise de vários motores de TA online. Um dos resultados é um *corpus* de traduções + traduções automáticas EN > PT, o CorTA. Essas ferramentas estão disponíveis em <http://www.linguateca.pt>.

O TrAva é uma ferramenta construída essencialmente para a criação de material de teste e foi desenvolvida como proposta inicial de avaliação conjunta para a tradução automática (TA). O TrAva permite traduzir frases do inglês para o português em quatro motores de TA disponíveis na Internet e apresenta um quadro de classificação de critérios lingüísticos utilizando dois sistemas gramaticais: o sistema de etiquetagem gramatical utilizado pelo British National Corpus (BNC) para a classificação das frases da língua de partida (inglês) e uma taxonomia baseada na sintaxe do português para a classificação dos resultados na língua de chegada. Os resultados da avaliação de traduções desenvolvidos com o TrAva são consultáveis através do *corpus* CorTA, Corpus de Traduções automáticas Avaliadas (Santos et al., 2004).

O TrAva é um serviço web que tem como objectivo permitir a recolha cooperativa de exemplos de tradução automática (i.e. frases originais e as respectivas traduções automáticas) de forma a criar uma colecção de casos que permita uma melhor compreensão dos problemas envolvidos na TA da língua portuguesa. O modo de funcionamento do TrAva é simples e próximo do serviço ao qual sucede, o EVAL (<http://poloclup.linguateca.pt/ferramentas/eval/>). O TrAva envia uma frase em inglês fornecida pelo utilizador a quatro serviços de tradução automática web para que esses a traduzam para português. Os serviços empregues são:

- FreeTranslation (<http://www.freetranslation.com>)
- Systran (<http://www.systranbox.com/systran/box>)
- E-Translation Server (http://www.linguatex.net/online/ptwebtext/index_en.shtml)
- Amikai (<http://www.amikai.com/demo.jsp>)

O TrAva apresenta posteriormente ao utilizador as traduções automáticas obtidas a partir desses quatro serviços para que sejam então o alvo de um processo de classificação. A classificação é feita através de um formulário próprio onde o utilizador é convidado a introduzir informação relativa aos problemas de tradução ou tipo de erros que ocorrem nas traduções e, simultaneamente, informação adicional sobre a frase original. Toda esta informação é guardada em base de dados, que é imediatamente consultável por todos os utilizadores.

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

O TrAva permite aos utilizadores pesquisar a informação da base de dados que é construída colectivamente. Os critérios de pesquisa disponíveis são todos aqueles que foram usados para a classificação das traduções e frases originais (nota: a identidade dos participantes é reservada pelo que não é apresentada nas consultas da base de dados, apesar de ficar armazenada no TrAva). Espera-se que essa compilação colectiva de exemplos de tradução automática, que se encontra acessível publicamente, possa ajudar a melhorar a compreensão dos fenómenos que estão envolvidos na tradução automática do português. Para isso convidamos todos os interessados a visitarem <http://www.linguateca.pt/trava/> e a participarem na recolha, após a inscrição no TrAva.

5. Comentário final

Com esta breve apresentação das várias ferramentas, esperamos ter despertado um interesse em utilizá-las. O seu interesse justifica os nossos esforços e qualquer “feedback” que queiram nos mandar será apreciado, porque acreditamos que a pesquisa só pode avançar se houver colaboração e troca de informação e ideais entre pesquisadores.

Parte 2 – Universidade de São Paulo, Brasil

6. O Projeto Lácio-Web

O Lácio-Web (LW) (Aluísio et al 2003, Aluísio et al 2004) é um projeto realizado com parceria entre o Núcleo Interinstitucional de Lingüística Computacional (NILC)², localizado no ICMC-USP, o IME-USP e a FFLCH-USP, sob a coordenação de Sandra Maria Aluísio. O objetivo desse projeto é divulgar e disponibilizar gratuitamente na Web: a) vários *corpuses*³ do português brasileiro escrito contemporâneo, representando bancos de textos adequadamente compilados, catalogados e codificados, em um padrão que possibilita fácil intercâmbio, navegação e análise; e b) ferramentas lingüístico-computacionais, tais como contadores de frequência, concordanciadores e etiquetadores morfosintáticos treinados em grandes *corpuses* anotados manualmente.

² www.nilc.icmc.usp.br/

³ Nesta parte do artigo, optamos por utilizar a forma aportuguesada *corpuses*, tanto para o singular, quanto para o plural, (a exemplo de *lápiz*) para as palavras latinas *corpus* e *corpora*, respectivamente.

O público-alvo do LW é heterogêneo: de um lado lingüistas, cientistas da computação, lexicógrafos, terminólogos etc. e, de outro, não especialistas em geral. O LW é acessado a partir de um portal (<http://www.nilc.icmc.usp.br/lacioweb/>), que informa os tipos de córpus, ferramentas, todo o material disponível e forma de contribuir com textos para a continuação do projeto, disponibiliza manuais e artigos relacionados e permite, após cadastramento do usuário, o acesso a seus córpus e ferramentas. O portal pretendeu ser didático ao público leigo em geral e estudantes em particular e, para isso, foram incluídos textos explicativos e de ajuda.

Dada a importância de um recurso de base como são os córpus de uma dada língua, para avançar estudos lingüísticos variados e também para a construção de sistemas computacionais de processamento de língua natural (PLN), justifica-se o sucesso que tivemos em conseguir permissão oficial para incluir materiais diversos, durante os 30 meses do projeto. Porém, para obter essa permissão, foi incluído, juntamente com o termo de autorização, um texto explicativo apontando o potencial dos recursos e a necessidade de obtenção de textos integrais para diversas pesquisas lingüísticas, por exemplo, a análise de discurso e tarefas como a tradução.

A primeira disponibilização pública do projeto ocorreu em 20/01/2004 e a segunda no final do projeto, em 30/06/2004. Parte do material adquirido ainda precisa passar por um processo de três fases para ser disponibilizada: a) compilação-formatação dos textos que vieram da Web e de CD-ROMs, b) nomeação sistemática dos arquivos e c) criação de um cabeçalho para os textos com diversas informações, a saber: bibliográficas comuns – título, autoria, local e data de publicação, fonte, editor, língua; de catalogação – tamanho do arquivo, tipo de amostragem, tipo de autoria, sexo do(s) autor(es); e da tipologia textual em quatro categorias: domínios, gêneros, tipos de texto e meios de distribuição, que serão apresentadas na Seção 6.2. Essa subcategorização detalhada é que permite ao usuário do LW a criação de subcórpus de estudo que atendam a suas pesquisas específicas.

6.1 A Constituição do LW

O Lácio-Web tenta preencher uma lacuna em termos de recursos para pesquisa e suporte à criação de ferramentas de PLN para a língua portuguesa do Brasil. Para tanto, é formado por seis córpus: Lácio-Ref, Mac-Morpho, Lácio-Dev, Par-C, Comp-C e Lácio-Sint, descritos abaixo:

- 1) **Lácio-Ref:** córpus aberto e de referência composto de textos escritos em português brasileiro, respeitando a norma culta, com 4278 arquivos,

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

totalizando 8.291.818 ocorrências. É um *corpus cru* (não anotado com informações morfossintáticas, sintáticas ou de nível mais elevado), mas possui anotações da existência de elementos gráficos e anotação de cabeçalho. A grande maioria dos textos está disponibilizada na íntegra.

- 2) **Mac-Morpho**: *corpus* fechado e anotado morfossintaticamente, formado por artigos jornalísticos retirados da Folha de São Paulo, ano 1994, dos cadernos Esporte (ES), Dinheiro (DI), Ciência (FC), Agronomia (AG), Informática (IF), Ilustrada (IL), Mais! (MA), Mundo (MU), Brasil (BR) e Cotidiano (CO). Composto de 1.167.183 ocorrências, o *corpus* foi etiquetado pelo parser Palavras de Eckhard Bick (<http://visl.hum.sdu.dk/>), revisado manualmente quanto à anotação morfossintática e serviu de treinamento de três etiquetadores morfossintáticos disponíveis na Web (Aluísio, Pelizzoni et al, 2003). O MAC-MORPHO é disponibilizado para *download* em 2 formatos: 1) adequado para pesquisas lingüísticas com o uso de contadores de frequência ou concordanciadores, por exemplo; 2) adequado ao treinamento de etiquetadores e que, por ter os polilexicais separados⁴, teve o tamanho do *corpus* alterado para 1.221.468 ocorrências.
- 3) **Lácio-Dev**: *corpus* projetado para ser um *corpus* aberto, isto é, de atualização contínua, e com textos que não foram revisados em relação à norma culta. Esse *corpus* destina-se à avaliação do desempenho de corretores gramaticais do português brasileiro, isto é, pretende servir de *benchmark* para a tarefa de correção gramatical, bem como para análise de inadequações lingüísticas nos textos.
- 4) **Par-C**: *corpus* aberto, paralelo, Português-Inglês, que possui, inicialmente, textos de 1 ano de edições da Revista Pesquisa Fapesp, num total de 646 textos em cada língua. O número total de ocorrências desse *corpus* é de 893.283.
- 5) **Comp-C**: *corpus* aberto, formado por textos originais de conteúdo comparável em inglês e português, inicialmente disponível apenas para o gênero jurídico. Conta com 29 textos, 61.149 ocorrências e será ampliado futuramente. Os *corpus* comparáveis são projetados para a avaliação de métodos de extração de termos para sistemas de PLN, para

⁴ “Rio=de=Janeiro_NPROP”, por exemplo, é separado em “Rio_NPROP de_NPROP Janeiro_NPROP”, em que NPROP é uma etiqueta para nomes próprios.

confeção de glossários e dicionários especializados e outras pesquisas lingüísticas.

- 6) **Lácio-Sint** (porção etiquetada do cópús Lácio-Ref): cópús fechado e etiquetado automaticamente com lemas, etiquetas morfossintáticas e sintáticas. Diferente do Mac-Morpho, esse cópús será composto por textos de diversos gêneros e contará com ferramentas desenvolvidas no NILC, tais como o *parser* Curupira (Martins et al, 2002).

No total, o Projeto LW possui 5.708 arquivos, totalizando 10.413.524 ocorrências. Os cópús Lácio-Dev e Lácio-Sint serão disponibilizados futuramente, como fruto de pesquisas de doutorado e mestrado, respectivamente.

6.2 Tipologia Textual do LW

O LW distingue seus textos em quatro categorias ortogonais: gênero, tipo de texto, domínio e meio de distribuição. A definição e a composição das categorias são detalhadas abaixo.

Gênero textual: o gênero discrimina o texto pela intenção comunicativa e pelo caráter discursivo, isto é, a comunidade (meio) em que circula e as atividades humanas que o tornam relevante. Convencionamos o uso de um supergênero, chamado Literário (LT), um conjunto de gêneros e um conjunto de subgêneros. Os gêneros e subgêneros são dados abaixo:

Gênero	Subgêneros
Científico (CI)	----
De referência (RE)	enciclopédico, lexicográfico, terminológico e outros.
Informativo (IF)	jornalístico e outros
Jurídico (JU)	----
Prosa (PR)*	biografia, conto, novela, romance e outros
Poesia (PO)*	----
Drama (DR)*	----
Instrucional (IS)	didático, procedimental e outros
Técnico-Administrativo (TA)	----

* Esses gêneros, especialmente, advêm do supergênero Literário.

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

Tipo textual: considera-se “tipo de texto” o modo específico de estruturação de um texto. Refere-se ao texto visto “de dentro”, ou seja, suas partes componentes, seu léxico, sua sintaxe, sua adequação ao tema etc. Trata-se de uma lista em constante atualização e que, no momento, é composta de 39 categorias (e “Outros” – tipos textuais não previstos), por ex.: apostila, manual, parecer, reportagem, súmula, testamento etc.

Domínio: é a “área de conhecimento” que tematiza a principal informação veiculada pelo texto. Temos 3 grandes linhas de domínio, denominadas “domínio geral”. A cada uma dessas linhas associam-se subdomínios, denominados “domínios específicos”. A divisão em termos de domínio geral apresenta as seguintes vertentes:

a) científica: classifica os textos tematizados pela ciência. Esse grupo é composto por seis áreas do conhecimento: Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas e da Terra, Ciências Humanas e Ciências Sociais Aplicadas;

b) **religião e pensamento:** envolve os temas metafísicos, espirituais e teológicos (ex.: livros de bruxaria, de auto-ajuda, etc.).

c) **generalidades:** absorve os textos com temas variados e, de modo geral, inseridos num campo tematizado pelo senso comum (ex.: entretenimento). Inclui, além disso, os textos que abordam, de forma não-analítica, temas considerados pela ciência (ex.: ciência e tecnologia, saúde, esporte, etc.).

Meio de distribuição: seleciona o canal por meio do qual o texto foi divulgado ao seu público-alvo, por ex.: Cd-rom (CR), Diário Oficial (DO), Internet (IN), Jornal (JO), Livro (LI), Tese (TE).

6.3 As Ferramentas do LW

O Projeto Lácio-Web disponibiliza várias ferramentas lingüístico-computacionais como concordanciadores, contadores de frequência e etiquetadores morfossintáticos, treinados com o cópulo do português do Brasil anotado manualmente – o MAC-Morpho – e futuramente pretende disponibilizar extratores de termos e alinhadores de textos paralelos. O objetivo é facilitar o estudo de características lingüísticas do português por pesquisadores da área de lingüística e lingüística computacional, assim como melhorar a qualidade dos sistemas desenvolvidos para o português, tais como, tradutores para o português, sumarizadores automáticos e engenhos de busca especializados no português.

As ferramentas podem ser usadas com o Lácio-Ref, com os subcópus criados pelo usuário ou ainda com o cópulo que o usuário tiver carregado para o

Crop, 10, 2004

LW. Existem três tipos de pesquisa para montagem de subcorpú: (1) a pesquisa simples que faz a busca de textos baseada em meio de distribuição e gênero; (2) a avançada que considera, primeiramente, meio de distribuição, supergênero e gênero textual. Considera também subgênero para os gêneros Informativo e Prosa e outros dados particulares de cada gênero textual, por exemplo: para o gênero Informativo, o nome do periódico e seção/caderno; para o Científico, o nome do autor e para o supergênero Literário, o nome do autor e da obra; e (3) a pesquisa personalizada que oferece grande parte dos dados de cabeçalho das amostras do Lácio-Ref. Por isso, é possível montar um subcorpú refinado em termos de detalhes da bibliografia e da classificação textual.

6.3.1 Contador de Frequência Padrão

Calcula a frequência de todas as palavras do corpú escolhido, informando o número total de arquivos, de ocorrências, de palavras únicas, e a variação do vocabulário por meio do cálculo da razão *type/token*. Apresenta também a tabela de frequência das palavras. A saída desse contador pode ser salva para leitura posterior.

6.3.2 Contador de Frequência por Palavra

Dada uma ordem decrescente de frequência das palavras do corpú escolhido, o Contador calcula a frequência de uma palavra escolhida, apresentando a frequência das palavras do contexto anterior e posterior a ela. A saída desse contador pode ser salva para leitura posterior.

6.3.3 Concordanciador para corpú sem anotação

O concordanciador implementado gera uma lista enumerada de todas as ocorrências de uma determinada palavra (ou expressão) escolhida de um corpú sem anotação. As opções para corpú no nosso projeto são, atualmente, o subcorpú de pesquisa gerado a partir do Lácio-Ref, o próprio Lácio-Ref ou o corpú do usuário, desde que sem anotação morfossintática.

6.3.4 Concordanciador para corpú anotado morfossintaticamente

O concordanciador implementado gera uma lista enumerada de todas as ocorrências a partir da seleção de uma palavra com a respectiva etiqueta. A opção para corpú no nosso projeto é, atualmente, o próprio MAC-Morpho e

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

as etiquetas são as que constam do Manual de Etiquetação do MAC-Morpho, disponível para *download* no portal LW.

6.3.5 Etiquetadores Morfossintáticos

Três etiquetadores disponíveis na Web – MXPOST (Ratnaparkhi, 1996), TreeTagger (Schmid, 1994) e o etiquetador de BRILL (Brill, 1995) – foram treinados com o córpus MAC-MORPHO, que possui 1.221.468 palavras. Esse córpus foi separado em uma parte para treinamento (977.161 palavras), que traz 80% de cada um dos 10 cadernos da *Folha de São Paulo* que compõem o córpus, e uma parte para teste (244.307 palavras) com os 20% restantes de cada caderno. A precisão dos etiquetadores por caderno pode ser vista no portal LW.

7. O Projeto COMET

Face à enorme possibilidade de pesquisas contrastivas no âmbito do ensino de línguas estrangeiras e da tradução e, principalmente, ao fato de não haver córpus bilíngües de áreas técnicas que envolvam o português brasileiro, está sendo construído o COMET, um Córpus Multilíngüe para Ensino e Tradução.

O COMET (www.fflch.usp.br/dlm/comet) é constituído basicamente de três subcórpus: um Córpus Técnico-Científico, um Córpus de Aprendizes e um Córpus de Traduções.

7.1 O Córpus Técnico-Científico (CORTEC)

O CORTEC constitui um córpus técnico-científico de âmbito geral, mas privilegia quatro grandes áreas, determinadas a partir de questionário que submetemos a diversos tradutores profissionais, via Internet, indagando das áreas mais carentes de material de apoio. As respostas apontaram para: Direito Comercial, Informática, Ortodontia e Meio Ambiente.

A outra parte do córpus está sendo construída com todos os córpus compilados pelos alunos do Curso de Especialização em Tradução (CETRAD) – Inglês e da Pós-Graduação da FFLCH-USP, resultantes de projetos diversos (Tagnin 2003⁵), alguns dos quais serviram para a elaboração de glossários destinados a tradutores (disponíveis em <http://www.fflch.usp.br/citrat>). Numa pri-

⁵ Observe-se aqui o paralelismo com as pesquisas realizadas por Maia, relatadas na outra parte deste artigo.

Crop, 10, 2004

meira etapa (2001), foram construídos *córpus* com aproximadamente 100.000 palavras em cada língua nas seguintes áreas:

- Biotecnologia: alimentos transgênicos
- Culinária: receitas
- Computação: segurança na Internet
- Moda: roupas
- Veterinária: doenças dos bovinos
- Ecologia: biodiversidade
- Odontologia: ortodontia
- Automação industrial: sensores
- Negócios: mercado financeiro
- Turismo: ecoturismo
- Engenharia genética: genoma

A esses, foram acrescentados, em 2003, 14 outros, totalizando atualmente 5.463.597 palavras em inglês e 2.928.940 em português.

Em sua grande maioria, os textos constituem *córpus* comparáveis, ou seja, são de tamanho semelhante (em geral, ao redor de 100.000 palavras), compostos de textos dentro de uma mesma área, seguindo padrões semelhantes, tais como: gênero, tipologia textual, extensão, fonte, data etc., em inglês e português. Essa composição permite observar o uso natural da linguagem, fornecendo ao tradutor subsídios para produzir uma tradução fluente e natural. Permite também avaliar a equivalência de significado e de uso de um termo ou palavra pela análise do seu contexto de ocorrência e, assim, produzir traduções naturais e glossários bilíngües confiáveis. Ao pesquisador, permite estudos sobre a fraseologia da área, aspecto praticamente ignorado na grande maioria dos glossários, que em geral se atêm a termos simples – recorremos aqui à Culinária, a título de exemplo – (*pimenta, caçarola*) e compostos (*pimenta-do-reino, pimenta calabresa, panela de pressão*), não registrando unidades de significado mais extensas como *pimenta-do-reino moída na bora*, ou colocações verbais como *untar uma forma, cortar (uma cebola) em rodelas*.

7.1.1 Problemas na construção dos *córpus*

Ao contrário do que se pode imaginar – que basta “baixar” da Internet textos de um determinado assunto para já se obter um *córpus* – a construção de um *córpus*, para que atenda os objetivos a que se propõe, deve seguir rigorosos critérios de compilação (Atkins et al 1992). O desconhecimento desses critérios,

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

em especial na primeira etapa do projeto, acarretou uma série de problemas, desde a delimitação da área de estudo e o balanceamento do cópús até a obtenção da permissão de inclusão dos textos, passando pela padronização dos textos e inserção do cabeçalho em cada um deles.

7.1.2 A composição atual do CORTEC

O material já coletado foi redistribuído da seguinte forma, dentro das quatro áreas prioritárias:

- **Informática**
 - Segurança na Internet
 - Impressoras
- **Ortodontia**
- **Direito Comercial**
 - Legislação americana/brasileira (Laranjinha 1999)
- **Meio Ambiente**
 - Ecologia: Biodiversidade
 - Turismo: Ecoturismo

Como há diversos cópús da área de Medicina, os seguintes serão agrupados dentro desse domínio:

- Dermatologia
- Insuficiência Cardíaca
- Nefrologia
- Hipertensão Arterial (Castanho 2003)

Por se tratar de um cópús técnico-científico, é essencial que seja atualizado constantemente, o que implica tanto o acréscimo de textos recentes, para garantir a atualidade da terminologia, quanto a criação de novas áreas ou renomeação de outras, como é o caso do Ecoturismo, hoje denominado Turismo Sustentável.

7.1.3 Um cópús paralelo – Revista Pesquisa FAPESP

Dentro do Cópús Técnico-Científico está também sendo construído um cópús paralelo (originais com respectivas traduções) com os textos eletrônicos da Revista Pesquisa da FAPESP, a partir da edição de número 60, do ano 2000, que foram gentilmente cedidos por aquela instituição. A revista é composta de diversas seções e cobre áreas como Política Científica e Tecnológica, Ciência, Tecnologia, e Humanidades. O cópús está sendo alinhado no nível da sentença

Crop, 10, 2004

por um alinhador sentencial desenvolvido no NILC (Caseli e Nunes, 2003). É importante salientar que nesse *cópus* os textos originais são em português e os traduzidos, em inglês. Além das pesquisas contrastivas léxico-gramaticais de praxe, a diversidade de tipologia textual da revista (reportagens, cartas, notícias, carta do editor, artigos) permitirá estudos também no nível do discurso. Uma parte desse material já está disponível no Par-C do LW.

7.3 O *Cópus* de Traduções

Esse *cópus* é constituído de a) um conjunto de textos paralelos (originais e respectivas traduções) de literatura estrangeira traduzida para o português brasileiro, que consiste de nove contos americanos e vinte contos canadenses traduzidos por alunos do CETRAD. Os contos canadenses foram publicados em 2002, sob o título *Lá do Canadá* (Tagnin 2002). O *cópus* está sendo aumentado com aproximadamente 25 contos australianos e suas respectivas traduções; b) um conjunto de textos paralelos de literatura brasileira vertida para idiomas estrangeiros. Esse tipo de *cópus* permite analisar processos e estratégias de tradução, bem como enseja toda sorte de estudos contrastivos, desde morfológicos, sintáticos e lexicais até textuais.

7.4 *Corpus* de Aprendizes

Esse *cópus* assemelha-se, em termos de objetivos, ao Lácio-Dev do Lácio-Web, mas é constituído de redações de aprendizes de línguas estrangeiras (alemão, espanhol, francês, inglês e italiano) dos cursos de graduação e extensão da FFCLH/USP e destina-se a pesquisas sobre os problemas mais recorrentes de aprendizagem das respectivas línguas. Permitirá pesquisas tanto horizontais, por exemplo, dificuldades de uma determinada classe ou de aprendizes de certo nível, quanto verticais, ou seja, pesquisas diacrônicas que acompanham o desenvolvimento de um grupo ou indivíduo ao longo de um período de tempo.

Também incluirá um *cópus* de aprendizes de tradução, onde serão recolhidas as primeiras versões de trabalhos, sem qualquer correção, dos alunos do Curso de Especialização em Tradução do inglês. Com ele pretende-se determinar as dificuldades enfrentadas pelos aprendizes, seja no que concerne a estratégias de tradução de determinados itens, tais como aqueles que indicam aspectos culturais ou dialetais, sejam problemas específicos do próprio vernáculo.

Ambos têm por objetivo último o aperfeiçoamento do ensino nas respectivas áreas.

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

7.5 Caracterização do COMET

7.5.1 Os objetivos

Face ao acima exposto, tornam-se claros os objetivos desse cópua no campo da tradução e do ensino. Em primeiro lugar, o COMET pretende ser uma fonte de linguagem natural atualizada em diversas áreas técnico-científicas. Com a configuração descrita, pretende compensar a falta de material lexicográfico e terminológico nas áreas contempladas. Acima de tudo, no entanto, pretende ser fonte representativa para a pesquisa, a prática e o ensino da tradução e das línguas estrangeiras, como já vem ocorrendo.

7.5.2 O público-alvo

O público a que se destina abrange desde aprendizes e professores de tradução e das línguas estrangeiras contempladas, tradutores nas duas direções (inglês e português), até lexicógrafos e terminólogos, além de quaisquer pesquisadores interessados nos vários aspectos lingüísticos desses idiomas, inclusive na análise do discurso.

7.5.3 Os textos

Em virtude dessa abrangência, os textos que compõem o COMET inserem-se, na sua grande maioria, em três gêneros: acadêmico, jornalístico e comercial, além dos textos dos aprendizes.

Os textos acadêmicos são aqueles escritos por especialistas para especialistas. Caracterizam-se por apresentarem a linguagem natural empregada por esses profissionais, ou seja, apresentam o termo em seu contexto natural, inclusive com suas colocações e coligações. Esse aspecto é essencial para o tradutor que, com freqüência, tem dúvidas quanto às palavras (verbos, adjetivos) que ocorrem com o termo em questão.

Os textos jornalísticos nas áreas técnico-científicas são, em geral, escritos por especialistas para um público leigo. Por essa razão, apresentam muitas vezes uma definição dos termos técnicos, aspecto de especial interesse para o terminólogo. O tradutor, porém, também se beneficia desse tipo de texto, pois o contexto de ocorrência pode assegurar-lhe a equivalência (ou não) de um termo sobre o qual esteja em dúvida.

Finalmente, os textos comerciais (folhetos, manuais, anúncios etc.), escritos por especialistas ou não-especialistas para um público leigo, são de grande valia pela alta concentração de termos técnicos e, muitas vezes, pelas ilustrações

Crop, 10, 2004

que os acompanham, o que contribui para esclarecer o significado de termos obscuros.

Os textos são inseridos na íntegra, não só para assegurar a possibilidade de análise textual, como também para servirem de fonte de referência para o estudo do assunto tratado. O público que mais se beneficia desse aspecto são os aprendizes de tradução, que podem, dessa forma, familiarizar-se com o assunto em que estão trabalhando. É fato que, quanto maior o conhecimento de uma área, mais apto estará o aprendiz para produzir uma tradução confiável.

Em suma, o COMET é um *cópus* multilíngüe destinado ao ensino e à pesquisa de línguas e de tradução. Sua configuração facilita seu uso para a resolução de questões práticas, tais como determinar o uso correto de certo termo, ou a palavra que usualmente co-ocorre com outra, assim como se presta para uma gama extremamente variada de estudos acadêmicos. No âmbito de nossa Universidade, está sendo usado para trabalhos sobre lexicologia, terminologia, construção de *cópus*, processos e estratégias de tradução, dificuldades dos aprendizes e análises contrastivas. Outros trabalhos envolvem a construção de *cópus* próprios que, ao término, serão incorporados ao COMET. Dessa forma, estabelece-se uma proveitosa troca acadêmica: o COMET alimenta diversos estudos acadêmicos, enquanto *cópus* resultantes de outros estudos retroalimentam o COMET. Assim, garante-se o enriquecimento e a constante atualização do *cópus*.

8. Os pontos de contato

Pelo acima exposto, podemos observar que os construtores de *cópus* técnicos do Projeto COMET, por exemplo, podem se beneficiar do Corpógrafo para a extração de termos e compilação de glossários. O Projeto COMET, parceiro do NILC na construção do LW, contribuiu e continuará contribuindo com os textos para os quais obtiver autorização de inclusão, tanto em inglês, quanto em português, para aumentar, principalmente, o *cópus* Par-C. O Lácio-Ref, por ser o único *cópus* de português brasileiro sistematicamente classificado em gênero, tipo de texto, domínio e meio de distribuição, que está gratuitamente disponível na Web, complementa recursos semelhantes existentes para a variante europeia, permitindo pesquisas contrastivas, em especial nas áreas de especialidade. O objetivo de ambas as partes é disponibilizar gratuitamente na Web recursos computacionais e lingüísticos que possam contribuir para o desenvolvimento e enriquecimento das pesquisas nos dois lados do oceano.

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

Referências Bibliográficas

- ALUÍSIO, S. M.; PELIZZONI, J. M.; MARCHI, A. R.; OLIVEIRA, L. H.; MANENTI, R.; MARQUIVAFÁVEL, V. (2003). An account of the challenge of tagging a reference corpus of Brazilian Portuguese. In: *PROPOR'2003, 2003, Faro. Lecture Notes on Artificial Intelligence. Proceedings of PROPOR'2003*. Springer Verlag, 2003. v. 1.
- ALUÍSIO, S., PINHEIRO, G.M., MANFRIM, A.M.P, OLIVEIRA, L.H.M. de, GENOVES Jr., L.C., TAGNIN, S.E.O. The Lácio-Web: Corpora and Tools to Advance Brazilian Portuguese Language Investigations and Computational Linguistic Tools. In: *LREC 2004. Proceedings of LREC, 2004, Lisboa, Portugal*, p. 1779-1782.
- ALUÍSIO, S.M., PINHEIRO, G., FINGER, M., NUNES, M.G.V., TAGNIN, S.E.O. The Lácio-Web Project: overview and issues in Brazilian Portuguese corpus creation. In: *CORPUS LINGUISTICS 2003, 2003, Lancaster, UK. Proceedings of Corpus Linguistics 2003 (Also as UCREL Technical Report, Vol 16 Part)*. Lancaster: 2003. v. 16, p. 14-21.
- ATKINS, S., CLEAR, J. & OSTLER, N. Corpus Design Criteria, *Literary and Linguistic Computing*, vol. 7, n. 1, 1992, 1-16.
- BRILL, E. Transformation-based error-driven learning of natural language: A case study in part of speech tagging, *Computational Linguistics 21* (1995), 543-565. Disponível em: <http://www.cs.jhu.edu/~brill/>.
- CASELI, H.M.; NUNES, M.G.V. Sentence Alignment of Brazilian Portuguese and English Parallel Texts. In: *Argentine Symposium on Artificial Intelligence (ASAI 2003)*. Buenos Aires, Argentine, September 2003.
- CASTANHO, R.M.C. *Proposta para a elaboração de um glossário de colocações na área médica – subárea hipertensão arterial*, dissertação de mestrado, Universidade de São Paulo, 2003.
- FRANKENBERG-GARCIA, Ana & Diana SANTOS. 2003 COMPARA, um corpus paralelo de português e inglês na Web. In: Stella TAGNIN (Org.). *Cadernos de Tradução*, n. 9 – 2002/1, Núcleo de Tradução – NUT, Universidade de Santa Catarina, 61-79.
- FRANKENBERG-GARCIA, Ana & Diana SANTOS. 2003. Introducing COMPARA, the Portuguese-English parallel translation corpus. In: Federico Zanettin, Silvia Bernardini & Dominic Stewart (Eds.). *Corpora in Translation Education*, Manchester: St. Jerome Pub. 71-87.
- LARANJINHA, A. L. T. *Para um Glossário Bilingüe Português-Inglês de Termos do Direito Comercial: Colocações Verbais*, dissertação de mestrado, Universidade de São Paulo, 1999.
- MAIA, Belinda 2003a Training Translators in Terminology and Information Retrieval using Comparable and Parallel Corpora. In: F. Zanettin, S. Bernardini & D. Stewart *Corpora in Translator Education*, Manchester: St. Jerome Pub. , 43-54.
- MAIA, Belinda. 2003b. Ontology, Ontologies, General Language and Specialised Languages. In: *Volume Comemorativo dos 25 anos do CLUP*. Porto: CLUP. 23-39.

Crop, 10, 2004

- MAIA, Belinda 2003c. What are comparable corpora? In: *Proceedings of pre-conference workshop Multilingual Corpora: Linguistic Requirements and Technical perspectives at Corpus Linguistics 2003*, Lancaster U.K., 27-34.
- MAIA, Belinda. 2003d. Ensinar como especializar-se. In: *Actas do V Seminário de Tradução Científica e Técnica em Língua Portuguesa*, Lisboa: União Latina.
- MAIA, Belinda 2003e. Using Corpora for Terminology Extraction: Pedagogical and Computational Approaches. In: B. Lewandowska-Tomasczyk, (Ed.). *PALC 2001 – Practical Applications of Language Corpora.*, Lodz Studies in Language. Frankfurt: Peter Lang, 147-164.
- MAIA, Belinda. 2003f. Terminology – where to find it, and how to keep it. (Keynote Speaker). In: *Proceedings of III Jornadas sobre la formación del traductor e intérprete*, Universidad Europea de Madrid. CD-ROM.
- MAIA, Belinda. 2003g. Do-it-yourself, disposable, specialized mini corpora – where next? Reflections on teaching translation and terminology through corpora. In: Stella Tagnin (Org.). *Cadernos de Tradução*, n. 9 – 2002/1 – Núcleo de Tradução – NUT, Universidade Federal de Santa Catarina, 221-235.
- MAIA, Belinda. 2002a. Nothing is inherently boring – reflections on training translators in terminology. In: Maia, B. J. Haller & M. Ulrych (Eds.). 2002, *Training the Language Services Provider for the New Millennium. Proceedings of Encontros III de Tradução da AsTra-FLUP 25-26 Maio de 2001*, 355-64.
- MAIA, Belinda. 2002b. The Industrialisation of Translation – will it work? In: *Génesis–Revista Científica do ISAI*, n. 2. Porto: ISAI., 17-26.
- MAIA, Belinda. 2002c. Corpora for terminology extraction – the differing perspectives and objectives of researchers, teachers and language services providers. In: *The Proceedings of the Workshop Language Resources for Translation Work and Research – held in conjunction with LREC 2002*, Universidad de Las Canárias, Spain, 25-28.
- MAIA, Belinda, 2000. Making corpora – a learning process. In: Bernardini, S. & Zanettin, F. (Eds.). 2000: *I corpora nella didattica della traduzione*. Bologna: CLUEB, 47-56.
- MAIA, Belinda 1997. Do-it-yourself corpora... with a little bit of help from your friends! In Barbara Lewandowska-Tomasczyk and Patrick James Melia (Eds.). *PALC '97 Practical Applications in Language Corpora*. Lodz: Lodz University Press, 403-410.
- MAIA, Belinda, & Luís SARMENTO. 2003a. The Pedagogical and Linguistic Research Applications of the GC to Parallel and Comparable Corpora. In: *Proceedings of CP3A 2003: Corpora Paralelos, Aplicações e Algoritmos Associados*. Braga: Universidade do Minho.
- MAIA, Belinda and Luís SARMENTO. 2003b. Constructing comparable and parallel corpora for terminology extraction – work in progress. Poster apresentado em Corpus Linguistics 2003, Lancaster U.K. (Vencedor do 1o. prêmio).
- MARTINS, R. T.; HASEGAWA, R.; NUNES, M.G.V. Curupira: um parser funcional para o português, *NILC-TR-02-26*, dezembro 2002.
- RATNAPARKHI, A. A Maximum Entropy Part-of-Speech Tagger, *Proceedings of the First Empirical Methods in Natural Language Processing Conference* (1996).

MAIA, Belinda; SARMENTO, Luís e TAGNIN, Stella E. O. e ALUÍSIO, Sandra Maria. *Idéias que cruzam o oceano*.

SANTOS, Diana, Belinda MAIA & Luís SARMENTO. Gathering empirical data to evaluate MT from English to Portuguese. In: Lambros Kranias, Nicoletta Calzolari, Gregor Thurmair, Yorick Wilks, Eduard Hovy, Gudrun Magnusdottir, Anna Samiotou & Khalid Choukri (Eds.). *Proceedings of LREC 2004 (Workshop on the Amazing Utility of Parallel and Comparable Corpora)* (Lisboa, Portugal, 25 May 2004), 14-17.

SANTOS, Diana. DISPARA, a system for distributing parallel corpora on the Web. In: Nuno Mamede & Elisabete Ranchhod (Eds.). *Portugal for Natural Language Processing (PorTAL 2002)* (Faro, Portugal, 23-26 June 2002), Berlin/Heidelberg: Springer-Verlag. *Lecture Notes in Artificial Intelligence*, 209-218.

SARMENTO, Luís & Belinda MAIA. Gestor de corpora – Um ambiente Web integrado para Lingüística baseada em Corpora. In: José João Almeida (Ed.). *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, Junho), Braga: Universidade do Minho, 25-30.

SARMENTO, Luís, Belinda MAIA & Diana SANTOS. The Corpógrafo – a Web-based environment for corpora research. In: Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (Eds.). *Proceedings of LREC 2004* (Lisboa, Portugal, 26-28 May 2004), 449-452.

SCHMID, H. Probabilistic part-of-speech tagging using decision trees, *Proceedings of International Conference on New Methods in Language Processing* (1994), 44-49.

TAGNIN, S. E.O. 2003. Os Corpora: instrumentos de auto-ajuda para o Tradutor. In: Stella Tagnin (Org.). *Cadernos de Tradução*, n. 9 – 2002/1, número especial sobre Corpus e Tradução, Universidade Federal de Santa Catarina, Florianópolis: Núcleo de Tradução, 191-219.

TAGNIN, S.E.O. (Org.). 2002. *Lá do Canadá – contos*. São Paulo: Olavobrás.

VARANTOLA, Krista. 2003. Disposable Corpora as Intelligent Tools in Translation. In: Stella Tagnin (Org.). *Cadernos de Tradução*, n. 9 – 2002/1 – Núcleo de Tradução – NUT, Universidade Federal de Santa Catarina, 171-189.

Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais

Marcia Epstein Fiker*
Stela Foley**

Resumo: *O artigo apresenta parte do trabalho final desenvolvido durante uma das matérias do Curso de Especialização em Tradução de Língua Inglesa da Universidade de São Paulo (CETRAD).*

O objetivo deste artigo é demonstrar de que maneira a utilização do corpus eletrônico pode auxiliar em traduções de colocações do idioma inglês para o português. A escolha de investigar a tradução de colocações deve-se em grande parte às muitas possibilidades de análise que estas nos oferecem por seus padrões geralmente arbitrários e independentes de significados inter e intralinguais. Essa peculiaridade faz das colocações um tema bastante rico e necessário para um estudo mais detalhado, e um levantamento das colocações e suas equivalências na língua alvo pode vir a ser uma ferramenta inestimável para o tradutor, pois as colocações com frequência tornam-se um empecilho durante os trabalhos de tradução.

Devido à extensão do tópico elegemos enfocar somente colocações verbais do tipo verbo + substantivo, selecionando algumas colocações com os verbos make e perform e algumas colocações verbais com o

* Mestranda do Programa de Semiótica e Lingüística Geral da FFLCH/USP.

** Mestranda do Programa de Estudos Lingüísticos e Literários em Inglês da FFLCH/USP.

substantivo attention. Duas abordagens de pesquisa foram adotadas: a primeira com o intuito de confirmar certas hipóteses de tradução de colocações verbais com os verbos make e perform, partindo dos diferentes substantivos que os acompanham em diferentes tipos de discursos, utilizando as ferramentas descritas no decorrer do artigo; a segunda abordagem parte do substantivo attention que sabidamente gera diferentes colocações importantes em inglês e em português, e busca os diferentes verbos associados a este substantivo, bem como o ambiente semântico em que ocorrem as colocações encontradas.

Palavras-chave: Tradução; Corpus; Colocações; Colocações Verbais.

Abstract: *This article presents part of a paper written for a course attended during the Specialization Course in Translation Studies, CETRAD, at the Universidade de São Paulo.*

The objective of this article is to demonstrate how an Electronic Corpus can aid in the translation of Collocations from English into Portuguese. The choice of working with collocations came from the great number of possible analyses that they offer, as they are arbitrary and independent of any inter and intra language meanings, a peculiarity which causes problems during the translation process. As the topic is extensive, we decided to focus our studies on the verb + noun verbal collocations, selecting some collocations with the verbs make and perform and some collocations with the noun attention. Two different investigating approaches were used: the first one had the purpose of confirming certain translation hypotheses of the collocations with the verbs make and perform in different types of discourse with the help of the described tools. The second approach took the noun attention, which is known to generate different collocations in English and in Portuguese, and searched for the different verbs associated with this noun, as well as for the semantic field which occurs in conjunction with the collocations encountered.

Keywords: Translation; Corpus; Collocations; Verbal Collocations.

Introdução

A Linguística de *Corpus* ocupa-se da coleta e exploração de corpora, ou seja, de dados lingüísticos textuais coletados criteriosamente, e dedica-se à

exploração da linguagem através de evidências empíricas, extraídas por meio de computador. A linguagem é vista como um sistema probabilístico, o qual pressupõe que, apesar de muitos traços lingüísticos serem possíveis em teoria, a frequência na qual eles ocorrem na prática não é igual nem aleatória, variando em função do contexto específico de cada texto. Assim, podemos afirmar que existe um padrão de comportamento regular e sistemático da linguagem, que possibilita a utilização do *corpus* como fonte de informação e que, através desta informação, podemos identificar a frequência de ocorrência de traços lingüísticos de várias ordens e, posteriormente, estimar as probabilidades teóricas dessas ocorrências.

Existem vários tipos de pesquisas privilegiadas na Lingüística de *Corpus*: a compilação de *corpus*, o desenvolvimento de ferramentas, a descrição da linguagem e a aplicação de corpora (ensino de línguas, reconhecimento de voz, tradução, etc). A área da descrição da linguagem, onde é estudado o comportamento de itens lexicais ou de estruturas gramaticais no meio lingüístico em que ocorrem, é a que contém maior atividade, e dentre as várias questões de que a área da descrição se ocupa, uma das principais está centralizada na descrição de colocações. É possível verificar quais agrupamentos lexicais são empregados pelos falantes, ou atestados pelo uso, o que se concretizou em um princípio de entendimento da linguagem denominado de idiomático. O usuário de uma língua tem à sua disposição “um grande número de frases pré ou semiconstruídas, que se constituem em escolhas únicas, embora pareçam analisáveis em segmentos” (Sinclair, 1987). Para Sinclair, é possível prescrever-se com muita precisão as probabilidades de certos itens ocorrerem em contextos específicos, tornando os níveis do léxico e da gramática supérfluos.

Atualmente, a Lingüística de *Corpus* é uma das áreas mais vibrantes voltadas ao estudo da linguagem, principalmente por não se dedicar a um assunto definido, mas se ocupar de vários fenômenos (léxico, sintaxe, estrutura, etc). Existem dificuldades envolvidas ao se retratar um campo tão vasto: a primeira refere-se à grande quantidade de trabalhos novos que surgem e a segunda refere-se ao debate de cunho teórico, ou seja, a Lingüística de *Corpus* necessita explicitar qual é o quadro teórico que lhe dá coerência e sustentação, para deixar de ser um tipo de “contabilidade lingüística”, apenas registrando as ocorrências lexicais e textuais.

Por esta razão, muitos autores a definem como uma perspectiva, uma maneira de se chegar à linguagem. Para Hoey (1997), por exemplo, a Lingüística de *Corpus* não é um instrumental, mas uma abordagem. Leech (1992), por outro lado, a define como “uma nova empreitada de pesquisa e, na verdade, uma nova abordagem filosófica”.

Nos trabalhos de tradução a utilização do *corpus* é uma ferramenta inestimável porque a tradução de um mesmo termo pode variar dependendo da área em que se trabalha e do contexto em questão: áreas técnica, acadêmica, jornalística, literária, etc. Assim, uma tradução feita com o auxílio do *corpus* nos permite escolhas melhores, o que não ocorre quando utilizamos o dicionário como nossa principal ferramenta, pois as opções de traduções são limitadas e apresentadas de forma descontextualizada. A pesquisa do melhor termo, feita com o auxílio do *corpus*, nos ajuda a decidir o termo mais adequado para cada contexto.

Como nos propusemos a investigar as traduções de colocações e, mais especificamente, de colocações verbais, é importante que definamos colocações como palavras que tendem a coocorrer com frequência bastante elevada em uma dada língua, ou seja, as colocações têm a ver com a probabilidade de certas palavras serem utilizadas com outras e com a naturalidade ou tipicidade das combinações resultantes.

As colocações verbais representam uma das categorias das combinações lexicais convencionais e podem ser definidas como as combinações convencionais e consagradas de verbos utilizados com certos substantivos, adjetivos ou advérbios com alto grau de ocorrência.

Alguns exemplos destas colocações em inglês e suas respectivas traduções para o português:

Verbo + substantivo: *pay attention* traduzido como **prestar atenção**.

Verbo + adjetivo: *turn pale* traduzido como **ficar pálido**.

Verbo + advérbio: *climb aboard* traduzido como **subir a bordo**.

Substantivo + verbo: *a river flows* traduzido como **um rio corre**.

Diretamente ligada à noção de fluência numa língua encontra-se a idéia de convencionalidade e na área de tradução é evidente a necessidade de se conhecer as estruturas convencionais da língua de chegada, ou seja, o tradutor deve resgatar aquele “jeito de dizer” peculiar à língua para a qual traduz. As convenções lingüísticas são os “jeitos” aceitos pela comunidade que fala determinada língua (Tagnin, 1989). Chama-se convencionalidade o aspecto que caracteriza a forma peculiar de expressão numa dada língua. Em uma tradução é importante recuperar as combinações léxicas convencionais, no intuito de manter a naturalidade da língua alvo.

Um exemplo que ilustra de maneira bem objetiva a dificuldade de traduzir colocações verbais é o do substantivo *button* que em inglês é utilizado com o verbo *push*. A expressão *push the button* é uma colocação consagrada em inglês, e é traduzida para o português como **apertar o botão**; apesar de o verbo *push* ser

traduzido como **empurrar**, em português, a colocação usual é **apertar o botão** e não **empurrar o botão**.

É possível que um tradutor menos cuidadoso e com pouco conhecimento sobre colocações verbais e sobre a língua de chegada, possa recorrer a um dicionário, procurar a palavra *push* e traduzi-la como **empurrar** sem dar-se conta de que o verbo **empurrar** não ocorre com o substantivo **botão** em português. Por outro lado, a colocação *push the door* em inglês é traduzida como **empurrar a porta** e não **apertar a porta**. Outros exemplos mostram que os verbos empregados diferem muitas vezes de uma língua para outra: *take a step* / **dar um passo**; *give a lecture* / **dar ou fazer uma palestra**.

As colocações são uma área interessante para os estudos contrastivos porque os padrões de colocação são geralmente arbitrários e independentes do significado tanto dentro de uma língua como entre as línguas. As diferenças dos padrões colocativos das línguas referência e alvo criam armadilhas em potencial que apresentam vários problemas de tradução. Baker (1992) alerta que “quando a tradução de uma palavra ou frase é criticada como imprecisa ou inapropriada em um dado contexto, a crítica pode se referir à inabilidade do tradutor de reconhecer um padrão colocativo com um significado único diferente da soma dos significados de seus elementos individuais”.

Outra vantagem da utilização de corpora eletrônicos é que muitos deles estão em constante mutação, da mesma forma que a língua; novos termos são criados a cada dia, principalmente na área técnica e, nesse aspecto, esse tipo de *corpus* é muito mais representativo do que o dicionário; porém, é importante trabalhar com corpora diferentes para confirmar a validade da informação obtida, não esquecendo que um *corpus* fechado também é uma ferramenta extremamente valiosa em pesquisa.

Nossa pesquisa acerca de colocações verbais demonstrou ser esse um tópico bastante extenso e, por essa razão, decidimos focar as colocações verbais do tipo *verbo + substantivo*. Algumas colocações com os verbos *make* e *perform*, foram selecionadas para um estudo mais detalhado. Também foram estudadas as colocações verbais com o substantivo *attention*.

A escolha dos verbos acima mencionados deu-se em função de suas diferenças: *make* é um verbo irregular de alta incidência em inglês, enquanto *perform* é regular e menos freqüente; assim nos pareceu relevante investigar de que forma estas diferenças poderiam influenciar os resultados de nosso estudo. Na pesquisa baseada em um substantivo optamos por *attention*, palavra que está presente em diversas colocações verbais. A decisão de adotar duas metodologias de pesquisa teve como objetivo primário enriquecer nossa pesquisa e torná-la mais abrangente.

Colocações verbais com o verbo *make*

O primeiro passo para a escolha das colocações verbais a serem estudadas com o verbo *make* foi o de entrar no site do *Cobuild* com o objetivo de checar as ocorrências de colocações com esse verbo. Parte dos resultados encontrados estão apresentados no Apêndice 1.

Como as ocorrências apareceram de forma aleatória nessa busca inicial, procuramos verificar como as colocações com *make* apareciam estatisticamente. Parte dos resultados obtidos podem ser verificados no Apêndice 2.

A decisão de trabalhar com as colocações: *make sure*, *make money*, *make a mistake*, *make a choice*, *make arrangements*, *make an appointment*, *make friends* e *make an impression* foi feita com base na frequência em que mesmas são encontradas no *Cobuild*, ou seja, procuramos escolher algumas colocações que apareciam com bastante frequência, outras com frequência relativa e outras ainda que ocorriam com pouca frequência para que o estudo se tornasse mais significativo.

As possíveis traduções para as colocações selecionadas sugeridas pelos livros, e dicionários consultados foram:

make arrangements - tomar providências	make friends - fazer amigos
make an appointment - marcar uma consulta	make money - ganhar dinheiro
make a mistake - cometer um erro	make sure - ter certeza/fazer com que

Como as traduções foram apresentadas fora de contexto, a próxima etapa da pesquisa foi a de confirmar, primeiramente através do COMPARA, que é um *corpus* literário e paralelo (inglês-português), se as mesmas poderiam ser traduzidas pelas colocações sugeridas. Trabalhamos com textos originais em inglês e suas correspondentes traduções para o português do Brasil. Inicialmente tivemos dificuldades em encontrar colocações com o verbo *make* em inglês, pois iniciamos a busca pelo verbo; posteriormente, fomos informados pela Profa. Diana Santos, uma das responsáveis pelo projeto COMPARA, que, como o verbo *make* é um verbo curto e irregular, seria muito difícil obter ocorrências ao efetuar a busca pelo verbo. Consequentemente a busca foi refeita utilizando os diferentes substantivos que acompanhavam o verbo *make*; a partir daí, a pesquisa se tornou produtiva. Parte dos resultados encontrados são apresentados no Apêndice 3.

As previsões de todas as traduções foram confirmadas no COMPARA, com exceção de *make arrangements*, porque não foi encontrada nenhuma ocorrência desta colocação. Portanto, decidimos concentrar a pesquisa nessa colocação, verificando em outros corpora eletrônicos, se **tomar providências** poderia realmente ser aceito como sua tradução.

Para isso, primeiramente pesquisamos a colocação *make arrangements* no *corpus WebCorp*, em inglês, e a seguir, sua possível tradução para o português **tomar providências**, no mesmo *corpus*. Foram encontradas 120 páginas contendo 256 ocorrências da colocação *make arrangements* e 120 páginas onde apareciam 272 ocorrências da colocação **tomar providências**. É interessante mencionar que esses resultados demonstram uma ocorrência bastante significativa dessa colocação verbal em inglês e também de sua tradução para o português.

A partir daí, o próximo passo foi o de buscar essa colocação em outros corpora, dessa vez corpora comparáveis, já que os resultados no *corpus* paralelo se mostraram infrutíferos.

No Apêndice 4 apresentamos uma seleção dos resultados obtidos em textos equivalentes encontrados em diferentes sites, em inglês e português comparando o uso dessa colocação em uma variedade de contextos: técnico (turismo, veterinária, informática, jurídico, esportivo), jornalístico, acadêmico, religioso e empresarial provenientes de diferentes corpora eletrônicos: *WebCorp*, *BNC*, e o *corpus* técnico dos Alunos do CETRAD. Para auxiliar nossa pesquisa, utilizamos os instrumentos de busca WordSmith Tools e KWIC Finder.

Com base nas informações obtidas nos diferentes corpora, pode-se dizer que a proposta inicial de traduzir *make arrangements* por **tomar providências** foi confirmada em todos os contextos pesquisados (Apêndice 4), ou seja, esta colocação está presente em contextos acadêmicos, empresariais, técnicos e literários em português.

Colocações verbais com o verbo *perform*

A pesquisa referente a colocações verbais com o verbo *perform* foi feita de forma mais simplificada do que a com o verbo *make*. Nosso objetivo principal foi o de trabalhar com um verbo menos freqüente, porém regular, além de ser uma palavra mais longa para verificar se esses fatos alterariam o processo de busca e os resultados obtidos. Tendo esse propósito em mente e para que nosso trabalho não se estendesse por demais, decidimos trabalhar com apenas uma colocação verbal com o verbo *perform*. Escolhemos a colocação verbal: *perform a contract*, para a qual tínhamos três sugestões de traduções encontradas em dicionários: **cumprir contrato / levar a cabo o contrato / fazer valer o contrato**.

O fato de não obtermos nenhuma ocorrência dessa colocação verbal no COMPARA, nem através da pesquisa pelo verbo e nem pelo substantivo era previsível, pois sendo esse um *corpus* literário, não seria esperado encontrarmos colocações relacionadas a contextos técnicos, como é o caso de **cumprir contrato**. A seguir apresentamos uma pequena amostragem dos exemplos encontrados

no *corpus WebCorp* e *Kwic Finder*. Mais exemplos podem ser encontrados no Apêndice 5.

Utilizando o *WebCorp* encontramos 121 ocorrências de *perform a contract* em 120 páginas acessadas. Pudemos observar que os contextos eram quase em sua totalidade jurídicos:

- 1) <http://www.judiciary.state.nj.us/charges/civil/419.htm>
that whether seller's refusal to perform a contract for the sale of a
- 2) <http://biz.findlaw.com/legal/contracts/faqs.html>
court order a party to perform a contract rather than pay damages?

Quanto às possíveis traduções dessa colocação verbal encontramos: 5 ocorrências de **fazer valer o contrato** em 7 páginas acessadas, o que significa poucas ocorrências da colocação em questão. Além disto, analisando os exemplos encontrados (Apêndice 6), verificamos que o uso dessa expressão implica dizer que o contrato não estava sendo respeitado como deveria, portanto **fazer valer o contrato** seria quase que forçar a sua continuidade, o que presume uma diferença de significado:

- 1) <http://www.estado.estadao.com.br/editorias/2002/02/14/eco043.html>
presidente da Transbrasil e quer fazer valer o contrato, assinado em janeiro, que lhe
- 2) <http://www.mj.gov.br/dpdc/clipping/clip1909.htm>
Mallet, o único meio para fazer valer o contrato e receber a indenização

Em 71 páginas acessadas obtivemos 77 ocorrências da expressão **levar a cabo**, porém somente uma relacionada a **contratos**, e mesmo assim colocada com o substantivo **cumprimento**:

- 1) <http://www.trekbrasilis.aidi.com.br/tng/episodios/tng003c/>
da idéia de levar a cabo o cumprimento do terrível contrato

Nas outras páginas acessadas encontramos a expressão **levar a cabo** colocada com outros substantivos, por exemplo: tarefas, obras públicas, estratégias, projetos, estudos, iniciativas etc. Portanto **levar a cabo o contrato** não é uma colocação verbal usual em português, e traduzir *perform a contract* por **levar a cabo o contrato** não equivale à “forma como falamos” e portanto não soaria natural.

Como uma forma de complementar nossa pesquisa e também por curiosidade, tentamos encontrar no COMPARA textos originais em português que contivessem os substantivos: tarefas, obras públicas, estratégias, projetos, estudos, iniciativas para ver se havia alguma ocorrência dos mesmos com a expressão **levar a cabo**, já que havia diversas ocorrências no *WebCorp*, porém, nada foi encontrado.

Crop, 10, 2004

A seguir prosseguimos nossa pesquisa, buscando ocorrências da colocação **cumprir contrato**, pois era outra dentre as possíveis traduções para a colocação *perform a contract*. Em uma pesquisa parcial no *WebCorp* encontramos 7 páginas contendo 32 ocorrências desta colocação, todas dentro de contextos jurídicos, ou empresariais, como ocorre com a colocação *perform a contract* em inglês, como podemos verificar nos exemplos abaixo: (Outros exemplos podem ser encontrados no Apêndice 7)

- 1) http://www.sfiec.org.br/artigos/energia/Na_energia_ningu%E9m_se_salva_sem_perdas.htm
contrato nenhum. Cobrar dívida é cumprir contrato. Que quebra de contrato é
- 2) <http://www.espacovital.com.br/obrigacoes.htm>
para reparo desobriga inquilino a cumprir contrato Editora Globo pagará passagens não
- 3) (<http://jornal.publico.pt/2004/03/26/Desporto/D13.html>)
Fernando Santos diz que vai cumprir contrato com o Sporting João Pinto
- 4) <http://an.uol.com.br/2001/jul/17/0moa.htm>
aval, condicionando a AmBev a cumprir contrato da nova indústria ou a

Com o auxílio da ferramenta de busca *Knix Finder* encontramos mais exemplos da colocação **cumprir contrato**, em contextos semelhantes aos encontrados no *WebCorp*. (Mais exemplos podem ser encontrados no Apêndice 8).

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

<http://www.terra.com.br/exclusivo/noticias/2002/08/01/022.htm>

Original

Query term 1, “cumpr* contrato”, matches: 1

Big Brother Brasil 2 »	cumprir	com Rede TV! Quinta, 01 de agosto de 2002,
Casa dos Artistas 3 »	contrato	12h44 O jornalista Marcelo Rezende (Foto:
Fama » Popstars		Divulgação/ Rede TV!) O jornalista Marcelo
Notícias Marcelo		Rezende declarou através de nota divulgada pela
Rezende afirma que vai		Rede TV! que irá cumprir seu contrato com a
		emissora, desmentindo os boatos de que iria para a
		Record assumir o Cidade

http://www.jurismidia.com/bdnot_03.htm

Original

Query term 1, “cumpr* contrato”, matches: 2

É obrigação da seguradora pagar	cumprir	Médico associado à cooperativa
indenização inicialmente concordada	contrato	que proíbe dupla militância não
em apólice Dirigir na contramão não		pode prestar serviços à outra
exclui obrigação da seguradora de		empresa STJ:

A conclusão à qual pudemos chegar segundo este nosso estudo é que das três hipóteses levantadas como possíveis traduções para a colocação verbal *perform a contract*, a colocação **cumprir contrato** pode ser aceita como a mais usual dentro dos contextos pesquisados.

Em resumo, podemos dizer que as nossas escolhas para as traduções das colocações *make arrangements* e *perform a contract* são confiáveis por resultarem de uma pesquisa da frequência e uso destas colocações dentro de corpora comparáveis e paralelos, compostos de textos aleatórios e de conteúdo variado. Esta diversidade oferecida pela tradução baseada em *corpus*, nos brindou com

traduções reais e comunicativas, proporcionando resultados representativos e que, portanto, podem ser estendidos a um todo.

Colocações Verbais com o substantivo *atenção* / *attention*

Em um enfoque distinto, buscamos ver quais possibilidades um mecanismo de busca como o *WordSmith Tools* poderia oferecer na geração (produção) de colocações, bem como tipos de palavras e de “ambiente” semântico que podem coocorrer com as colocações geradas. Escolhemos partir do substantivo **atenção** que sabidamente gera diversas colocações importantes no inglês e em português.

As questões que nortearam esta pesquisa foram:

- 1 Como **identificar** uma colocação? Como distingui-la das combinações livres* (combinações menos coesas entre os termos)?
- 2 Na tradução, uma colocação na língua fonte sempre deve corresponder a uma colocação na língua alvo para manter o “jeito de dizer” natural da língua alvo?
- 3 Que problemas de tradução as colocações podem oferecer?

Para Benson³, as combinações livres são previsíveis, pois podem ser criadas por qualquer aprendiz que conheça o vocabulário e a gramática de uma língua. Na terminologia de Saussure, pertencem à *parole*. Já as colocações são combinações não previsíveis, arbitrárias e pertencem ao sistema da língua, a *langue*. Benson define as colocações como “combinações lexicais arbitrárias recorrentes”.

Nossas fontes de busca foram:

- 1) O *Corpus* Jornalístico do CETRAD, que contém 1.971.630 palavras (*tokens*) oriundas das diversas seções dos jornais *Financial Times*, *Los Angeles Times*, *The Guardian*, *The New York Times*, *USA Today* e *Washington Post*. Podemos inferir uma grande riqueza lexical a partir do *Type Token Ratio* que varia de 6 a 10.
- 2) O *CetenFolha* (**Corpus** de **Extractos** de **Textos** **Electrónicos** **NILC/Folha** de S. Paulo) é um *corpus* de cerca de 24 milhões de palavras em português brasileiro, criado pelo projeto Processamento Computacional do Português com base nos textos do jornal *Folha de S. Paulo*.

Escolhemos dois corpora jornalísticos pelo fato de os jornais cobrirem uma ampla e diversificada gama de assuntos podendo assim observar-se o comportamento das palavras que co-ocorrem com “atenção” em vários contextos.

- 3) O COMPARA, um *corpus* paralelo bi-direcional português-inglês baseado em um conjunto em aberto de textos-fonte e traduções de português-inglês e inglês-português, foi usado para verificar se a tradução

de colocações da língua fonte sempre gera colocações na língua alvo ou outras possibilidades que se aproximam mais do “jeito de dizer” da língua alvo e também para comparar com os equivalentes traduzidos encontrados em dicionários bilíngües.

- 4) O *WebCorp*, um conjunto de ferramentas que permite acesso a toda a Rede da Internet enquanto *corpus*, de onde fatos sobre a língua podem ser extraídos para estudo – no nosso caso interessava ver o contexto e o número de vezes que uma colocação ocorria para testar sua **recorrência**, considerada uma característica distintiva das colocações por alguns autores.
- 5) *Activator*: um dicionário de produção, baseado em extensos *corpora*, conjuntamente denominados *Longman Corpus Network*. Os editores salientam a importância atribuída às colocações. Este dicionário foi usado para distinguir o significado das diferentes colocações geradas com o substantivo *attention*.
- 6) *Barsa*: usamos um dicionário bilíngüe tradicional para averiguar os equivalentes que serão encontrados por um tradutor que usa este recurso somente.

A ferramenta de busca *WordSmith Tools* é um conjunto de programas para o exame de como as palavras se comportam nos textos. Usamos a ferramenta *Concord*, que busca a palavra desejada e pode apresentá-la de forma centralizada, exibindo-a em um contexto lexical com um certo número de palavras à esquerda e à direita. Esta forma de apresentação permite analisar os padrões de co-ocorrência da palavra-chave.

A geração de colocações a partir do substantivo *atenção* e do mecanismo de busca *WordSmith Tools*

Optamos por procurar não apenas as possíveis colocações, mas também algumas palavras que recorriam com as colocações, o “ambiente” semântico que cercava as combinações lexicais.

Em inglês, através da ferramenta *Concord* do *WordSmith Tools*, obtivemos as seguintes combinações lexicais e palavras que acompanham o substantivo *attention* em 188 ocorrências no *corpus* jornalístico do CETRAD:

Accord attention; Attract (undivided) attention; Bring (to) attention; Capture (the) attention (of); Deflect attention; Demand attention; Deserve attention; Devote attention; Direct attention (to); Distract attention; Draw (the most) attention (from / to); Focus (refocus) attention; Gain attention; Garner attention; Get (a lot of) attention;

Crop, 10, 2004

Give (scant) attention; Grab attention; Keep (the audience) attention; Lavish attention; Lift (their) attention (away); Pay (much attention; a whiff of attention; closer attention); Stretch (the) attention; Turn attention (to); Warrant attention

Em português, através do *CetenFolba*, obtivemos 3093 ocorrências de atenção, entre as quais selecionamos algumas pela sua frequência:

Agradecer a atenção; Atrair (especial) atenção; Chamar (particular) atenção; Concentrar atenção; Dar atenção (especial); Dedicar (uma) atenção (considerável); Despertar a atenção; Desviar a atenção; Dispensar atenção (atenção dispensada); Exigir atenção; Focar atenção; Ganhar a atenção; Justificar a atenção; Manter a atenção; Merecer (tanta / pouca/particular) atenção; Mobilizar a atenção; Pedir (a sua) atenção; Precisar de (muita) atenção; Prender a atenção; Prestar atenção; Querer atenção; Receber atenção; Redobrar a atenção; Voltar (a) atenção (a).

Colocações estudadas

Como foge ao escopo deste trabalho o exame de todas essas combinações lexicais, resolvemos nos concentrar, através do exame das palavras que colocam com *attention* (Apêndice 9), nas combinações que certamente são colocações pela grande incidência de ocorrências, como *pay attention*, *attract attention*, *draw attention*. O Apêndice 10 mostrará o contexto geral dos verbos que mais co-ocorrem com *attention* e também o contexto específico de cada uma das colocações. Examinamos brevemente também *call attention* e *give attention*, por serem traduzidas respectivamente por **chamar atenção** e **dar atenção**, duas prováveis colocações de alta incidência em português.

Em português examinamos **prestar atenção**, **atrair atenção**, **chamar atenção**, **dar atenção** por serem combinações lexicais mais recorrentes e prováveis colocações (Apêndice 11).

Note-se, não obstante, o fato de que para alguns autores a recorrência não é um indicador absoluto de que uma determinada combinação lexical seja ou não uma colocação. Mesmo o exame destas poucas prováveis colocações foi superficial, devido ao espaço exíguo do qual dispomos aqui.

Ademais, achamos interessante a combinação *lavish attention*, que aparece duas vezes no *corpus* jornalístico e três vezes no *Cobuild*, mas um número muito alto de vezes no *WebCorp*, o que faz suspeitar de uma colocação comum em inglês que pode oferecer problemas de tradução, e resolvemos estudá-la mais a fundo (Apêndice 12).

Quanto ao “ambiente semântico” concentramo-nos em *scant attention* e *close attention*, sendo *scant* e *close* dois adjetivos em inglês que ocorrem com grande frequência no *WebCorp*, embora com pouca frequência no *corpus* jornalístico.

Resultado

Pelo exame de alguns exemplos com as colocações *attract attention*, *draw attention* e *call attention* vimos que ao serem traduzidas elas às vezes podem apresentar o mesmo tipo de produção em português. Vejamos os seguintes exemplos:

- 1 The women’s movement *drew people’s attention* to the discrimination against women in traditional jobs;
- 2 I don’t want to *draw attention* to myself. I think I’ll sit in the back row

Nestas duas sentenças, *draw attention* poderia ser traduzido por **chamar a atenção das pessoas** e **chamar atenção para mim mesmo**. Na sentença: “It may be a good idea to *call your employer’s attention* to those broken windows”, também em português poderíamos traduzir por: **chamar a atenção de seu empregador**, e em: “Tall and beautiful, she *attracted attention* wherever she went”, poderíamos ter **atrair a atenção** e **chamar a atenção**. Até na sentença com a (provável) colocação *get attention*: “Unhappy kids often steal just *to get attention*” a solução em português seria **Crianças infelizes muitas vezes roubam para chamar atenção**. Assim, com várias colocações em inglês produzindo a mesma tradução em português, o uso dos dicionários às vezes pode confundir, mais do que esclarecer.

O Apêndice 13 aponta para as soluções dos dicionários *Activator* e *Houaiss*. Vemos que muitas vezes é difícil dizer com precisão quais colocações em inglês são mais bem traduzidas pelas colocações correspondentes em português. Pelo exame do Apêndice 10 vemos que, pelos menos na linguagem jornalística, *draw attention* repete-se 21 vezes, muito mais que *call attention* (duas vezes), embora o significado das duas colocações seja semelhante (Apêndice 15), o que indica que *call attention* é muito menos usada. Em português a colocação **chamar atenção para** parece ser a solução melhor para *draw attention* e *call attention*, pelo menos no contexto jornalístico.

E quanto a *pay attention* e *give attention*? Será que invariavelmente serão traduzidos por **prestar atenção** e **dar atenção**? Se depender dos dicionários comuns, sim. Mas no COMPARA encontramos diversas soluções interessantes e outras possibilidades de equivalência, algumas das quais soam muito mais naturais em português. Vejamos o Apêndice 14:

Chamar atenção: *bear in mind*, *catch the eye* (2), *attract attention*:

Crop, 10, 2004

O recato **chamaria mais atenção** que a nudez: *modesty would attract more attention than nudity*. (Vemos aqui **chamar atenção** corretamente traduzida como *attract attention*).

The headword “Dover’s Powder” *caught my eye*: a palavra “pó de Dover” **chamou-me a atenção**.

A parrot’s perch *catches the eye*: Um poleiro de papagaio **chama-nos atenção**.

There are a few points you might like *to bear in mind*: Há alguns detalhes para os quais gostaria de **chamar a sua atenção**.

Dar atenção: *take notice, pay attention, listen*:

Took no notice of anyone: **Não dava atenção a ninguém**.

Give a special thought: **Dar atenção especial**

pay attention to one another: **Dar atenção uns aos outros**

Vemos aqui uma equivalência de *pay attention* com **dar atenção** e não com **prestar atenção**

Prestar atenção: *listen*:

Você **não presta atenção** no que eu digo: *you don’t listen* when I talk

Reparar: *pay attention*:

Ninguém mais **repara** em mim: *Nobody pays any attention* to me anymore

Nhonhô, **não repares** nesse grande manhoso – Nhonhô, *don’t pay attention* to that...

É interessante como **não repare** ou **ninguém repara**, sem ser uma colocação, soam muito mais natural em português do que **não preste atenção**, dependendo do contexto.

Com os adjetivos *scant* e *close* encontramos:

Scant attention: **fazer pouco caso** (também uma colocação em português): ver Apêndice 15.

Aos que antes **faziam pouco caso de**: *Who would pay scant attention* to him.

Nosso **fazer pouco caso** pode substituir *scant attention* aqui, resgatando a naturalidade da língua, embora esta solução jamais apareceria num dicionário comum!

Não encontramos sugestão em português do Brasil para *close attention*, mas apenas uma tradução para o português de Portugal de **atenção cerrada**. Podemos sugerir **atenção concentrada** (ver final do Apêndice 11) no *Webcorp*. Tampouco encontramos equivalente para *lavish attention*, o que torna esta co-

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

locação difícil de traduzir. No *Concise Oxford Dictionary* encontramos para *lavish*: **profuse, excessive, over abundant**. Na *Barsa*: **pródigo, generoso, liberal**. Parece também co-ocorrer com *detail*: *lavish attention to detail*. Poderia ser traduzido como **cumular de atenção** ou **prestar extrema atenção**? Eis uma colocação com **atenção** que merece um estudo **atento**!

Conclusão

Não dispomos de espaço suficiente neste trabalho para refletir com mais profundidade sobre todas as possibilidades que estes exemplos obtidos nos corpora vêm acrescentar ao ofício da tradução. Mas é fascinante pensar que indiscutivelmente podemos chegar mais perto do “jeito que a gente diz”, ou da convencionalidade da língua, através da utilização deste recurso de tradução que são os diferentes corpora e as ferramentas de busca. Trata-se de ferramenta que nos põe em guarda contra a “rigidez determinista” do dicionário e o “hábito literalista” do tradutor desatento.

As abordagens que adotamos permitiram a execução de uma série de tarefas: a busca e localização de uma colocação específica no *corpus* paralelo obtendo assim uma ou várias opções de tradução; a confirmação de que se está de fato diante de uma colocação verbal, através da constatação de seu uso, a confirmação de uma acepção corrente de uma colocação e a constatação de uma segunda acepção até então desconhecida. O uso de corpora na pesquisa das colocações verbais com vistas à tradução tornou possível e fascinante o que talvez sem eles fosse quase inviável ou extremamente árduo.

Enfim, é interessante refletir, além disso, sobre a observação de Kennedy (1998, p. 93) de que a análise de um *corpus* pode revelar, e freqüentemente revela, fatos a respeito de uma língua que nunca se pensou em procurar. Isso indica que a Linguística de *Corpus* não reflete apenas um meio mais rápido de descrever como funciona a linguagem, mas torna possível também a obtenção de dados novos sobre a língua, antes insuspeitados, descortinando assim horizontes que merecem ser explorados.

Fontes de Referência

1. Sites:

Alta Vista:

<http://www.altavista.com>

BNC:

<http://sara.natcorp.ox.ac.uk/lookup.html>

Cobuild:

<http://collins.co.uk/Corpus/CorpusSearch.aspx>

Crop, 10, 2004

COMPARA:	http://www.linguateca.pt
Google:	http://www.google.com http://www.google.com.br
Webcorp:	http://www.webcorp.org.uk/webcorp.html
WordSmith Tools:	http://www.liv.ac.uk/~ms2928.wordsmith.index.htm

2. Corpus em CD-ROM:

Corpus da Folha de S. Paulo – **EDIÇÃO 2000**
Corpus jornalístico do Cetrad
Corpus Técnico dos Alunos do CETRAD, 2001

3. Bibliografia:

- AUBERT, F.H. *As Infidelidades da Tradução*. Campinas: Edit. da Unicamp, 1994.
- BAKER, M. *In Other Words*. New York: Routledge. 1992.
- BENSON, M. et al (1986) *The BBI dictionary of English word combinations*. Amsterdam / Philadelphia: John Benjamins, 1986
- BERBER SARDINHA, A. P. Lingüística de Corpus: histórico e problemática. *DELTA – Revista de Documentação de Estudos em Lingüística Teórica e Aplicada*. São Paulo: v. 16, n. 2, p. 323-367, 2000
- COLLINS. *Dictionary of English Phrasal Verbs and their Idioms*. London: Collins. 1990.
- COLLINS. *English Dictionary Millenium Edition*. UK: Harper Collins Publishers.1999.
- HOEY, M. From concordance to text structure: New uses for computer corpora. In: B. LEWANDOSWKA-TOMASZCZYK & P. J. MELIA (Org.). *PALC'97 – Practical Applications in Language Corpora*. Lodz: Lodz University Press, 1997
- HOUAISS, A. *Dicionário Houaiss da Língua Portuguesa*. 2001. 1. ed. Rio de Janeiro: Objetiva. 2001.
- KENNEDY, G. 'Between' and 'through'. The company they keep and the functions they have. In: K. Aijmer e B. Altenberg (Org.). *English Corpus Linguistics – Studies in honor of Jan Svartvik*. London/New York: Longman, 1991
- LANGUAGE ACTIVATOR. Essex: Longman. 1994.
- LEECH, G. Corpora and theories of linguistic performance. In: SVARTVIK, J. (Org.). *Directions in Corpus Linguistics*. Berlin, New York: De Gruyter, 1992
- LONGMAN *Dictionary of English Idioms*. England, Longman. 1979.

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

LTP Dictionary of Selected Collocations. England: LTP. 1990.

NOVO DICIONÁRIO BARSÁ. New York: Appleton-Century-Crofts. 1968.

OXFORD Collocations Dictionary for Students of English. Oxford: Oxford University Press. 2002.

OXFORD Dictionary, The Concise. Oxford: Clarendon Press. 1990.

SINCLAIR, J. *Corpus, Concordance and Collocations*. Oxford: Oxford University Press. 1991

SWAN, M. *Practical English Use*. Oxford: Oxford University Press. 2002.

TAGNIN, S. O. *Expressões Idiomáticas e Convencionais*. São Paulo: Ática. 1989.

TAGNIN, Stella O. *Convencionalidade e Produção de Texto: Um Dicionário de Colocações Verbais Inglês/Português*. Tese de Livre-Docência, não publicada. Universidade de São Paulo. 1998.

WEBSTER Unabridged Dictionary. New York: Prentice Hall Press. 1972.

Crop, 10, 2004

Apêndice 1 – *Cobuild* – Ocorrências de colocações com *make* + substantivo

wait if that's **the choice that they make**. Totenberg: the one you book, not a similar one; **make a note** of the still waiting for someone special to **make a commitment**, and bureaucracy, terrified they might **make a mistake**. The PC business must either **make a profitable transition** to the best of my ability and probably **make a good job** of it. turning away only long enough to **make a brief notation** on the intermediary structures would have **to make certain** beliefs an opportunity to ask questions or **make comments** about local the perplexing vagaries of politics **make friends** of former foes you regularly. It may be necessary to **make limited changes** being a spy come to **make a report**. To make matters worse, where the farmer decides he'll **make more money** by letting which was an unpleasant **choice to make**," she says. the manufacturer's instruction. **Make sure** your fuses My popularity with the fans helped me **make this decision**

Apêndice 2 – *Cobuild* – Tabela de frequência de colocações com *make* + substantivo

<u>Collocate</u>	<u>Corpus Freq</u>	<u>Joint Freq</u>	<u>Significance</u>
Sure	16016	3515	57.519008
Difference	4895	852	28.091225
Sense	10847	699	23.752850
Decision	9834	616	22.225552
decisions	2670	509	21.786302
Money	25274	698	20.157272
changes	6965	357	16.481302
Mistake	2153	293	16.293852
Effort	4520	309	15.895121
Use	25110	523	15.681465
contribution	1784	224	14.186321
mistakes	1161	212	14.038232
Things	32931	521	13.380862
progress	3499	220	13.288109
Way	52936	693	13.161123
Love	18553	365	12.747808
choice	6869	233	12.318483
arrangements	1566	168	12.170561
choices	949	150	11.740205
debut	2192	165	11.728127
impact	3225	173	11.547845
move	12416	266	11.325981
appointment	1897	141	10.828530
statement	4683	173	10.822190
profit	2967	152	10.753426
point	19526	319	10.703856
mind	12985	252	10.519779
contact	5679	174	10.372568
friends	11713	227	9.977302
order	12647	234	9.884839
attempt	4629	149	9.724052
amends	150	96	9.697740
impression	1769	116	9.695116
matters	3026	129	9.613723

Crop, 10, 2004

Apêndice 3 – COMPARA – Exemplos com a colocação *make* + substantivo

MAKE A MISTAKE

EBDLIT2 (1445):	I wrote her out a cheque for twenty, once, but this only caused more embarrassment because she ran after me in the car park and said I'd made a mistake .	Uma vez, passei um cheque de vinte, mas isso só causou mais constrangimento, porque ela correu até o estacionamento para dizer que eu tinha cometido um erro .
-----------------	--	---

MAKE MONEY

EBDLIT2 (1896):	In London he worked for one of the big stores in the West End and never made enough money to buy a decent car or a house big enough for his growing family.	Em Londres, trabalhou para uma dessas grandes lojas no West End e nunca ganhou dinheiro bastante para comprar um carro ou uma casa decente para sua família.
-----------------	--	---

MAKE SURE

EBDLIT2 (1586):	Looking after a twelve-year-old boy whose chief topic of conversation is computer games, and making sure he does his homework, is obviously not her natural vocation.	Tomar conta de um garoto de doze anos, cujo interesse maior são os jogos de computador, e fazer com que ele faça seu dever de casa está claro que não é sua vocação natural.
EBDLIT2 (1709):	Well, the first thing I'm going to do tomorrow morning is find out who the hell they were and who invited them, and make sure they never come to a recording again.	Bom, a primeira coisa que vou fazer amanhã é descobrir quem foi o diabo que os convidou e fazer com que nunca mais apareçam para uma gravação aqui de novo.

MAKE AN APPOINTMENT

EBDLIT2 (1446):	When I was dressed she came back into the room and we made an appointment for two weeks' time.	Quando me vesti, ela voltou à sala e marcamos a próxima consulta para daqui a duas semanas.
-----------------	---	--

MAKE FRIENDS

ESNG1 (572):	He'll make friends .	Ele fará amigos.
--------------	-----------------------------	------------------

MAKE ARRANGEMENTS Procura: [word="arrangements" & _texto="E.*"& ((_texto="PB.*") | (_variantetrad="brasileira"))].
Pedido de : concordância em contexto.
Corpus: COMPARA_ING
Nenhum caso foi encontrado

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

Apêndice 4

WebCorp – Exemplos com as colocações *make arrangements* e tomar providências em diferentes contextos

I) <http://www.captlew.com/contact.htm>

contact the Captain Lew and make arrangements for a fabulous day at

<http://www.sportnet.com.br/artigos/wmprint.php?Art/D=285>

e irrita torcedores que prometem tomar providências pelo desrespeito total

II) <http://www.nbcci.org/objective.htm>

U.K. To mediate or make arrangements to mediate in any dispute

http://universoespírita.org.br/textos%20interessantes/licao_vida.htm

será prejudicada se ninguém tomar providências para ajudá-lo. A professora

III) <http://www.uiowa.edu/admissions/graduate/faqs.html>

program (link) . How do I make arrangements to take the GRE/GMAT

http://www.abel.com.br/edu_on/historia/trabalhos/trab4/m5.html

que autorizava o exército a tomar providências para a construção.

IV) <http://e-newsletters.internet.com/samples/intmedia-research-alert-html.html>

Analyst Michael J. Pastore may make arrangements by contacting Mary Ann Boland

<http://www.ointernauta.com.br/conexao/link10.htm>

muitos já estavam mobilizando-se para tomar providências reais, pois acreditaram no e-mail

Kwic Finder: Exemplos com as colocações *make arrangements* e tomar providências em diferentes contextos

V) <http://www.physical.com/jchat/messages/50315.htm>

I'll make arrangements to be there on Saturday 11 March 2000 at 6.25 a.m.; in response to you, I'll make arrangements to be there! Rhonda! Oh Rhonda! I'm so excited! I cannot wait! You are so sweet!

Crop, 10, 2004

<http://dantoplace.blog.ig.com.br/200206.html>

me empolguei um pouco... eu acho melhor você não deixar mais seus comentarios ou então vamos ter que tomar providências sobre sua pessoa

BNC: Exemplos com as colocações *make arrangements* e tomar providências em diferentes contextos

VI) AD1 Gentleman and ladies. Hill, Susan. London; Hamish Hamilton Ltd, 1969, pp. 5-138.

Alida Thorne wiped away the tears with the back of her soft hand, only wanting to be taken to bed, like a child, and soothed, to have someone decide and make arrangements, tell her that all would be well, she should have her way.

Corpus Técnico dos Alunos do CETRAD (utilizando *WordSmith Tools*) – Exemplos com as colocações *make arrangements* e tomar providências em diferentes contextos

VII) d:\corpus~7\corpus\doenç~11\ingl_s~5\origem\ioma)~88.txt

The sole supplier of pertussis vaccine which uses bovine digest. You should also be aware that DH has made arrangements for meningococcal vaccine to be available, on a named patient basis, from SKF and Merieux.

d:\corpus~7\corpus\genoma\portug~7\origem\poendo.doc21

Queira o senhor perito relatar as providências tomadas para não contaminar ou violar o material durante o transporte e armazenamento

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

Apêndice 5 – *WebCorp* – Exemplos com a colocação *perform a contract*

WebCorp – perform a contract
Accessed 120 web pages, 121 concordances were generated.

http://www.islandcounty.net/auditor/vendor_list_application.htm
firm's ability to satisfactorily perform a contract:

<http://www.manches.com/dpa/part4.htm>
Is the transfer necessary to perform a contract between you and the individual

<http://www.morgans-law.com/examples.htm>
of the buyer's failure to perform a contract to scrap a ship. 7

<http://www.discoveraz.com/buyers/earnmoney.html>
show ability and intent to perform a contract. In real estate, it represents

<http://www.carleton.ca/law/outlines/s00/203v-s00.htm>
legal grounds for failing to perform a contract, in Chapter 12); and then

<http://isd.co.la.ca.us/pcs/purchasing/bids/pcsterms.cfm>
quality, fitness or capacity to perform a contract with the County or any

<http://cobrands.smallbiz.findlaw.com/legal/contracts/faqs.htm>
court order a party to perform a contract rather than pay damages? Is

<http://www.judiciary.state.nj.us/charges/civil/419.htm>
that whether seller's refusal to perform a contract for the sale of a

<http://www.sba.gov/regulations/testimony/archive/ballentine10-21-99.html>
as the financial capability to perform a contract, plus other commitments. In 1977

http://www.buyer-agent.net/why_use_an_EBA.html
ability and/or willingness to perform a contract to acquire seller's property that

<http://biz.findlaw.com/legal/contracts/faqs.html>
court order a party to perform a contract rather than pay damages? Is

<http://www.sportslawnews.com/archive/history/HT1.html>
one cannot be forced to perform a contract against his will, so the

Crop, 10, 2004

<http://netec.mcc.ac.uk/WoPEc/data/Papers/zuriewwp052.html>

monotonic impact on behavior. Individuals perform a contract when enforcement is strong or

<http://courtlink.utcourts.gov/opinions/supopin/aquagen.htm>

compelling Calrae to continue to perform a contract that is unenforceable

<http://www.pwgsc.gc.ca/sos/corporate/sm/text/ch11c-01-e.html>

When a contractor fails to perform a contract, or when a claim is

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

Apêndice 6 – *WebCorp* – Exemplos com a colocação ‘fazer valer o contrato’ em diferentes contextos

Using the Google search engine: accessed 7 web pages
5 concordances were generated.

<http://www.jornalopcao.com.br/destaques.asp?idjornal=94>

podem recorrer à Justiça para fazer valer o contrato assinado nas urnas. E ao

<http://www.estado.estadao.com.br/editorias/2002/02/14/eco043.html>

presidente da Transbrasil e quer fazer valer o contrato, assinado em janeiro, que lhe

<http://www.estado.estadao.com.br/suplementos/agri/2002/03/13/agri029.html>

o há mecanismo legal para fazer valer o contrato em seu rompimento.

<http://www.sbda.org.br/revista/Anterior/1661.htm>

termo (meio acessório de fazer valer o contrato principal), mas de direitos reais

<http://www.mj.gov.br/dpdc/clipping/clip1909.htm>

Mallet, o único meio para fazer valer o contrato e receber a indenização

Crop, 10, 2004

Apêndice 7 – *WebCorp* – Exemplos de ocorrências da colocação ‘cumprir contrato’

Using the Google search engine: accessed 7 web pages
32 concordances were generated.

http://www.sfiec.org.br/artigos/energia/Na_energia_ningu%27E9m_se_salva_sem_perdas.htm

contrato nenhum. Cobrar dívida é cumprir contrato. Que quebra de contrato é

<http://www.futebolinterior.com.br/news/newsclube.php?nid=43485>

hoje a minha idéia é cumprir contrato até o final”, afirmou. O

<http://www.radialistasp.org.br/juridico.htm>

O radialista não terá que cumprir contrato de experiência nos casos de

http://www.pellon-associados.com.br/boletim/boletim_mai02/Not%20EDcias%20Imprensa%20mai%202002.htm

exclui obrigação da seguradora de cumprir contrato A Quarta Turma do Superior

http://www2.correioweb.com.br/cw/EDICAO_20030605/pri_esp_050603_186.htm

no Olímpico. Sua idéia é cumprir contrato com o Fluminense. Editor: Paulo

http://www.stj.gov.br/webstj/Noticias/detalhes_noticias.asp?seq_noticia=8016

mulher de endossante condenado a cumprir contrato A Quarta Turma do Superior

<http://jessicamenezes.turmadobar.com.br/index.asp?op=9>

Aerosmith lançou pela Geffen, para cumprir contrato, a coletânea “Big Ones”. Na

<http://www.esportes.terra.com.br/esportes/2002/04/18/166.htm>

gente ÚLTIMAS NOTÍCIAS Dodô vai cumprir contrato com o Bota, diz procurador

<http://www.soccerage.com/pt/22/00068.html>

seleção da França Rivaldo quer cumprir contrato com o Barcelona até o

<http://www.espacovital.com.br/obrigacoes.htm>

para reparo desobriga inquilino a cumprir contrato Editora Globo pagará passagens não

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

<http://www.anfip.org.br/linha-direta/lin1503.php>

março. Justiça determina Unimed-SJRP a cumprir contrato A ANFIP impetrou ação judicial

<http://www.geocities.com/alefadi/his5.html>

que só o fez para cumprir contrato. Roland seguiu na turnê “Elemental

<http://www.todapalavra.jor.br/edicoes/05/musica.htm>

é um disco burocrático, para cumprir contrato. E não são poucas as

<http://www.sapereaudare.hpg.ig.com.br/esportes/texto05.html>

o Fiore Gigliotti) sem nunca cumprir contrato in totum com todos eles

<http://www.hyldon.com.br/Lancamentos.htm>

de “disco secreto”, lançado para cumprir contrato e sem divulgação. Hyldon, um

<http://www.estadao.com.br/divirtase/noticias/2001/nov/04/12.htm>

deixar a Globo. “Vou só cumprir contrato nesses dois meses que restam

<http://pt04.net/euro2004/euro2004-figuras.php?figuraid=390>

Turquia 2003-09-27 Euro2004 Eriksson vai cumprir contrato até 2006 2003-09-04 Inglaterra As escolhas

http://www.animegaiden.com.br/blogs/cavaleiro/archives/2003_04.html

um piloto que tem que cumprir contrato, não sei a Willians forçou

http://5minutos.blogs.sapo.pt/arquivo/2003_06.html

Adriana Calcanhoto : “Casa Cheia” = cumprir contrato e se não gostarem podem

<http://www.samba-choro.com.br/s-c/tribuna/samba-choro.9808/0470.html>

então LP “Chico Canta”, para cumprir contrato com sua gravadora. O LP

<http://www.portalf1.com/noticias/noticia.asp?noticia=1103>

está acabado e que irá cumprir contrato até 2006. Vão ter de

<http://www.florestanegra.com.br/pt/noticias/index.php?cdnoticia=118>

Parque São Jorge e prometeu cumprir contrato até o final do ano

<http://www.abcnatal.com.br/resenha45.htm>

me indicou ao ABC. Vim cumprir contrato até o final de 73

Crop, 10, 2004

<http://www.londoncalling.com.br/scripts/destaque.asp?ID=102>
apenas como um álbum a cumprir contrato, já que Neil Young estava

<http://www.caetenews.com.br/caetur/bomtempo001.html>
de pacote que terá que cumprir contrato e roteiro da viagem, mas

<http://www.topgol.com.br/competicoes/brasileiro/estaduais/paraiba/noticias.htm>
do jogador que terá de cumprir contrato até o final de agosto

<http://www.esportes.terra.com.br/2003/09/30/108.htm>
de São Januário Vampeta promete cumprir contrato e admite renovar por dois

<http://www.an.com.br/1999/mai/04/0esp.htm>
de novo técnico Parreira pretende cumprir contrato com o Flu. Atletas fazem

<http://www.oparaense.com/esporte-17.htm>
Belém, outra vez, depois de cumprir contrato com o Remo; no entanto

<http://www.macua.com/biografias/domingosmarques.html>
papéis para tenores e, após cumprir contrato de quatro anos no Maxime

<http://www.rudeawakening.hpg.ig.com.br/historia.htm>
banda estivesse acabando, ou fosse cumprir contrato com a gravadora. Infelizmente aconteceu

<http://modernosdescobrimientos.inf.br/desc/lobato/lobatonotempo1937.htm>
requerera passaporte a fim de cumprir contrato comercial com uma Empresa editora

Apêndice 8 – *Kwic Finder* – Exemplos de ocorrências da colocação ‘cumprir contrato’

<http://www.estado.com.br/edicao/pano/98/07/30/fut538.html>

Original

Query term 1, “cumpr* contrato”, matches: 1

Sexta-feira, 31 de julho de 1998 Irmãos De Boer terão de	cumprir contrato	ZEIST - Os irmãos Frank e Ronald de Boer perderam, ontem, a apelação feita para romper o contrato com o Ajax, clube que defendem. O comitê de arbitragem da Associação Holandesa de Futebol alegou que os gêmeos, de 27 anos, não apresentaram argumentos consistentes para quebrar o contrato de seis anos.
---	---------------------	--

<http://www2.estado.com.br/jornal/suplem/tele/99/05/16/tele012.html>

Original

Query term 1, “cumpr* contrato”, matches: 2

Tiazinha. Uma emissora convidar um profissional que se destaca em outra é regra do mercado. Assim fazem os times de futebol e empresas de todas as áreas. A Abert (Associação Brasileira de Emissoras de Rádio e TV) tem um código de ética pelo qual uma empresa não pode convidar quem	cumpre contrato	em outra. Flávio Cavalcanti Jr., vice-presidente de Televisão da Abert, reconhece que a questão é delicada. “Há um pacto entre as redes em respeitar contratos, mas não se pode impedir o crescimento profissional das pessoas”, diz. “As multas contratuais impõem limites, mas a desigualdade no poder financeiro dos canais
Angel Mix foi ao ar com Mudanças que não agradaram à direção da líder. Seu empresário, Marcos Saraiva, diz que ouviu de José Paulo Vallone propostas com mundos e fundos. “Angélica está feliz, mas não há motivo para não ouvir uma proposta interessante”, diz Saraiva. A	cumprirá contrato	até 2002. “Há vários interesses em jogo e nada será decidido a curto prazo”, insiste Saraiva. Astrid Fontenelle diz que vive atormentada com notícias sobre convites que recebeu da Globo e da Band. “Detesto isso, parece que você está leiloando seu salário”, diz. Ela afirma que foi procurada por um

Crop, 10, 2004

http://www.jurismidia.com/bdnot_03.htm

Original

Query term 1, "cumpr* contrato", matches: 2

Iguatemi Shopping de devolver alugueis reajustados durante Plano Real Viuva de avalista não é obrigada a pagar débito de locação Cláusula limitando obrigações do plano de saúde deve ter destaque STJ condena União a pagar 200 mil pinheiros por não	cumprir contrato	assinado há meio século STJ reconhece possibilidade de revisão de cláusula de contrato já quitado STJ: Posto deverá indenizar a Esso por descumprir contrato Segunda Seção aprova súmula sobre cobrança de direitos autorais.
--	------------------	---

<http://www.anoregpr.org.br/noticias.htm>

Original

Query term 1, "cumpr* contrato", matches: 1

22/05 - Dirigir na contramão não exclui obrigação da seguradora de	cumprir contrato	22/05 – STJ nega habeas corpus a advogado acusado de cometer denúncia caluniosa 22/05 – Nilson Naves- decisões do Judiciário serão publicadas on line no Diário da Justiça.
--	------------------	---

Apêndice 9 – *WordSmith Tools* – Palavras que co-ocorrem com *attention*

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	ATTENTION	195	5	2	1	1	1	1	1	188	2	0	0	0	0
2	THE	129	55	74	12	7	9	12	15	0	1	38	10	8	17
3	AND	41	22	19	8	1	5	4	4	0	2	2	2	7	6
4	THAT	26	14	12	3	5	4	1	1	0	3	1	3	3	2
5	FOR	13	8	5	1	0	1	3	3	0	3	0	0	1	1
6	FROM	13	1	12	0	0	0	1	0	0	10	1	0	0	1
7	HAS	13	8	5	1	0	4	3	0	0	2	0	1	1	1
8	HAVE	13	10	3	3	1	4	2	0	0	0	1	0	1	1
9	HIS	13	6	7	0	2	0	1	3	0	0	3	2	1	1
10	ARE	12	5	7	1	0	3	1	0	0	0	0	1	2	4
11	WAS	12	4	8	1	1	2	0	0	0	1	1	0	4	2
12	WITH	11	7	4	1	1	0	5	0	0	1	0	0	1	2
13	ATTRACTED	10	10	0	1	1	1	5	2	0	0	0	0	0	0
14	BUT	10	2	8	1	1	0	0	0	0	0	4	3	1	0
15	MORE	10	7	3	1	0	0	0	6	0	1	1	0	1	0
16	PAY	9	8	1	0	0	0	4	4	0	0	0	0	1	0
17	BECAUSE	8	1	7	1	0	0	0	0	0	3	1	3	0	0
18	ITS	8	5	3	1	0	0	0	4	0	1	0	1	0	1
19	SAID	8	0	8	0	0	0	0	0	0	0	5	3	0	0
20	MEDIA	7	6	1	0	0	3	0	3	0	0	0	0	1	0
21	THEY	7	2	5	0	2	0	0	0	0	0	1	1	2	1
22	THIS	7	2	5	0	0	1	0	1	0	2	3	0	0	0
23	DRAW	6	6	0	0	1	0	2	3	0	0	0	0	0	0
24	GET	6	3	3	0	0	0	2	1	0	0	1	0	1	1
25	PAID	6	5	1	0	1	0	4	0	0	1	0	0	0	0
26	SOME	6	2	4	2	0	0	0	0	0	0	2	0	1	1
27	DRAWN	5	4	1	0	0	0	3	1	0	0	0	1	0	0
28	FIRST	5	3	2	2	0	1	0	0	0	1	0	0	1	0
29	GETTING	5	2	3	2	0	0	0	0	0	2	0	1	0	0
30	LOT	5	5	0	0	0	1	4	0	0	0	0	0	0	0
31	MUCH	5	5	0	0	0	0	0	5	0	0	0	0	0	0
32	NOT	5	4	1	1	1	1	1	0	0	0	0	0	0	1
33	SHOULD	5	4	1	1	0	3	0	0	0	0	0	0	0	1
34	SPECIAL	5	5	0	0	0	0	0	5	0	0	0	0	0	0
35	THEIR	5	4	1	0	1	0	0	3	0	0	0	0	1	0

Crop, 10, 2004

Apêndice 10 – *WordSmith Tools* – Concordância em contexto geral e em contextos específicos com *attention*

Draw Attention

N	Concordance
1 .	"The higher the fee, the more attention the recruiter will pay
2 4) highlights the need to draw	attention to the fact that the Police Commission
3 , "Doctor in court" (January 19), draws	attention sadly but truthfully, to the "battered m
4 ference last year in Berlin that drew critical	attention to Iran. The two cases became lin
5 ocracy in America, she pays obsessive	attention to what Americans wear, what they ea
6 cribed in psychological and physical detail.	Attention is given to her personality as well as h
7 supported Internet providers have attracted more	attention many small, local provid
8 e was invited to help draw international	attention to the Festival of Britain.
9 better or worse, his office isn't able to pay full	attention to the fountain. "It's our property.
10 ting.	Only last week Peter Riddell drew attention to polling evidence of
11 tion on lawyers that has drawn particular	attention in the profession is a two-year-
12 t, Democratic leaders worked hard to draw	attention to Greenspan's warnings that long-ter
13 P, Ms Johnson, 44,	attracted attention for her unswerving loyalty to the
14 complained that Zhao 'never paid a whit of	attention to people like us. . . . What he re
15 record amounts of funding and attracted national	attention with the mapping of the human g
16 the vagaries of male affections	- a little attention-getter we can only hope will be limited
17 e which the Office of Fair Trading drew	attention to in its recent report on the industry.
18 ired Saturday. Rich's pardon has attracted	attention in part because it did not go thro
19 ut scandal-free. Mr King draws	attention to "doubts over the integrity" of at
20 entury. On the scientific side, he gives Jue	attention to Einstein's contribution to the develo
21 strous, so some Uruguayans 'aven't paid much	attention And because the president insis
22 wever, hat attracted the most positive	attention from the Jordanian media and
23 Other spats have the effect of both drawing	attention to an important research question whil
24 from the mainstream, it attracted serious	attention from the surrealists, and in p
25 Rangers forward escaped the	attention of Thuram and Lassissi to meet Reyna

Call Attention

N	Concordance
1 as "a great place to begin."	Bush called attention to his shared understanding with the De
2 rights was unimpeachable. Calling	attention to her own race and g

Pay Attention

N	Concordance
1 ore Euro-supportive ind ess Francophobe,	paid much attention to this baton change. The Financial Times publish
2 with a knowing and admonitory eye, and invites us to pay	attention have a care, take heed, live life as if it matter
3 Liverpool • Nancy Banks-Smith should pay closer	attention (Going underground, December 27). If she h
4 is why Brown and co are paying	attention. The chance
5 ore than game film. "That's something to pay	attention to," McNabb said of the weather, "b
6 today," Yary said. "This was the first time I	paid close attention to this. . . . I didn't know if I was deserving. I
7 cker searches and generally increase the time and	attention paid to student safety. The Education De
8 in a pool of blood. The gunman seemed	to pay no attention. As his first targets scrambled
9 candidates.	"The higher the fee, the more attention the recruiter will pay to the assi
10 lead of democracy in America, she pays obsessive	attention to what Americans wear, what they eat, the bizar
11 that for better or worse, his office sn't able to pay full	attention to the fountain. "It's our property. It's our fou
12 n't pay much	attention to her last
13 lder Wang complained that Zhao 'never paid a whit of	attention to people like us. . . . What he really wants
14 but not monstrous, so some Uruguayans 'aven't paid much	attention. And because the president insists that his
15 but not monstrous, so some Uruguayans 'aven't paid much	attention. And because the president insists that his
16 the warnings.	"I try to make my daughter pay attention to the signs, but with a 9-year-old or any- 17 ndow, playing with the venetian blinds, and not paying attention. So she took the boy by the arm and told hi 18 sic, watches television, is a voracious reader and is paying attention to the new wave of American women, who will pla 19 debated in Congress, where the nation is paying attention to it," spokesman David Bernstein said. 20 re that sounds are parts of words, to follow commands, pay attention and express ideas in a story line. Dr. Tallal

Give Attention

N	Concordance
1 ming the neat balancing act of giving due	attention to both life and science. But the devil
2 specially impressed by the	special attention given to Frank and Lucy's anniversary
3 mey-general, failed to give full	attention to the investigation
4 e given environmental issues	scant attention in Texas, he is clearly not Mr. Reaga
5 ribed in psychological and physical detail.	Attention is given to her personality as well as
6 ntury. On the scientific side, he gives due	attention to Einstein's contribution to the develo
7 Newcastle upon Tyne • It is timely that	attention is being given to the implications of

Draw Attention

N	Concordance
1 ercial aircraft, that drew the most	attention from Mr. Mbeki — and where the big
2 its material and to the point. Our	attention first is drawn to the artist's general u
3 last election, and McCain's success in drawing	attention to the issue during the campaig
4 create an arts scene that will draw	attention back up to their neighborhood, this
5 d at least have drawn special	attention to the prophetic, al
6 medical care here, and drawn the	attention of several Russian cities hoping to
7 ions of dollars. But what has drawn	attention to Al-Ribat is not Muhammad bin
8) highlights the need to draw	attention to the fact that the Police Commissio
9 rence last year in Berlin that drew critical	attention to Iran. The two cases became l
10 "Doctor in court" (January 19), draws	attention, sadly but truthfully, to the "battered
11 was invited to help draw international	attention to the Festival of Britain.
12 g. Only last week Peter Riddell drew	attention to polling evidence of huge
13 on on lawyers that has drawn particular	attention in the profession is a two-yea
14 Democratic leaders worked hard to draw	attention to Greenspan's warnings that long-ter
15 which the Office of Fair Trading drew	attention to in its recent report on the industry.
16 scandal-free. Mr King draws	attention to "doubts over the integrity" of at
17 Other spats have the effect of both drawing	attention to an important research question wh
18 short while to draw the art world's	attention The spotlight shifted whe
19 tyle that serves Battle well at home and draws	attention overseas, said Doyenart, an oc
20 tyle that serves Battle well at home and draws	attention overseas, said Doyenart, an oc
21 , last week), I would like to draw	peoples' attention to the fact that it was not an accusati

Crop, 10, 2004

Apêndice 11 – *Corpus Nilc* – Concordância das colocações mais comuns em português com *atenção*

O corpus NILC Universidade de São Carlos, contendo textos brasileiros do registo jornalístico, didáctico, epistolar e redacções de alunos, encontrava-se em vários ficheiros em formato de texto, numa estrutura de directorias indicando a fonte.

O CETENFolha (Corpus de Extractos de Textos Electrónicos NILC/Folha de S. Paulo) é um corpus de cerca de 24 milhões de palavras em português brasileiro, criado pelo projecto Processamento computacional do português com base nos textos do jornal *Folha de S. Paulo* que fazem parte do corpus NILC/São Carlos, compilado pelo Núcleo Interinstitucional de Linguística Computacional (NILC).

Procura: “**atenção**”.

Pedido de uma concordância em contexto

Corpus: NILC/São Carlos v. 7.0

3093 ocorrências.

Resultados da procura

Thu Dec 12 23:48:05 CET 2002

Alguns exemplos:

ATRAIR ATENÇÃO

Procura: “**atrair**” “**a**” “**atenção**”.

Pedido de uma concordância em contexto

Corpus: NILC/São Carlos v. 7.0

25 ocorrências.

Concordância

Procura: “**atrair**” “**a**” “**atenção**”.

par 12125: «O julgamento vai **atrair a atenção** internacional, e não vai interessar aos poderosos», assinala a artista plástica, que no julgamento vai contar novos detalhes da matança .

par 23974: Outros 140 outdoors foram espalhados em várias capitais para **atrair a atenção** dos turistas .

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

par Agrofolha-94a-agr-1: José Artur de Abreu Martins, da Secretaria de Agricultura do RS, diz que a raça está em «franco desenvolvimento e deve **atrair a atenção** dos criadores que visitarem a exposição» .

par Brasil-94a-pol-1: Na era da TV e do computador, o elemento gráfico é fundamental para **atrair a atenção** de crianças e adolescentes .

CHAMAR ATENÇÃO

Resultados da procura

Thu Dec 12 23:56:38 CET 2002

Procura: “**chamar**” “**a**” “**atenção**”.

Pedido de uma concordância em contexto

Corpus: NILC/São Carlos v. 7.0

219 ocorrências.

Procura: “**chamar**” “**atenção**”.

par 18048: Em nome da discricção, «para não **chamar atenção**», pediu oito ingressos, para hoje, nas mesas 701 e 702, ao lado da mesa de som .

par 31500: Na cidade de Curnavaca, o bispo Luis Reynoso, assessor jurídico do episcopado mexicano, pediu em homilia a perseguição a esse grupo, «integrado por pessoas anormais, que procuram **chamar atenção** sobre si» .

par 132394: Esse estado de espírito geral ficou claro, quando uma despretenhosa emenda apresentada por um deputado do PMDB de Goiás, Dante de Oliveira, em março de 1983, passou a **chamar atenção** do público, sobretudo após às declarações favoráveis do Cardeal Paulo Evaristo Arns e de Don Ivo Lorscheiter, secretário geral da CNBB .

par Brasil-94a-pol-1: FHC passará o réveillon em Brasília só com a família para não **chamar atenção** .

DAR ATENÇÃO

Procura: “**dar**” “**atenção**”.

Pedido de uma concordância em contexto

Corpus: NILC/São Carlos v. 7.0

31 ocorrências.

Crop, 10, 2004

par 11340: Muito assustado, Leonardo teria corrido para pular o muro da casa, sem lhe **dar atenção** .

par 22323: Mas não posso simplesmente dizer que não estou com vontade de **dar atenção** .

par 22529: Mas, prefiro não **dar atenção** a essas cobranças .

par 40364: Resolvemos, hoje que a nossa obra está firme, sair da senda que sempre seguimos de não **dar atenção** às mais pérfidas das dificuldades por que tivemos de passar .

PRESTAR ATENÇÃO

Pedido de uma concordância em contexto
Corpus: NILC/São Carlos v. 7.0
103 ocorrências.

par 8608: A doença sinaliza que tem que **prestar atenção** em você», apregoa o escultor.

par 8776: Percília Gomes Soares, de 35 anos, que ensina Ciências para as turmas da quinta série, contou que, em vez de **prestar atenção**, os alunos conversavam sobre o acidente .

par 10447: Além disso, quem for pegar táxis em direção ao Sambódromo deve **prestar atenção** aos pontos criados pela Secretaria Municipal de Transportes Urbanos (SMTU) nas imediações do Sambódromo e ao preço das tarifas estabelecidas para as corridas, bairro por bairro .

par 12128: Perigo – Para Ivone, existe outro perigo que o Ministério Público e a polícia devem **prestar atenção**: no ano passado, Vágner reconheceu mais quatro policiais que participaram da chacina .

WebCorp – Concordância das colocações mais comuns em português com atenção

http://www.gnosis.org.br/sawpage/livros/educ_fundamen/pcap4.htm

Document Dated: Tue, 28 Nov 2000 02:12:54 GMT

[Plain Text](#) [Word List](#) 1853 tokens, 12 types

- ensina aos estudantes que devem prestar atenção durante a aula, e os

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

- elaborado pelo organismo. Se soubermos prestar atenção conscientemente, poderemos economizar energia criadora
- da consciência. Há que saber prestar atenção sem se identificar. Quando prestamos
- Os alunos e alunas devem prestar atenção nas aulas sem se esquecerem
- não ensina os estudantes a prestar atenção conscientemente. A disciplina é uma

Statistics

Using the Google search engine WebCorp accessed 97 web pages, 7 of which returned errors.

203 concordances were generated.

<http://www.bdt.fat.org.br/publicacoes/padct/bio/cap10/elopla.html>

Document Dated: Unknown

Plain Text Word List 3281 tokens, 46 types

- mas que eventualmente começam a atrair atenção do exterior Lapachol 1.8
- e que eventualmente começam a atrair atenção do exterior Andiroba, óleo da

Statistics

Using the Google search engine WebCorp accessed 95 web pages, 5 of which returned errors.

83 concordances were generated.

<http://www.umaformadeamar.com.br/comportamentos/comportamentoF01.htm>

Document Dated: Wed, 11 Sep 2002 11:16:15 GMT

Plain Text Word List 594 tokens, 13 types

- as beldades. Você pode inclusive chamar atenção para um corpo bonito. 11-

<http://www.maisprojetos.com.br/mktverde/oquee.htm>

Document Dated: Thu, 28 Nov 2002 22:30:56 GMT

Plain Text Word List 1803 tokens, 32 types

- forma criativa que desenvolvemos para chamar atenção dos vários segmentos da sociedade

Statistics

Using the Google search engine WebCorp accessed 95 web pages, 5 of which returned errors.

120 concordances were generated.

Crop, 10, 2004

<http://www.oruminante.com.br/LstTxt.asp?CodTxt=85>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 752 tokens, 21 types

- mulheres vêm se esquecendo de dar atenção a essa parte tão importante

<http://www.evirt.com.br/artigos/antonio03.htm>

Document Dated: Wed, 29 May 2002 00:42:32 GMT

[Plain Text](#) [Word List](#) 1934 tokens, 21 types

- de contentamento. Esse doar-se significa dar atenção ao outro, dar estímulos positivos

Statistics

Using the [Google](#) search engine WebCorp accessed 93 web pages, 2 of which returned errors.

120 concordances were generated.

<http://www.uol.com.br/diariodovale/arquivo/2001/marco/02/page/fr-esporte4.htm>

Document Dated: Thu, 01 Mar 2001 23:41:41 GMT

[Plain Text](#) [Word List](#) 1269 tokens, 17 types

- o treino da manhã, preferem concentrar atenção nos astros do momento. Roberto

<http://www.ergon.com.br/opodercap1.htm>

Document Dated: Mon, 02 Dec 2002 15:51:44 GMT

[Plain Text](#) [Word List](#) 1236 tokens, 8 types

- pode, por exemplo, ajudar a concentrar atenção e energia. Mas a pressão

http://www.sebrae-sc.com.br/newart/mostrar_materia.asp?cd_noticia=4433

Document Dated: Unknown

[Plain Text](#) [Word List](#) 763 tokens, 18 types

- pontos críticos do negócio e concentrar atenção na superação de seus entraves

http://www2.correioweb.com.br/cw/EDICAO_20020509/pri_eco_090502_165.htm

Document Dated: Unknown

[Plain Text](#) [Word List](#) 839 tokens, 15 types

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

- Brian Dyson. A decisão de concentrar atenção no maior país da América Statistics

Using the Google search engine WebCorp accessed 39 web pages, 2 of which returned errors.

33 concordances were generated

Crop, 10, 2004

Apêndice 12 – *Cobuild* – Combinação LAVISH + ATTENTION

Cobuild

Query Results

NOTE: no more than 40 lines will be displayed here, since a threshold has been implemented. If there were more than 40 instances found, a random selection will have been applied.

[p] Clearly a wise parent won't lavish attention and toys on one child and works of the Western tradition to lavish attention on material that is the naval high command continued to lavish attention on this apparently vital

Corpus Jornalístico CETRAD

Suggested Retail \$ 6,800.00

Sale Price \$ 5,440.00

Diplomat

Model #65325 (blue ocean)

Model #65225 (antique ocean)

Height 48", Diameter 32",

Weight 122 lbs.

A handcrafted masterpiece, this extraordinary globe features lavish attention to both form and function. Distinctive details include touch-on illumination, rich, ten-color cartography and a solid brass, hand-engraved meridian. The hand-carved, genuine mahogany cradle mounting rubbed to a lustrous walnut-finish further emphasizes the Diplomat's uncommon beauty. This globe features over 20,000 place names.

Parte inferior do formulário

Parte superior do formulário

#65225 antique ocean ▼

Parte inferior do formulário

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

WebCorp

<http://www.globeworld.com/globesfloor32.html>

Document Dated: Wed, 04 Dec 2002 21:38:29 GMT

[Plain Text](#) [Word List](#) 291 tokens, 14 types

- masterpiece, this extraordinary globe features [lavish attention](#) to both form and function

http://gta.nitshade.net/gta3_story.shtml

Document Dated: Unknown

[Plain Text](#) [Word List](#) 483 tokens, 24 types

- and as EGM said “[the] [lavish attention](#) to detail [is] on par

<http://anyhotelanywhere.com/findhotel.cfm/172789>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 1154 tokens, 18 types

- Inn was created with a [lavish attention](#) to detail. Additional Pictures: Enlarge
- Inn was created with a [lavish attention](#) to detail. This award-winning property

<http://www.bbonline.com/ga/planters/index.html>

Document Dated: Sat, 21 Sep 2002 12:13:26 GMT

[Plain Text](#) [Word List](#) 252 tokens, 16 types

- Inn was created with a [lavish attention](#) to detail. This award-winning property

<http://www.jmrads.com/Playback.html>

Document Dated: Wed, 03 Apr 2002 23:03:48 GMT

[Plain Text](#) [Word List](#) 683 tokens, 17 types

- Playback Equipment Although we do [lavish attention](#) and expense on our CDs’

Crop, 10, 2004

<http://www.innsite.com/inns/A003846.html>

Document Dated: Unknown

Plain Text Word List 423 tokens, 25 types

- are nearby. Enjoy the hosts' lavish attention to detail, from homemade hazelnut

Statistics

Using the AltaVista search engine WebCorp accessed 120 web pages, 9 of which returned errors.

117 concordances were generated.

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

Apêndice 13 – DICIONÁRIOS *ACTIVATOR* E *HOUAISS* – SOLUÇÕES

Apenas no dicionário *Activator* temos alguma orientação sobre o significado individual das diferentes colocações:

Pay attention: to listen to and watch carefully what you are doing, what is happening, or what someone is saying. Ex.: ***John never pays attention in class.***

Draw attention to: to tell people about a problem, unpleasant subject etc that you think they should pay attention to.

His article drew attention to the lack of funding in modern scientific research

Call attention to: to make people conscious of a bad problem or situation that you think people should pay attention to so that they can help:

The only way to raise money is by calling attention to the famine in Africa

No *Houaiss* temos também algumas pistas sobre as colocações:

Chamar a atenção: dar na vista; ser visível ou vistoso

Chamar a atenção de: fazer advertência, repreender

Chamar a atenção para/sobre: colocar em destaque, focalizar, mostrar, alertar

Dar atenção a: ouvir e tratar com respeito e consideração

Prestar atenção a/em: olhar, ouvir, sentir algo com atenção aumentada, concentrada

Apêndice 14 – COMPARA – SOLUÇÕES DE TRADUÇÃO PARA O PORTUGUÊS DE COLOCAÇÕES COM *ATTENTION*

PBMA 2(156):	Nobody pays any attention to me anymore.	Ninguém mais repara em mim.
PBMA 2(160):	Virgília sensed this and told her son: «N honhô, don't pay any attention to that big trickster there. He doesn't want to talk so he can make you think that he's at death's door.»	– N honhô, não repara nesse grande manhoso que aí está; não quer falar para fazer crer que está à morte.
EBDL3T 2(937):	Há alguns detalhes para os quais gostaria de chamar a sua atenção .	There are a few points you might like to bear in mind .
EBJB1(1041):	Comia o que lhe levavam, mas não dava atenção a ninguém.	He would eat whatever food was brought him, but took no notice of anyone.
EBJB1(1082):	Um poleiro de papagaio chama-nos a atenção .	A parrot's perch catches the eye .
EBJT 1(731):	Dizia que podia ouvir isso durante o dia inteiro, no Land Rover, no tractor, que a hora das refeições era para se comunicar, para darem atenção uns aos outros, para conversar.	She said he could hear that all day, in the Land Rover, on the tractor, that mealtimes were for communication, for paying attention to one another, for talking.
EBJT 1(1995):	– Não estava a prestar muito atenção .	I only half was .
PBMA 3(591):	Mamãe é boa demais; dá-lhe atenção demais.	Mamma's too good; she pays him too much attention .
PBPM1(868):	Wilmer, atenção , a sinopse é longa.	Wilmer, pay attention , this is a long outline.
PBPM1(1005):	... ele perguntou, é a terceira vez que você me fala dessa raquete, ela disse, você não presta atenção no que eu digo.	... ' he said. ' It's the third time you've asked about the racket, ' she said. ' You don't listen when I talk. '

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbiais*

PBRF1(1973):	O baile do Scala estaria cheio de mulheres nuas, Liliansa sabia que o recato chamaria mais atenção do que a nudez.	The Scala ball would be full of nude women. Liliansa knew that modesty would attract more attention than nudity.
EBDL1T1(1797):	A bri o dicionário para confirmar a ortografia da palavra <i>perscrutar</i> e ao folhear as páginas, a palavra <i>pó-de-dover</i> chamou-me a atenção .	I just opened the dictionary to check the spelling of «glowered», and as I flipped the pages the headword «Dover's Powder» caught my eye .
EBDL3T1(106):	Raras vezes terminara a bibliografia preliminar, sem que algo novo e totalmente diferente viesse chamar-lhe a atenção .	Seldom, indeed, had he drawn up a preliminary bibliography before his attention was distracted by some new or revived interest in something entirely different.
EBDL1T2(2134):	Não me lembro de ter dado atenção especial essa cláusula na época, e	I can't recall giving this clause any special thought at the time, but I'm not surprised that I agreed to it.
PBMA1(410):	He felt the urge to stay, to shine where he'd been in the shadows, to get one up on the people who 'd paid scant attention to him before, mostly the ones who 'd laughed at his friendship with Quincas Borba.	Sentia cócegas de ficar, de brilhar onde escurecia, de quebrar a castanha na boca aos que antes faziam pouco caso dele, e principalmente aos que se riram da amizade do Quincas Borba.
PPMC1(228):	And it wasn't just because of a lack of opportunity, given Mara's close attention ;	E não foi apenas por falta de oportunidade, devido à atenção cerrada de Mara.

Crop, 10, 2004

Apêndice 15 – AMBIENTE SEMÂNTICO: RECORRÊNCIA DE *CLOSE ATTENTION* E *SCANT ATTENTION*

WordSmith Tools:

N	Concordance
1," Yary said. "This was the first time I	paid close attention to this. . . . I didn't know if I was deserving. I

N	Concordance
1 ave given environmental issues	scant attention in Texas, he is clearly not Mr. Reagan, t

WebCorp

<http://www.pdinet.org/sgo/v16no4/doctrine.html>

Document Dated: Fri, 11 May 2001 20:00:17 GMT

Plain Text Word List 1910 tokens, 27 types

- to be reminded to pay close attention to his life and doctrine
- Christians and Christian leaders paying close attention to today? They're paying close
- attention to today? They're paying close attention to dreams, visions, and prophecy
- unusual physical manifestations. They're paying close attention to trends, marketing techniques, and
- of church growth. They're paying close attention to the size of their
- staff, and building. They're paying close attention to sociology, psychology, and psychiatry
- psychology, and psychiatry. They're paying close attention to the latest hot Christian
- they are not paying equally close attention either to their lives or
- grace to help us pay close attention whenever we are faced with
- or to glorify God? Pay close attention to these things. Don't overestimate
- wage that battle wisely. Pay close attention to your words. Listen to
- how we say it. Pay close attention to your actions, especially when

Using the Google search engine WebCorp accessed 120 web pages, 12 of which returned errors. 206 concordances were generated.

<http://www.taiwandc.org/twcom/91-no3.htm>

FIKER, Marcia Epstein; FOLEY, Stela. *Os Corpora como Ferramenta para Solução de Problemas de Tradução de Colocações Verbais*

Document Dated: Wed, 03 May 2000 02:44:16 GMT

[Plain Text](#) [Word List](#) 2235 tokens, 29 types

- Chinese nationalists. The West paid scant attention to the plight of the
 - years the world still paid scant attention to the plight of the
-

<http://www.islamicthought.org/pp-zb-leader.html>

Document Dated: Wed, 13 Sep 2000 10:39:44 GMT

[Plain Text](#) [Word List](#) 8721 tokens, 61 types

- re-order Muslim societies [1]. Such scant attention to the fundamental aspects of

Statistics

Using the [Google](#) search engine WebCorp accessed 120 web pages, 5 of which returned errors.

144 concordances were generated.

WebCorp – FAZER POUCO CASO

<http://www.observatoriodaimprensa.com.br/artigos/fd170720022.htm>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 2682 tokens, 31 types

- desinformação ou má-fé, mas, principalmente, fazer pouco caso da inteligência das pessoas. Ora
-

<http://www.naum.com.br/literatura/Livros/Parabolas/19.html>

Document Dated: Sun, 07 Apr 2002 04:11:44 GMT

[Plain Text](#) [Word List](#) 2236 tokens, 21 types

- 1-5. Mas, não se deve fazer pouco caso do pecado. O Senhor nos
-

<http://www.presbiterianismo.com.br/Documento/CMW1.htm>

Document Dated: Mon, 20 May 2002 16:29:18 GMT

[Plain Text](#) [Word List](#) 6108 tokens, 26 types

- nossa fé e consciência; o fazer pouco caso e desprezar a Deus
-

Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Lingüística de *Corpus*

Helmara Febeliana Real de Moraes*

Resumo: *A convencionalidade lingüística ainda é um aspecto pouco explorado por lexicógrafos, pelo menos no que diz respeito ao uso de colocações adverbiais. Por meio da Lingüística de Corpus, apresentaremos alguns resultados que demonstram que a tradução, associada ao uso dessas colocações tanto na língua de partida (inglês) quanto na língua de chegada (português), pode causar problemas para o tradutor desatento a essas unidades de significado várias vezes integrantes de associações de palavras maiores. Além de uma pequena amostra de um glossário bilíngüe de colocações adverbiais baseado em corpus, apresentaremos um panorama geral sobre Lingüística de Corpus e comentaremos as etapas percorridas para a obtenção de nossos resultados: a pesquisa em dicionários, algumas vantagens para uma abordagem baseada e direcionada pelo corpus, os corpora e as ferramentas utilizadas, bem como os critérios estabelecidos para a análise de um dado grupo de colocações adverbiais.*

Palavras-chave: *Lingüística de Corpus; colocações adverbiais; unidades de significado; pesquisa baseada em corpus; pesquisa direcionada pelo corpus.*

* Mestranda do Programa de Estudos Lingüísticos e Literários em Inglês da FFLCH/USP.

Abstract: *Linguistic conventionality is still a topic which is not investigated in depth by lexicographers, at least as far as adverbial collocations are concerned. Through Corpus Linguistics, we are presenting some results which show that the translation (associated to the usage) of such collocations both in the source language (English) and in the target language (Portuguese) may cause problems to the translator who is not attentive to these units of meaning and which, in several situations, are part of extended word associations. Along with a small sample of a corpus-based bilingual adverbial collocations glossary, we will provide an overview on Corpus Linguistics and comment on the steps we have gone through in order to arrive at our results: the research in dictionaries, some advantages of a corpus-based and a corpus-driven approach, the corpora and the tools used, as well as the criteria established for the analysis of a given group of adverbial collocations.*

Keywords: *Corpus Linguistics; adverbial collocations; units of meaning; corpus-based research; corpus-driven research.*

1. Colocações adverbiais: problema para o tradutor? Por quê?

Em muitos casos, a tradução de colocações adverbiais passa despercebida pelo tradutor, principalmente pelo fato de, em geral, essas associações não apresentarem problemas de compreensão. A maior parte dessas colocações é composicional, ou seja, conhecendo o significado das partes conseguimos compreender o todo. No entanto, como ressalta Tagnin (2002/1: 194), a dificuldade encontra-se na produção dessas unidades, uma vez que não foram memorizadas de forma consciente pelo falante. De fato, o que muitas vezes ocorre é que há uma associação na língua de chegada que soa mais natural e é comumente utilizada naquele dado contexto. O tradutor não consegue, no entanto, resgatar aquela forma equivalente. Quando ele depara, por exemplo, com a colocação **lavishly illustrated**, vasculha seus arquivos mentais e, muitas vezes, não chega a **fartamente ilustrado**, forma comumente utilizada em português. Desse modo, acaba utilizando uma expressão semelhante, mas que não reflete, na língua de chegada (neste caso, o português), as mesmas características de uso da colocação na língua de partida. Nesse processo, encontram-se vários aspectos que, devido à própria facilidade de compreensão de uma expressão, podem ser ignorados pelo tradutor, como é o caso da tipologia textual e do contexto, apenas para citar dois pontos importantes na escolha de uma forma equivalente. O emprego

inadequado de uma colocação adverbial pode comprometer tanto a fluência do texto quanto a carga semântica expressa no original, tornando-se um problema para o tradutor muitas vezes preocupado com outros aspectos de sua produção, como é o caso da terminologia em textos técnicos.

Dependendo da colocação empregada (tanto no original quanto na tradução), o autor/tradutor pode estar utilizando o advérbio apenas para enfatizar o adjetivo/verbo que modifica mas, em outros casos, o advérbio apresenta carga semântica específica e não está apenas intensificando o sentido da palavra que modifica. É o caso, por exemplo, da colocação **lie outright**. Por meio da pesquisa em *corpora*, observamos que **mentir descaradamente** é a forma equivalente mais freqüente, mas o tradutor desatento pode, por exemplo, utilizar um advérbio mais geral, como é o caso de **muito**. Dessa forma, chega à associação **mentir muito**, que modifica o sentido do original: **mentir muito** não significa **lie outright**, uma vez que a associação em português atém-se à freqüência da ação e aquela em inglês ao modo como a ação é realizada. Com isso, o tradutor estará comprometendo o sentido expresso pelo texto original. Um outro exemplo é a colocação **sexualmente transmissível**. Ao traduzi-la para o inglês, o tradutor desavisado pode utilizar **sexually transmissible**, principalmente pelo fato de as palavras serem consideradas cognatas e poderem por isso ocorrer primeiro na memória do tradutor. Apesar de gramaticalmente correta e utilizada pelos falantes de língua inglesa, **sexually transmissible** não equivale, em termos de freqüência de uso, à colocação em português. Nesse caso, a tradução mais adequada seria **sexually transmitted**, utilizada na língua de chegada com a mesma freqüência encontrada na língua de partida. Inclusive, ambas as colocações fazem parte de unidades de significado maiores consideradas equivalentes: **doenças sexualmente transmissíveis**, cuja sigla é DST, e **sexually transmitted diseases**, para a qual encontramos STD.

Face a essas dificuldades, justifica-se a elaboração de um glossário de colocações adverbiais bilíngüe cujo objetivo é agilizar o trabalho do tradutor, que com uma simples consulta tem acesso a informações resgatáveis apenas por meio da pesquisa baseada em *corpus*. Dessa forma, ele economiza um tempo tão precioso para a realização de seu trabalho.

2. A pesquisa em dicionários

Antes de iniciarmos a pesquisa em *corpus*, decidimos verificar como são abordadas as colocações adverbiais em alguns dicionários atualmente disponíveis, tanto em inglês quanto em português. As conclusões às quais chegamos realmente justificam a necessidade de uma investigação lingüística baseada em *corpus*, como podemos observar nos seguintes resultados:

REAL DE MORAES, Helmara Febeliana. *Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Lingüística...*

- apenas um único dicionário (Macmillan, 2002)¹ deixa clara a preocupação em registrar as colocações adverbiais, apresentando a maioria das colocações em negrito e no verbete para a base;
- em língua portuguesa, ainda não há preocupação, por parte dos lexicógrafos, em enfatizar o uso de colocações adverbiais, mesmo em dicionários mais atuais;
- os dicionários bilíngües são ineficazes como fonte de pesquisa para as colocações adverbiais;
- não há coerência na apresentação das colocações. Quando encontradas, ora estão sob o verbete do verbo ou do adjetivo, ora sob o advérbio, ora sob ambos. Em alguns casos, apresentam-se como verbetes, com definição e exemplos. Em outros, são incluídas apenas nos exemplos da palavra pesquisada, sem qualquer ênfase, o que pode ter ocorrido de forma aleatória. Os dicionários Macmillan (2002) e Activator (1993) parecem preocupar-se um pouco mais com a sistematização das informações, apresentando a maioria das colocações nos verbetes para a base e para o colocado;
- mesmo um excelente dicionário de colocações pode mostrar-se ineficiente para o tradutor: o Oxford Collocations (2002)², por exemplo, não apresenta as colocações contextualizadas. Nesse caso, o tradutor deve valer-se de outras fontes para exemplificação, como a Internet.

Esses resultados, que demonstram a ineficiência do material hoje disponível para o tradutor, corroboram a necessidade da elaboração de um glossário de colocações adverbiais bilíngüe que apresente o uso dessas associações em contexto, seguidas de informações pragmáticas, semânticas e culturais que venham a agilizar o trabalho do tradutor.

¹ Os dicionários monolíngües investigados são: Dicionário eletrônico Houaiss da língua portuguesa (2001), Dicionário Aurélio eletrônico (1999), Collins COBUILD (1987), Longman Language Activator (1993), Macmillan English Dictionary (2002) e Oxford Collocations Dictionary (2002). Dicionários bilíngües: Dicionário inglês-português Vallandro & Vallandro (1954), Dicionário inglês-português Houaiss (1982), Novo Michaelis – dicionário ilustrado português-inglês (1985) e Portuguese-English Dictionary Taylor (1970).

² Apesar de monolíngüe, esse dicionário foi pesquisado por dois motivos: além de ser exclusivamente de colocações, sabemos que fontes monolíngües são freqüentemente utilizadas pelo tradutor, muitas vezes em busca da contextualização e do uso de uma dada palavra ou expressão.

3. Lingüística de corpus: alguns conceitos básicos

Para uma melhor compreensão dos resultados aqui apresentados, faz-se necessário definir alguns conceitos subjacentes à pesquisa realizada. Apresentaremos, resumidamente, o que vem a ser Lingüística de Corpus, englobando os conceitos de *corpus*, colocação e alguns outros padrões lingüísticos.

Muitos estudiosos já definiram a Lingüística de Corpus. Dentre as definições encontradas, vale destacar a apresentada por Laviosa (2002: 6)³:

“A Lingüística de Corpus dos anos 80 e 90 pode ser definida como um ramo da lingüística geral que envolve a análise de grandes *corpora* de textos corridos, legíveis por computador, utilizando uma gama de programas computacionais elaborados especificamente para análise textual.”⁴

No entanto, a definição de seu status ainda gera discussões: alguns a consideram teoria, outros metodologia, uma vez que pode contribuir para vários domínios como a Lexicografia, o Ensino de Línguas, a Tradução, a Lingüística Computacional, etc., e outros ainda uma abordagem filosófica, como é o caso de Leech (1992: 106). Para Halliday (Tognini-Bonelli, 2001: 48), os *corpora* fazem com que teoria e dados empíricos caminhem juntos, atribuindo à Lingüística de Corpus (doravante LC) um papel que vai além do metodológico: ele conduz a uma mudança qualitativa de nossa compreensão do que seja a língua. Para este estudo, a LC revela-se como uma combinação desses vários conceitos: além do aspecto teórico, apresenta metodologia própria com ferramentas que induzem a uma nova forma de “abordar” os estudos lingüísticos. O enfoque da LC não é o estudo da palavra isolada, como podemos observar em obras lexicográficas⁵; com a manipulação de grandes quantidades de dados, torna-se possível observar a co-ocorrência de determinados termos, conduzindo-nos a unidades de significado maiores. É exatamente este aspecto que nos interessa para o estudo da convencionalidade da língua, conforme veremos a seguir.

Mas afinal, o que é *corpus*?

Segundo Sanchez (1995: 8-9 apud Berber Sardinha, 2000: 338), *corpus* é:

³ Todas as traduções apresentadas foram feitas pela autora deste artigo.

⁴ “Corpus Linguistics of the 80s and 90s can be defined as a branch of general linguistics that involves the analysis of large machine-readable corpora of running text, using a variety of software tools designed specifically for textual analysis.”

⁵ Para maiores detalhes, vide Teubert 2001: 140-144.

“Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.”

Essa definição enfatiza aspectos fundamentais para a elaboração de um *corpus*: a autenticidade dos dados, sua composição, representatividade e manuseio por computador. Além disso, um *corpus* deve ser construído para fins de investigação e estudo lingüístico. Sua elaboração irá, indubitavelmente, refletir-se nos resultados apresentados em uma dada pesquisa. Cabe ao pesquisador determinar os critérios antes da elaboração do *corpus* a ser investigado. Em linhas gerais, deve primeiramente definir quais os objetivos dessa investigação, ou seja, qual/quais aspecto(s) lingüístico(s) deseja focar. Deve também delimitar a tipologia textual, o que evitará a coleta de material que, numa etapa seguinte, poderá ser descartado. Há ainda outras questões a serem definidas nesta etapa, tais como o gênero lingüístico, o registro (formal/informal), o tipo de fonte que será investigada, como é o caso de artigos publicados apenas em revistas médicas, resumos de dissertações/teses, etc, e mesmo se os textos serão digitados, escaneados e/ou baixados da Internet.⁶ Uma outra questão é o tamanho desse *corpus*, que já deve ser considerado no início da coleta de dados e está totalmente relacionado aos objetivos do pesquisador, como atestam Tagnin & Teixeira (2004: 321-322) e Berber Sardinha (2004: 29). Pode-se também utilizar um *corpus* já construído, mas deve-se observar se ele apresenta as características necessárias para o estudo em questão. Se a intenção do pesquisador é investigar, por exemplo, o uso de um dado termo em uma área de especialidade, um *corpus* de língua geral só será útil como *corpus* de referência, não como *corpus* de estudo⁷.

Para McEnery (1997: 12), a pesquisa baseada em *corpus* apresenta várias vantagens ao pesquisador/lingüista que objetiva investigar e analisar a língua em uso, em contraste com o analista introspectivo:

- os dados são observáveis e verificáveis por todos aqueles que queiram fazê-lo;
- o *corpus* evidencia quaisquer pontos de vista que tenham sido utilizados para embasar uma dada teoria, o que não ocorre quando os dados são obtidos por meio da introspecção;

⁶ Essas informações são amplamente discutidas por Berber Sardinha (2004).

⁷ Para mais detalhes sobre tipologia de *corpus*, vide Berber Sardinha (2004: 20-22).

- o lingüista de *corpus* observa dados autênticos, enquanto o analista introspectivo vale-se de dados artificiais, por meio da manipulação do informante;
- a freqüência só é computável com os *corpora*, o que não ocorre via introspecção.

Para o estudo de uma dada língua, corroboramos a idéia de que se deve considerar fatores extra-lingüísticos. Segundo esse mesmo autor, a abordagem baseada em *corpus* e a abordagem baseada na introspecção devem estar numa relação de complementaridade, pois “a Lingüística de Corpus é, e deveria ser, uma síntese de técnicas de introspecção e observação, baseando-se numa mistura de observação artificial e natural.”⁸ (McEnery, 1997: 16)

Essa relação de complementaridade pode também ser observada nas duas abordagens utilizadas dentro da própria LC: a abordagem baseada em *corpus* e a abordagem direcionada pelo *corpus*. Segundo Tognini-Bonelli (2001: 65-66), “(...) o termo *baseado em corpus* [*corpus-based*] é utilizado para referir-se a uma metodologia que se vale do *corpus* principalmente para expor, testar ou exemplificar teorias e descrições formuladas antes de os grandes *corpora* tornarem-se disponíveis para estudo lingüístico”⁹, ou seja, o *corpus* é visto como uma fonte de exemplos utilizados para embasar teorias pré-existentes ou mesmo como um recurso para informações quantitativas sobre um sistema já bem definido. A pesquisa direcionada pelo *corpus* (*corpus-driven research*), no entanto, irá conduzir o pesquisador a aspectos da língua antes desconhecidos ou impensados para investigação. Por meio da observação da língua em uso pode-se levantar hipóteses que serão posteriormente generalizadas e em seguida unificadas em uma teoria. Diferente do lingüista tradicional, que pode deixar passar despercebido um grande número de padrões potencialmente significativos, esse pesquisador deve estar aberto a novas descobertas e ser treinado para identificar quaisquer evidências que o *corpus* possa oferecer.

Dentre os padrões que o pesquisador pode investigar, encontramos as colocações. O termo colocação foi originalmente introduzido pelo lingüista J. R. Firth para descrever o fato de algumas palavras “andarem juntas” (Tagnin 1989: 30). Segundo o Oxford Collocations Dictionary (2002: vii), “colocação é o modo como as palavras se combinam numa língua para produzir um discurso

⁸ “Corpus linguistics is, and should be, a synthesis of introspective and observational techniques, relying on a mix of artificial and natural observation.”

⁹ “(...) the term *corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study.”

natural, tanto falado quanto escrito”¹⁰. Uma definição mais detalhada para esse termo foi apresentada por Tagnin (1998, 41): “Uma colocação é uma combinação lexical recorrente, não-idiomática, coesa, cujos constituintes são contextualmente restritos e de coocorrência arbitrária.”

Essas recorrências lexicais podem ocorrer entre várias classes de palavras, tais como substantivos, verbos, adjetivos e advérbios. Segundo Hausmann (1985 apud Tagnin 1998: 41-42), uma colocação é formada por uma base, que é a palavra de conteúdo ou referência, e um colocado, cuja categoria gramatical irá determinar o tipo de colocação em questão. Embasando-se na terminologia apresentada por esse autor, Tagnin (1998: 42) apresenta-nos as **colocações adverbiais**, objeto de estudo desta pesquisa, que podem ser classificadas como a associação entre um verbo e um advérbio, como em **thank gratefully** e **pedir encarecidamente**, e um adjetivo e um advérbio, como em **lavishly illustrated** e **perdidamente apaixonado**¹¹.

Conforme já mencionado, a LC permite o estudo de padrões lingüísticos que podem tornar o discurso do falante (nativo ou não) mais natural e fluente.

Enquanto as colocações definem-se como a associação entre itens lexicais, o termo coligação é utilizado para definir a associação entre itens lexicais e gramaticais¹². Como exemplo, podemos observar o padrão coligacional da palavra *cases* (Stubbs, 2001: 65), em inglês, que freqüentemente ocorre com a categoria gramatical dos quantificadores, em expressões como *in some cases, in many cases (cases <some, many, most, more, both, several>)*.

Relacionado ao aspecto semântico, um outro padrão que podemos investigar é a prosódia semântica de uma palavra ou expressão. Berber Sardinha (2000: 359) define esse termo como “(...) [a] associação entre itens lexicais e conotação (negativa, positiva ou neutra) de campos semânticos. O nome deve-se ao fato de certas palavras prepararem o ouvinte ou o leitor para o conteúdo semântico que está por vir (...)” Como exemplo, podemos citar o verbo *cometer*, que é geralmente utilizado com palavras de conotação negativa, tais como *crime, adultério e suicídio*.

Ao investigarmos as colocações adverbiais em busca desses padrões lingüísticos, observamos que várias dessas associações faziam parte de estruturas maiores, como é o caso da colocação **acreditar piamente**, encontrada nas unidades **acredito piamente + que** (oração subordinada substantiva objetiva direta)

¹⁰ “Collocations is the way words combine in a language to produce natural-sounding speech and writing”.

¹¹ Para a definição dos outros tipos de colocação, vide Tagnin (1998).

¹² Vide Sinclair (1996), Hoey (1997) e Berber Sardinha (2004).

Crop, 10, 2004

e **acredito piamente + sintagma preposicionado** introduzido pela preposição **em** (seguida ou não de determinante). Exemplos: *Acredito piamente que esse é um nicho de mercado em processo de franca consolidação.* (www.sagarana.uai.com.br/sec_entrevista.htm) e *Sou um brasileiro paulistano e acredito piamente nas previsões meteorológicas.* (super.abril.com.br/aberta/colunas/index_california_06_02_02.html).

Com relação às associações lexicais, um exemplo a ser citado é a colocação **estupidamente gelado**. Após a investigação, descobrimos que essa colocação é comumente utilizada no feminino singular referindo-se ao substantivo cerveja. Observamos que a associação **cerveja stupidamente gelada** apresenta altos índices de co-ocorrência, caracterizando uma unidade de significado maior do que a originalmente concebida, ou seja, apenas a associação entre o advérbio e o adjetivo. Segundo Tognini-Bonelli (2001: 19), partindo de uma palavra considerada núcleo, outras são a ela associadas por meio da observação do co(n)-texto. A frequência de co-ocorrência irá formar um determinado padrão lingüístico, tanto em nível lexical quanto gramatical, conforme acima exemplificado: “Essas unidades representam uma tendência em direção à idiomaticidade e à fraseologia, uma vez que são escolhidas como uma única unidade. Apenas ao atingir sua função pragmática (...) é que podem ser vistas como ‘funcionalmente completas’ (Tognini-Bonelli 1996b).” (Tognini-Bonelli 2001: 19)¹³

Além dessas unidades de significado maiores, há também a questão do registro e da tipologia textual. Algumas colocações apresentaram restrição de uso, como é o caso de **agradecer penhoradamente**. Essa colocação foi geralmente identificada em textos de agradecimento cuja linguagem era mais formal, relacionando-se, por exemplo, a aspectos políticos (discursos, atas de sessões, pronunciamentos e informativos). Exemplo: *Agradeço, penhoradamente, a concessão que me foi feita pelos eminentes Senadores Magno Malta e Mozarildo Cavalcanti, que prezo tanto.* (www.senado.gov.br/web/senador/JFonsecapant23.htm).

A partir desse panorama, passemos ao objeto de estudo desta pesquisa (colocações adverbiais), inserido na convencionalidade lingüística.

¹³ These units represent a shift towards idiomatcity and phraseology in that they are chosen as single units. Only when they have reached their pragmatic function (...) can they be seen as ‘functionally complete’ (Tognini-Bonelli 1996b).”

4. A convencionalidade lingüística e a tradução das colocações adverbiais

Como sabemos, a convencionalidade está intimamente ligada ao comportamento social e, como não poderia deixar de ser, à língua falada por uma determinada comunidade. Com isso, considera-se fluente o falante (nativo ou não) que domina e utiliza expressões convencionais e/ou idiomáticas compartilhadas por um dado grupo lingüístico, tanto no processo de decodificação quanto no de codificação de uma dada mensagem. No caso das colocações, como observa Tagnin (2002/1: 194),

a dificuldade pode residir no fato de, em geral, não constituírem problema de compreensão, de modo que tendem a passar despercebidas. Em outras palavras, por serem em grande parte posicionais, as colocações são de fácil compreensão. Entretanto, quando se trata de produzi-las, não são facilmente buscadas na memória, uma vez que não houve um esforço consciente para memorizá-las.

Com isso, a tradução de colocações adverbiais pode parecer um processo fácil e simples para o tradutor do par de línguas inglês-português. No entanto, além do problema da reprodução de uma forma já consagrada, o uso dessas unidades relaciona-se a aspectos sintáticos, semânticos, pragmáticos e até mesmo culturais que podem (ou não) variar de uma língua para outra. Basta observar, por exemplo, o aspecto cultural implícito na colocação **estupidamente gelado**. O seu uso freqüente na forma feminina associada ao substantivo *cerveja* (**cervejaupidamente gelada**) deve-se ao fato de essa bebida ser bastante consumida dessa forma no Brasil, fato que não necessariamente ocorre em outros países. Mesmo na Austrália, cujo clima é parecido com o nosso e a cerveja também é consumida gelada, não encontramos uma forma cujo uso fosse equivalente ao da colocação em português. Segundo uma falante nativa daquele país, uma expressão que poderia ser utilizada é **outrageously cold**, mas que ela de antemão não classificaria como convencional. Ao pesquisarmos no Google, confirmamos os dados: **outrageously cold** apresentou apenas 72 ocorrências em toda a *Web*¹⁴, sem quaisquer co-ocorrências com *beer*.

Em busca da equivalência semântica, o tradutor deve estar atento ao co(n)-texto em que uma colocação se insere, como é o caso de **piously believe** e **firmly believe**. Para a primeira colocação, a forma equivalente é **acreditar piamente**.

¹⁴ Resultados coletados em 27/09/04.

Crop, 10, 2004

Já **firmly believe** pode ser traduzida tanto por **acreditar firmemente** quanto por **acreditar piamente**, fato que será determinado pelo contexto e registro apresentados no texto original. Por meio da investigação nos *corpora*, observamos que **piously believe** apresenta uso mais restrito do que sua forma equivalente em português. Enquanto **acreditar piamente** é utilizada em vários tipos de discurso, **piously believe** é mais freqüentemente ligada a crença/religião. Exemplos:

Christian people may piously believe in the help which the souls of brothers and members, who have departed this life in charity, have worn throughout life (...). (www.sspcx.ca/Rosary_Crusade/issue_07.htm)

A tendência natural brasileira é acreditar piamente em tudo que vem de fora, quer sejam equipamentos, produtos acabados ou tecnologias. (www.gpca.com.br/gil/art77.htm)

O cenário empresarial mundial parece começar a acreditar firmemente que investimento cultural é um fator de significativa importância no desenvolvimento (...).

(www2.uol.com.br/JC/_1999/0107/art0107.htm)

A escolha lexical também deve ser criticamente analisada pelo tradutor. Ao traduzir a colocação **hermetically sealed**, por exemplo, deve observar que **fechado** é muito mais utilizado com **hermeticamente** do que **selado**. Em pesquisa com o buscador Google¹⁵, por exemplo, a co-ocorrência entre **hermeticamente** e **fechado** é de 648, enquanto entre **hermeticamente** e **selado** temos apenas 77 ocorrências. A pesquisa no CD-ROM FOLHA 99, também utilizado para este estudo, corrobora esses dados: **hermeticamente fechado**: 8 ocorrências; **hermeticamente selado**: nenhuma ocorrência. Vale observar que, em pesquisa apenas pelo advérbio, os 37 casos registrados na FOLHA apresentam apenas duas associações ao verbo **selar**: uma a **selado** e outra a **selada**. Desse modo, atestamos que **hermeticamente fechado** é a forma mais freqüente e, portanto, deve ser aquela utilizada para a tradução de **hermetically sealed**, que apresenta alta freqüência de uso (92.500 ocorrências registradas na *Web* pelo Google em 16/03/04 e 19 casos no BNC¹⁶). Alguns exemplos são:

¹⁵ Realizada em 20/10/04.

¹⁶ Vide item 5.

REAL DE MORAES, Helmara Febeliana. *Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Lingüística...*

Toyota's Crown Majesta will incorporate new hermetically sealed lead batteries (...)." (BNC); "Não exponha o frasco à luz solar direta. Mantenha as tiras reagentes sem uso dentro do frasco original, hermeticamente fechado. (www.lilly.com.br/saude/bulas/Glicofita%20Plus.pdf).

Sobre a questão gramatical, deve-se observar que pode haver identidade semântica sem haver identidade gramatical entre duas línguas, como sugere Lindquist (1989). É o caso das colocações **blindingly obvious** (advérbio+adjetivo) e **pay dearly** (verbo+advérbio), traduzidas respectivamente por **óbvio ululante** (substantivo+adjetivo) e **pagar caro** (verbo+adjetivo com função adverbial). Além dessas, podemos também citar **explain in detail** (verbo+sintagma preposicionado), traduzida por **explicar detalhadamente** (verbo+advérbio). Ainda com relação a esse aspecto, pode haver a omissão do colocado da língua de partida, como ocorre com **highly successful**, geralmente traduzida pela expressão **bem-sucedido** (raramente modificada por um advérbio em língua portuguesa). Exemplos: *The 25 Sales Habits of Highly Successful Salespeople.*" (www.growingresults.com/prod/1558503919.html); *101 segredos para ser um vendedor bem sucedido.* (www.secbrm.org.br/vendedor.htm).

Traduzir os advérbios terminados em **-ly** em inglês não é tão óbvio quanto parece. Em geral, o sufixo **-mente** é a primeira opção que passa pela cabeça do tradutor. No entanto, nem sempre essa solução é viável, como é o caso de **chronically ill** (não dizemos **cronicamente doente** em português!). O inverso também ocorre: para **mentir descaradamente**, temos **lie outright**.

Deve-se também observar a função do advérbio em cada colocação. No estudo já desenvolvido, verificamos que o advérbio pode ser um intensificador, como é o caso de **totally convinced** e **completely blind**; pode atribuir uma outra carga semântica ao verbo ou adjetivo que modifica, como é o caso de **descaradamente** em **mentir descaradamente**, cujo uso é diferente daquele apresentado apenas pelo verbo; pode relacionar-se ao próprio ato de fala, como em **speak frankly**, **speak fluently**, etc. Na tradução, o advérbio escolhido deve exercer o mesmo papel a ele atribuído na língua de partida. Como podemos observar, o tradutor deve estar ciente dessas associações que, muitas vezes, passam despercebidas durante o ato tradutório.

A seguir, apresentaremos os corpora, o banco de dados e as ferramentas utilizadas para o desenvolvimento desta investigação lingüística.

5. Os *corpora*, o banco de dados e as ferramentas utilizadas

A pesquisa baseada em *corpus* apresenta várias vantagens para o tradutor. Dentre elas, vale destacar:

- acesso à língua em uso, tanto na Língua de Partida quanto na Língua de Chegada (dados autênticos);
- levantamento de vários padrões lingüísticos, tais como colocações, coligações, prosódia semântica e campo semântico;
- observação do registro/tipologia textual em que a unidade analisada geralmente ocorre.

Neste estudo, utilizamos as seguintes fontes para investigação:

- **BNC (British National Corpus)**: *corpus* fechado de 100 milhões de palavras de inglês britânico (falado e escrito), cuja versão *online* (aqui utilizada) permite acessar gratuitamente parte desses dados. Disponível em <http://sara.natcorp.ox.ac.uk/lookup.html>;
- **CD-ROM FOLHA EDIÇÃO 99**: banco de dados¹⁷ de 156 milhões de palavras, formado por textos integrais publicados pelo jornal Folha de São Paulo entre 1994 e 1998.
- A **WEB** como um *corpus* aberto, tanto para o inglês quanto para o português, acessada com o buscador **Google** (www.google.com.br).

Para a manipulação desses resultados, necessitávamos de outras ferramentas que, além de confirmar a relação de colocabilidade entre os termos, facilitassem a análise por parte do pesquisador. Utilizamos então algumas ferramentas desenvolvidas pelo Prof. Dr. Tony Berber Sardinha, da Pontifícia Universidade Católica de São Paulo (PUC-SP), disponíveis em <http://lael.pucsp.br/corpora>. São elas:

- a) a **calculadora on-line**, cuja função é apresentar os resultados para as fórmulas Informação Mútua e Escore T, valores importantes para a definição do grau de fixidez da colocação, ou seja, se há recorrência sistemática de termos que co-ocorrem (informação privilegiada pela Escore T) ou mesmo co-ocorrência freqüente entre dois ou mais termos que pouco ocorrem em um dado *corpus* (privilegiada pela Informação Mútua);
- b) o **KWIC Google**, um formatador de resultados do buscador Google que alinha esses dados, apresentando a palavra de busca centralizada e as demais uniformemente distribuídas à direita e à esquerda da mes-

¹⁷ Chamar o CD-ROM FOLHA EDIÇÃO 99 de *corpus* é inadequado, uma vez que os dados são fechados e só podem ser acessados pela ferramenta de busca do próprio CD, fato que limita sua manipulação.

ma. Esse formato intitula-se concordância KWIC, e neste caso apresenta-se seguido de duas listas com os colocados mais freqüentes à direita (com a respectiva freqüência de ocorrência para cada termo) e à esquerda. Essa ferramenta é bastante útil na medida em que facilita a visualização dos dados e a identificação de padrões lingüísticos.

Além desse suporte tecnológico e dos dados quantitativamente coletados e analisados, fizemos uma análise qualitativa criteriosa de cada colocação. Essa análise consistiu da observação de um grupo de colocações adverbiais nas diferentes fontes de pesquisa, pela pesquisadora, que se valeu também de sua intuição enquanto falante de ambas as línguas em questão, numa relação de complementaridade. Nessa análise, foram observados e anotados os padrões colocacionais e coligacionais em que cada associação ocorria, bem como a prosódia e o campo semântico. Além disso, observamos também questões relacionadas ao registro (linguagem formal/ informal) e à tipologia textual. Todos os dados observados foram devidamente anotados durante esta fase da pesquisa.

Em seguida, passaremos à descrição da metodologia de busca adotada, uma vez que a pesquisa baseada em *corpus* é bastante ampla e necessita ser norteada e direcionada para a produção de resultados comparáveis (no caso da pesquisa bilíngüe) que contribuam para o estudo das línguas envolvidas.

6. A metodologia de busca para o levantamento e análise de dados

Para a investigação das colocações a serem incluídas no glossário, criamos inicialmente uma metodologia de pesquisa para cada fonte de dados, de acordo com os recursos que cada uma oferecia. No BNC, que possui ferramenta de busca própria e não permite quaisquer alterações além dos recursos já disponibilizados, analisamos os resultados apresentados, que se limitam ao máximo de 50 ocorrências na versão *online*. Nesse caso, foram seguidos os critérios de busca para cada colocação adverbial explicitados a seguir. Já no buscador Google, que permite alguns ajustes em “pesquisa avançada”, alteramos os seguintes itens:

- número de resultados a serem apresentados: 100;
- idioma: inglês ou português, dependendo da colocação;
- domínio: .br para o português, visando eliminar os resultados apresentados em .pt.

Vale observar que as colocações foram digitadas no *box* para “Procurar resultados com a expressão”, que viabiliza a busca pela expressão exatamente como foi inserida, em um mesmo texto, sem quaisquer palavras intervenientes. Caso o pesquisador queira utilizar o *box* “com todas as palavras”, basta digitar a colocação entre aspas que obterá os mesmos resultados.

O CD-ROM da FOLHA, que também possui ferramenta de busca própria, disponibiliza tanto uma busca simples quanto uma pesquisa avançada. Neste estudo, optamos pela “busca avançada” justamente por filtrar melhor os resultados. Além da opção “a frase exata”, podemos utilizar apenas uma parte da palavra de busca (que pode ser o radical), seguida de asterisco (*). Exemplo: *gelad**, que nos traria as ocorrências com *gelado*, *gelada*, *gelados* e *geladas*. Apesar de bastante útil, esse recurso não se aplica a mais de uma palavra, pois ao clicarmos em “todas as palavras”, opção necessária devido ao uso do asterisco em um dos termos da colocação para um levantamento geral, a ferramenta apresenta todos os textos em que as formas pesquisadas ocorrem, independente da posição em que se encontram no texto, o que dificulta o levantamento de dados e demanda análise de cada texto para exclusão daqueles em que os termos não co-ocorram. Após a busca, a ferramenta lista todos os textos em que a expressão inserida ocorreu (independente de utilizarmos ou não o asterisco); ao clicar no título, temos acesso ao texto integral, devidamente identificado com autoria, data e seção do jornal onde foi publicado. Nesse momento, o pesquisador passa a analisar, individualmente, cada texto. Apesar de ser um processo mais trabalhoso, podemos realizar uma análise mais rica e detalhada de cada colocação, observando o assunto em questão e melhor definindo a prosódia semântica e/ou campo semântico de uso. Nesta pesquisa, limitamos a análise da forma mais freqüente para cada colocação, bem como apenas para o advérbio (vide critérios abaixo discriminados) a 100 textos.

Consideradas as limitações e recursos oferecidos pelas fontes de pesquisa utilizadas, estabelecemos alguns critérios mais específicos para a busca de cada colocação adverbial. São eles:

Para a associação entre verbos e advérbios:

Em inglês:

- verbo no infinitivo (sem a partícula *to*), que já engloba a conjugação do presente simples, com exceção da terceira pessoa do singular, além de outros tempos verbais, como o futuro;
- verbo conjugado na terceira pessoa do singular do presente simples;
- verbo conjugado no passado simples.

Em português:

- verbo no infinitivo;
- verbo conjugado na primeira e na terceira pessoas do singular e do plural do presente do indicativo;
- verbo conjugado na primeira e na terceira pessoas do singular e do plural do pretérito perfeito.

Para a associação entre adjetivos e advérbios:

Em português:

- adjetivo no masculino singular e plural;
- adjetivo no feminino singular e plural.

Para ambas as línguas, testamos a possibilidade de anteposição e posposição do advérbio junto ao verbo ou adjetivo. Para a colocação *firmly believe*, por exemplo, investigamos tanto essa forma quanto *believe firmly*. Como *firmly believe* apresentou maior frequência, foi considerada o ponto inicial para a busca das outras formas, que seriam *firmly believes* e *firmly believed*. Além desse levantamento, fizemos também uma investigação apenas com o advérbio que compunha a colocação, com o objetivo de observar se os resultados dessa busca apresentariam a colocação estudada de forma significativa, tal que corroborasse a fixidez da associação. Além disso, visávamos observar com quais outras palavras o advérbio poderia co-ocorrer, induzindo a novas investigações (pesquisa direcionada pelo *corpus*). Com relação à colocação, depois de levantada a frequência de cada forma, analisamos minuciosamente aquela que havia apresentado maior frequência; dependendo dos resultados obtidos, investigamos também as outras formas para confirmar (ou não) alguma observação já feita. No caso da colocação acima mencionada, obtivemos as seguintes frequências de ocorrência:

- *firmly believe*: BNC: 54 vezes; Google: 373.000 vezes;
- *firmly believes*: BNC: 31 vezes; Google: 94.900 vezes;
- *firmly believed*: BNC: 28 vezes; Google: 47.600 vezes.
- *firmly*: BNC: 3.819 vezes; Google: 4.580.000 vezes.

Com esses resultados, a forma definida para análise mais detalhada foi *firmly believe*.

Tanto para o inglês quanto para o português, a metodologia adotada para a investigação lingüística foi a mesma. No entanto, como cada colocação pode apresentar usos e co-ocorrências diversas, a pesquisa foi sendo direcionada pelos resultados obtidos nas fontes utilizadas.

7. Resultados obtidos por meio da pesquisa baseada em *corpus*

A título de exemplificação, apresentaremos a seguir os resultados das análises quantitativa e qualitativa para a colocação **deeply hurt** e sua forma equivalente **profundamente magoado**, escolhidas por apresentarem dados relevantes para o tradutor observáveis apenas pela pesquisa em *corpora*. Ambas as colocações apresentaram resultados acima do ponto de corte para as fórmulas Informação Mútua (mínimo de 3) e Escore T (mínimo de 2), que atestam o grau

de fixidez da associação analisada. Para esses cálculos, os dados necessários foram coletados por meio do buscador Google. Abaixo, os resultados:

Colocação	Data*	Informação Mútua	Escore T
deeply hurt	30/08/03	8.13	175.15
profundamente magoado	02/09/03	11.92	15.64

*Refere-se a quando os dados foram coletados e o cálculo feito.

Nesta análise, além da comprovação de dados previamente intuídos, pudemos descobrir novas informações, como é o caso da tendência à formação de binômios tanto em inglês quanto em português. Abaixo, os resultados do levantamento para cada colocação, seguidos de uma análise contrastiva de dados.

Deeply hurt

Foram encontradas 21 ocorrências no BNC. Quanto à função da palavra *hurt*, ela apresenta-se tanto como verbo quanto como adjetivo (forma mais frequente). Enquanto verbo, apresentou-se 7 vezes na voz passiva, com menção do agente. Exemplos:

“Don’t lie!”, she said bitterly, deeply hurt by his endless obsession with power because it attacked the very roots of her growing feelings for him.”

“The late rock star Freddie Mercury was deeply hurt by media coverage of his battle against Aids, a friend said yesterday.”

É também utilizada em oração adjetiva, seguida de vírgula, para ênfase: “Deeply hurt, Vincent jumped to the conclusion that news of his friendship with Sien had reached him and upset the man.”

A palavra **hurt**, em função de adjetivo, foi identificada nas seguintes associações: **BE**¹⁸ + **deeply hurt** (8 vezes) e **FEEL** + **deeply hurt** (3 vezes). Com relação ao aspecto semântico, a colocação é utilizada para descrever situações relacionadas a sentimentos (em praticamente todos os casos). Em português, a forma equivalente seria **profundamente magoado**, mas **profundamente feri-**

¹⁸ As palavras em letras maiúsculas estão lematizadas. As demais estão na forma em que foram encontradas.

do também pode ser utilizada, uma vez que refere-se a ressentimentos na maioria dos exemplos observados com a ferramenta Google.

Seguindo os critérios estabelecidos para a análise das colocações, investigamos as 50 ocorrências apresentadas pelo BNC apenas para o advérbio *deeply*. Nesses exemplos, a co-ocorrência com *hurt* foi mínima. Os dois casos encontrados são:

“This is why men who have been deeply hurt in an extra marital affair often have a compulsion when it ends to tell all to their wives in a great burst of weeping.”

“Pat was deeply hurt, but Ken maintained a look on his face that seemed to be as shiny as the brown shoes he wore under those immaculately pressed trousers.”

Por meio do buscador Google, que identificou 104 ocorrências para a colocação, os padrões encontrados foram¹⁹:

- ~ **hurt** + **by** + **SN (sintagma nominal)**, caracterizando o uso da voz passiva – 17 casos (16,3%): “He was deeply hurt by the unkind remarks.”;
- **BE** + ~ **hurt** – 44 casos (42,3%), com 27 ocorrências não seguidas da preposição *by*, utilizada para indicar o agente da passiva. Nesses casos, torna-se difícil discriminar o uso do verbo na passiva do uso do adjetivo, pois além da própria estrutura temos a limitação do contexto. Ex.: “I would be deeply hurt but I would stay calm and refuse.”;
- **FEEL** + ~ **hurt** – 3 casos: “Someone had said something about me that seemed unkind, unfair, and untrue, and I felt deeply hurt.”;
- **SN (inclusive nomes próprios)** + ~ **hurt**, caracterizando o estilo jornalístico geralmente observado em manchetes de jornais e revistas – 10 casos (9,6%): “European Baptists ‘deeply hurt’ about possible split in world alliance (...)”;
- ~ **hurt**_[verbo] + **SN** – 7 casos: “I wondered if someone or something deeply hurt you.”;
- ~ **hurt**_[adj.] + **substantivo** – 7 casos: “He has done a remarkable job of giving the reader insight to the Primary Caregiver’s (mother’s) relationship or lack of, with this deeply hurt child.”;
- ~ **hurt** + **clause (introduzida por when ou that)** – 7 casos: “I felt deeply hurt when you tried to take my GAP.”.

¹⁹ Neste artigo, convencionou-se utilizar o símbolo ~ (til) para substituir o colocado.

Em 12 ocorrências, a colocação apresenta a conjunção *and* anteposta ou posposta. Isso pode significar que *hurt* é freqüentemente acompanhada de adjetivo ou particípio para enfatizar uma dada idéia, o que caracterizaria certa tendência à formação de binômios. Exemplos:

“I am deeply hurt and humiliated.”

“I am deeply hurt and offended by your homosexual myths article (...).”

“(...) you have deeply hurt and disappointed Gloria who really did want your help.”

“I can be drunk and happy, or drunk and mad, but drunk and deeply hurt is what I’m struggling with.”

“(...) and then proceed to be rejected and deeply hurt.”

A prosódia semântica é essencialmente negativa, acompanhando o sentido do próprio verbo. Como exemplo, podemos citar: *rejected and deeply hurt*, *deeply hurt and humiliated*, e *drunk and deeply hurt*. Quanto ao campo semântico, a colocação relaciona-se a sentimentos, sempre considerados negativos devido ao significado da colocação. É interessante observar que **deeply hurt** raramente refere-se a problemas físicos, caso em que o tradutor deveria utilizar *injured*.

Profundamente magoado

Na FOLHA, encontramos apenas 4 ocorrências que apresentaram os seguintes resultados:

- **ESTAR ~ magoado + com SN (alguém/algo)** – 2 vezes: “(...) Simon afirmou que o presidente estava ‘profundamente magoado’ com o abuso de preços.”
- **DEIXAR + SN (alguém) + ~ magoado**: 1 vez.
- **estou + ~ magoado**: 1 vez.

Apesar de a forma masculina singular ter sido a mais freqüente, investigamos também a forma feminina e o plural para ambas as associações. Porém, nada foi encontrado. A maior presença na forma masculina pode indicar que o homem fica mais **profundamente magoado** do que a mulher. No entanto, para comprovar tal suposição faz-se necessário um estudo mais aprofundado também em outros *corpora*.

Na pesquisa para o advérbio, que apresentou 1.835 ocorrências, identificamos apenas os 4 casos já analisados com a palavra **magoado**. Nas 50 primeiras

REAL DE MORAES, Helmara Febeliana. *Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Linguística...*

ocorrências, foram observadas outras associações: ~ *DECEPCIONADO* (4 vezes), ~ *ENRAIZADO*, *CONHECER* ~ *SN*, *LAMENTAR* ~ *SN*, *DESEJAR* ~ *SN*, *AFETAR* ~ *SN*, *MUDAR* ~, *ALTERAR* ~ *SN*, ~ *DIVIDIDO*, todas com frequência 2. Ao observar todas as ocorrências (inclusive as de frequência 1), verificamos que tanto os verbos quanto os adjetivos utilizados com esse advérbio pertencem a campos semânticos semelhantes:

- sentimentos: ~ enojado, ~ entristecido, ~ ofensivo, ~ arrependido, ~ *DECEPCIONADO*, ~ comprometido, ~ *MARCADO*; *PERTURBAR* ~, *LAMENTAR* ~, *TOCAR* ~, *INCOMODAR-SE* ~, *AFETAR* ~; *EMOCIONAR* ~;
- mudanças: ~ *DIVIDIDO*; *MODIFICAR* ~, *AFETAR* ~, *AVANÇAR* ~, *ALTERAR* ~;
- conhecimento: *CONHECER* ~, *ENTENDER* ~.

Alguns exemplos:

Estou profundamente arrependido por tudo o que fiz de errado.

Isso implica que um defeito em apenas um gene pode afetar profundamente as habilidades linguísticas.

Para quem quer conhecer profundamente a obra e vida de O'Keeffe, um simpósio sobre a artista acontece na cidade em 21 e 22 de julho.

No Google, obtivemos 175 ocorrências da colocação na forma masculina. No entanto, vale observar que a forma feminina também apresentou resultados significativos: 157 ocorrências. Abaixo, os resultados da análise desenvolvida.

- **oração reduzida de participio, entre vírgulas** (16 vezes): *Profundamente magoado, deixou a política após ter seu nome vetado pelo imperador (...)*. O fato de a colocação apresentar-se entre vírgulas indica a ênfase atribuída ao fato. Essas orações podem também classificar-se como adjetivas;
- **FICAR + ~ magoado** – 18 oc.: *Apesar de ter sido absolvido por unanimidade pela Relação do Maranhão em 12-MAR-1872, ficou profundamente magoado com o fato (...)*;
- **DEIXAR + alguém + ~ magoado** – 14 oc.: *(...) do governo só fez com que a venda do livro disparasse – “saiu-lhes o tiro pela culatra”, brinca Saramago – mas deixou o autor ‘profundamente magoado’ por viver (...)*;
- **ESTAR + ~ magoado** – 31 oc. (17,8% do total): *O presidente do Senado está profundamente magoado. Coisa de dor no coração mesmo.*;
- **SENTIR-SE + ~ magoado** – 5 oc.: *Sente-se profundamente magoado e humilhado, ofendido, agredido, maltratado, desrespeitado, despojado, atoradoado, enfrentado, ilícitado e por fim caçado!*;

- **SER +~ magoado** – 2 oc.: *Arranjou e gravou todo mundo da bossa. Hoje é [sic] profundamente magoado de não ser reconhecido e não continuar na indústria, que nem o recebe mais.*

Em português, observamos também a formação de binômios, com a associação da colocação a outro verbo ou adjetivo. No Google, foram identificados 10 casos, abaixo exemplificados:

- José estava profundamente magoado e confuso.*
- Profundamente magoado e desapontado, Mosbé orou a Deus.*
- Tô com RAIVA! Profundamente magoado e irritado.*

No aspecto semântico, as co-ocorrências atestam uso relacionado a sentimentos e reações, conforme já mencionado.

Abaixo, apresentamos uma análise contrastiva das colocações observadas. No caso do glossário, esses dados seriam distribuídos de acordo com a língua em que se encontra cada colocação analisada. Caberia ao consulente contrastar as informações obtidas, se necessário.

Comentários:

- a palavra **hurt** apresenta-se tanto como verbo quanto como adjetivo, sendo esta última a função mais freqüente. Em português, **magoado** também é utilizado com maior freqüência na função adjetiva;
- as associações mais freqüentes são com os verbos **be** e **feel**, em inglês, e **ficar**, **deixar** e **estar**, em português;
- tendência à formação de binômios, tanto em inglês quanto em português: *deeply hurt and humiliated, deeply hurt and offended; profundamente magoado e confuso, profundamente magoado e irritado;*
- prosódia semântica essencialmente negativa, relacionando-se a sentimentos, talvez pelo próprio significado da colocação. Na maioria dos exemplos, não há referência ao aspecto físico.

A seguir, apresentaremos a organização de um glossário de colocações adverbiais, bem como uma exemplificação com os verbetes **lavishly illustrated** e **fartamente ilustrado**, uma vez que se trata de uma fonte de referência bilíngüe.

8. Um glossário de colocações adverbiais bilíngüe baseado em *corpus*

Este glossário objetiva facilitar e agilizar o trabalho do tradutor tanto em textos literários quanto técnicos, uma vez que trata de aspectos lingüísticos e extra-lingüísticos presentes em qualquer tipologia textual ou gênero discursivo. Para a elaboração deste glossário, as estruturas definidas foram:

Macroestrutura: ordem alfabética, com entrada pelo verbo (infinitivo ou participípio passado, de acordo com a forma mais freqüentemente utilizada) ou adjetivo (masculino singular).

REAL DE MORAES, Helmara Febeliana. *Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Lingüística...*

Microestrutura:

Artigo = {Entrada + Enunciado lexicográfico (+PI, +PFE, +PP)}

Onde:

PI = paradigma informacional (categoria gramatical);

PFE = paradigma das formas equivalentes;

PP = paradigma pragmático (exemplificação, observações quanto ao uso e notas para o tradutor, se necessário).

Para exemplificar, observemos os verbetes para as colocações **lavishly illustrated** e sua forma equivalente **fartamente ilustrado**.

Na direção inglês-português temos:

illustrated, lavishly – adv.+part. passado – fartamente ilustrado

Lavishly illustrated, especially in the restoration chapters, this book is a worthy tribute to one of the classic transport aircraft of all time. (BNC)

The exhibition is lavishly illustrated with colour photographs and features a moss garden composed entirely of local Ulster mosses. (BNC)

Obs. pragmáticas

Padrões lingüísticos:

~²⁰ + subst.: ~ *book, volume, publication, biography, account, history, guide, catalogue, chapters* (mesmo campo semântico).

~ + adj. + book: ~ *story book, picture book.*

~ with + sintagma nominal ((det.)+(adj.)+subst.): ~ *with many colour photographs.*

~ in + sintagma nominal: ~ *in full colour.*

♦ Apresenta-se tanto em posição atributiva (mais freqüente) quanto predicativa:

The handsome, lavishly illustrated catalogue costs £19.95 in hardback and £12.95 in paperback. (BNC)

As part of the Scientific American Library series, this book is lavishly illustrated in full colour. (BNC)

Na direção português-inglês, a forma equivalente:

²⁰ No caso do glossário, convencionou-se utilizar o símbolo ~ (til) para substituir a colocação estudada.

ilustrado, fartamente – adv.+part. passado – lavishly illustrated
(...) O livro é *fartamente ilustrado* com trabalhos feitos pelo autor e contém 50 reproduções em cores de obras de diversos artistas de diferentes estilos que (...).

(www.cybermind.com.br/hallawell/amao.htm)

(...) *Fartamente ilustrado, em cores, o livro é impresso em papel de primeira qualidade, no formato 160x260mm, brochura, com 311 páginas. 60,00.*

(www.vet.ufmg.br/Publicacoes/Livros)

Obs. pragmáticas

Padrões lingüísticos:

substantivo+ ~ : *site, livro, volume, álbum, guia, manual* ~ (mesmo campo semântico). O subst. *livro* apresentou maior frequência de co-ocorrência.

~ +com+sintagma nominal (subst. de um mesmo campo semântico: *fotos, fotografias, mapas, figuras, etc.*): ~ *com fotos, mapas e documentos oficiais.*

~ em+sintagma nominal (subst.): ~ *em cores, ~ em policromia.*

Gênero textual predominante: propaganda.

- ◆ Apresenta-se como adjunto adnominal e em função predicativa (mais freqüente):

(...) *Os lugares religiosos mais importantes do mundo são apresentados num guia *fartamente ilustrado* com mapas, fotos aéreas e detalhes de tesouros históricos.* (www.univap.br/cultura/dialogo/60/07.htm)

O livro é *fartamente ilustrado* com trabalhos feitos pelo autor (...)

(www.cybermind.com.br/hallawell/amao.htm)

- ◆ É freqüentemente encontrada entre vírgulas, com o objetivo de enfatizar a informação apresentada:

*Mas o livro, *fartamente ilustrado*, fornece, de forma bastante didática, subsídios para a identificação de famílias, gêneros e espécies (...).* (www.jornalda paisagem.com.br/publi/publicacoes.htm)

- ◆ É mais freqüente no masculino singular.

Ao final do glossário, deverá ser apresentado um índice remissivo apenas com os advérbios para que o consulente possa verificar com quais adjetivos e/ou verbos um dado advérbio co-ocorre.

REAL DE MORAES, Helmara Febeliana. *Traduzir Colocações Adverbiais: Tarefa Fácil? Parece, Mas Não É – Um Estudo Contrastivo de Colocações Adverbiais sob o Enfoque da Lingüística...*

9. Conclusões

Como podemos observar, a pesquisa baseada e/ou direcionada pelo *corpus* é fundamental para o trabalho tanto do pesquisador quanto do tradutor. Por meio da análise de colocações adverbiais, item que passa despercebido pela maioria das obras lexicográficas hoje disponíveis, descobrimos quão rica pode vir a ser uma investigação lingüística embasada na Lingüística de Corpus. A tradução de colocações adverbiais encontra, nessa nova forma de estudar a língua, recursos antes não disponíveis que podem responder a questões relacionadas ao uso de cada associação. No entanto, para que os resultados sejam produtivos, deve-se observar a necessidade de uma metodologia de pesquisa definida logo no início do processo.

A proposta de um glossário de colocações adverbiais bilíngüe decorre dos resultados obtidos durante o levantamento de dados. Por que não partilhar com tradutores informações tão detalhadas e observáveis apenas por meio da pesquisa em *corpus*? Já que o tempo desse profissional é tão escasso para ater-se a apenas um item lingüístico, nada melhor do que utilizar uma ferramenta que agilize seu trabalho. Afinal, traduzir colocações adverbiais parece fácil, mas depois de investigarmos o que os dados autênticos tem a nos revelar, descobrimos que a questão é muito mais complexa do que parece. Além disso, a pesquisa baseada e/ou direcionada pelo *corpus* proporciona várias outras vantagens para o pesquisador: ele observa/confirma (ou não) dados levantados intuitivamente e tem a oportunidade de fazer novas inferências e descobrir novas informações, trabalhando com a intuição e a observação numa relação de complementaridade.

Como pudemos observar, há sempre novos dados a serem descobertos, e a Lingüística de Corpus vem ao encontro das expectativas desse novo pesquisador, que sempre busca aprimorar sua investigação lingüística.

Referências

Literatura teórica

- BERBER SARDINHA, Tony P. 2004. *Lingüística de Corpus*. Barueri: Manole. 410 p.
- _____. 2000. Lingüística de Corpus: histórico e problemática. *D.E.L.T.A.*, São Paulo, v. 16, n. 2, p. 323-367.
- LAVIOSA, Sara. 2002. *Corpus-based translation studies: theory, findings, applications*. Amsterdam – New York: Rodopi. 138 p.
- LEECH, Geoffrey. 1992. Corpora and theories of linguistic performance. In: SVARTVIK, Jan. (Ed.). *Directions in Corpus Linguistics*, The Hague: Mouton de Gruyten, p. 105-122.
- LINDQUIST, Hans. 1989. *English Adverbials in Translation: a corpus study of Swedish renderings*. Lund: Lund University Press.

- McENERY, Tony; WILSON, Andrew. 1997. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- STUBBS, Michael. 2001. *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- TAGNIN, S. E.O. 2003. Os Corpora: instrumentos de auto-ajuda para o Tradutor. In: Stella E. O. Tagnin (Org.). *Cadernos de Tradução*, n. 9 – 2002/1, número especial sobre Corpus e Tradução, Universidade Federal de Santa Catarina, Florianópolis: Núcleo de Tradução, 191-219.
- TAGNIN, Stella E. O. 1998. *Convencionalidade e produção de texto: um dicionário de colocações verbais inglês/português português/inglês*. Tese (Livre Docência em Língua Inglesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.
- TAGNIN, Stella E. O. 1989. *Expressões idiomáticas e convencionais*. São Paulo: Ática.
- TAGNIN, Stella E. O.; TEIXEIRA, Elisa Duarte. 2004. Lingüística de Corpus e tradução técnica – relato da montagem de um corpus multivarietal de culinária. *TradTerm*, n. 10, São Paulo: Humanitas/USP, p. 313-358.
- TEUBERT, Wolfgang. 2001. Corpus Linguistics and Lexicography. *International Journal of Corpus Linguistics*, v. 6 (Special Issue), p. 125-153.
- TOGNINI-BONELLI, Elena. 2001. *Corpus Linguistics at work*. Amsterdam/ Philadelphia: John Benjamins.

Dicionários

- COLLINS COBUILD *English language dictionary*. 1987. London: Collins Publishers.
- FERREIRA, Aurélio Buarque de Holanda. 1999. *Dicionário Aurélio eletrônico: Século XXI*. Versão 3.0 [S.I.], Lexicon Informática Ltda. 1 CD-ROM.
- HOUAISS, Antonio. *Dicionário eletrônico Houaiss da língua portuguesa*. 2001. Versão 1.0. [S.I.], Editora Objetiva Ltda. 1 CD-ROM.
- HOUAISS, Antonio (Ed.). 1982. *Webster's dicionário inglês-português*. Rio de Janeiro: Record.
- LONGMAN LANGUAGE ACTIVATOR. 1993. Essex: Longman.
- MACMILLAN *English dictionary for advanced learners of American English*. 2002. Oxford: Macmillan. Inclui 1 CD-ROM.
- NOVO MICHAELIS *dicionário ilustrado português-inglês*. 1985. 37. ed., v. 2, [s.l.]: Melhoramentos.
- OXFORD COLLOCATIONS *dictionary for students of English*. 2002. Oxford, Oxford University Press.
- TAYLOR, James L. 1985. *Webster's Portuguese-English dictionary*. 4. ed. Rio de Janeiro: Record.
- VALLANDRO, Leonel & VALLANDRO, Lino. *Dicionário inglês-português*. 1954. 1. ed. Rio de Janeiro – Porto Alegre – São Paulo: Editora Globo.



Uma Análise Contrastiva Preliminar do Verbo *Limp* em Inglês e suas Traduções em Português do Brasil

Kátia Hanna*
Tereza Cristina Hilst**

Resumo: Baseando-se em estudos desenvolvidos por Váradi e Kiss sobre a bidirecionalidade semântica em dicionários mono e bilingües, este trabalho tem por objetivo observar o comportamento bidirecional do verbo inglês *limp* em relação à equivalência gramatical e semântica no par de línguas inglês-português. Para desenvolver a pesquisa, utilizaram-se duas ferramentas eletrônicas, o corpus paralelo alinhado de textos literários COMPARA, que proporcionou efetuar um “pingue-pongue” entre textos originais e traduções em inglês britânico, americano e sul-africano e em português do Brasil, e o WebCorp, no qual se buscou verificar o comportamento gramatical e semântico que o verbo *limp* e suas traduções assumem em textos originais das línguas inglesa e portuguesa.

Palavras-chave: tradução; lingüística de corpus; lingüística aplicada.

Abstract: Based on studies conducted by Váradi & Kiss about the semantic bidirectionality in mono and bilingual dictionaries, this paper aims at observing the bidirectional behavior of the English verb *limp* concerning its grammatical and semantic equivalence in English and Portuguese. To achieve our goal, we used two electronic tools, the

* Mestranda do Programa de Estudos Lingüísticos e Literários em Inglês, FFLCH/USP.

** Ex-aluna do Curso de Especialização em Tradução Lato sensu da FFLCH-USP.

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

Portuguese-English parallel corpus COMPARA, that enabled a “ping-pong” between original and translated texts in British, American and South-African English and Brazilian Portuguese, and WebCorp, which was helpful to give examples of the grammatical and semantic behavior of the verb limp and its translations in texts written originally in English and Portuguese.

Keywords: *translation; corpus linguistics; applied linguistic.*

1. Introdução

O tradutor no exercício de seu trabalho confronta-se com várias dificuldades inerentes a seu ofício, dentre elas os diferentes valores semânticos do léxico das línguas de partida (L1) e de chegada (L2). Um dos recursos por ele utilizado para saná-las é a consulta a dicionários mono e bilíngües. Acreditava-se que as “definições” oferecidas pelos dicionários eram suficientes para o tradutor efetuar a escolha da palavra mais adequada ao contexto. No entanto, a prática mostrou-nos o contrário. Devido ao reduzido espaço físico do material lexicográfico optou-se, comercialmente, pela supressão e/ou redução do contexto nos quais os termos aparecem, fornecendo assim tão somente um rol de equivalências sem muitas indicações de qual a melhor opção para o caso pesquisado. Outra dificuldade constatada apresentou-se na validade da bidirecionalidade das equivalências entre L1 e L2 sugeridas pelos dicionários. Wolfgang Teubert¹ (1999) em sua pesquisa em dicionários bilíngües mostrou-nos que ela raramente ocorre e sugere como material de apoio ao tradutor *corpora* lingüísticos monolíngües para observação da convencionalidade² na língua de chegada.

Um dos pontos investigados por Váradi e Kiss³ (2001) foi a bidirecionalidade da tradução de três termos num *corpus* paralelo alinhado composto por um texto original em inglês e sua tradução para o húngaro. Através dos dados coletados em sua pesquisa, corroboraram as observações anteriormente feitas por Teubert, nas quais relata a baixa freqüência de bidirecionalidade nas equivalências e salienta que a necessidade de consulta a um *corpus* monolíngüe é diretamente proporcional ao uso metafórico do termo a ser traduzido.

¹ Apud Váradi e Kiss (2001), p. 167-168.

² Tagnin (1989), p. 12.

³ Váradi e Kiss (2001), p. 169.

Crop, 10, 2004

Pretendemos neste trabalho dar continuidade a essa linha de pesquisa observando um termo num *corpus* paralelo bidirecional de textos originais em inglês e em português e suas respectivas traduções.

2. Objetivo

O verbo de movimento *to limp* foi o escolhido para nossa pesquisa. Buscamos verificar como esse verbo se comporta em relação à frequência de bidirecionalidade na passagem de uma língua para outra, assim como analisar as opções fornecidas nas traduções em relação à escolha de equivalência palavra-palavra ou palavra-conjunto de palavras e a manutenção ou não da classe gramatical.

A fonte por nós escolhida foi o *corpus* paralelo alinhado de textos literários COMPARA. Os textos originais e traduzidos são excertos de obras escritas em inglês britânico, americano e sul-africano e português do Brasil.

Tendo em mente para *to limp* – e todas as traduções coletadas no “pingue-pongue” efetuado no COMPARA – sua acepção de verbo de movimento humano, fomos ao WebCorp atrás de ocorrências nesse sentido para constatar sua frequência de uso e fazer uma breve comparação com o material encontrado no COMPARA.

3. Metodologia

Observando o método de Teubert⁴ coletamos no COMPARA as traduções feitas do verbo *to limp* em obras literárias. A pesquisa se deu até a segunda tradução do inglês, como vemos no esquema: texto original em inglês britânico, americano e sul-africano (TOI1) > texto traduzido em português do Brasil (TTPB1) > texto original em português do Brasil (TOPB2) > texto traduzido em inglês britânico, americano e sul-africano (TII2) > texto original em inglês britânico, americano e sul-africano (TOI3) > texto traduzido em português do Brasil (TTPB3). (Tabela 1)

Posteriormente tratamos os dados coletados no COMPARA (Tabelas 2 a 6) e confrontamo-los com os dados fornecidos pelo WebCorp.

⁴ Op. cit., p. 167-168.

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

4. Gráficos de Teubert⁵

4.1 Das traduções encontradas e pesquisadas:

→ Sentido da pesquisa					
TOI1	TTPB1	TOPB2	TTI2	TOI3	TTPB3
Limp	Ser manco	Mancar	Be lame	Lame	Sem muito ânimo
	Arrastar	Arrastar	Drag	Drag	Puxar
					Segurar
			Forçado a ir		
			Arrastar		
			Subir		
Haul	Haul	∅			
Crawl	Crawl	Andar como tartaruga			
Endless	Endless	Sem parar			
		Infinitas			
		Infinito			
Lethargic	Lethargic	∅			

Tabela 1 – Traduções encontradas e pesquisadas

Legenda:

TOI1 – texto original em inglês britânico, americano e sul-africano

TTPB1 – texto traduzido em português do Brasil

TOPB2 – texto original em português do Brasil

TTI2 – texto traduzido em inglês britânico, americano e sul-africano

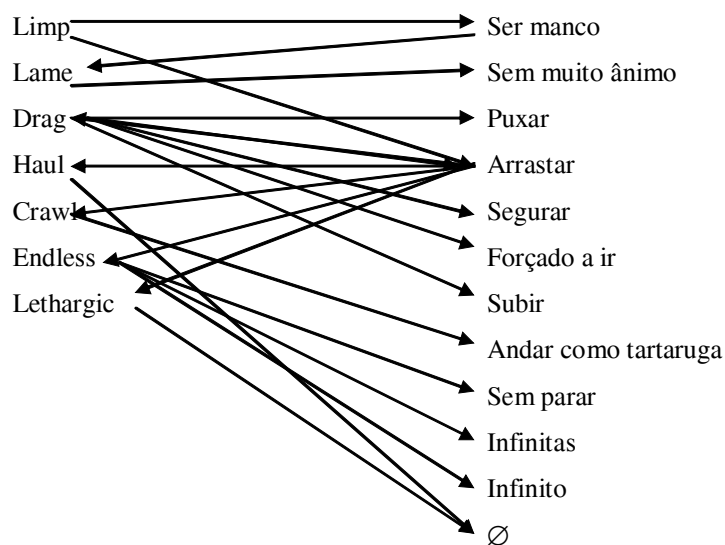
TOI3 – texto original em inglês britânico, americano e sul-africano

TTPB3 – texto traduzido em português do Brasil

∅ – nenhum exemplo foi encontrado

⁵ Op. cit., p.169.

4.2 Gráfico de Teubert⁶ com os termos escolhidos para análise:



5. Análise do verbo escolhido

5.1 Do verbo de movimento

O primeiro sentido esperado para o verbo *to limp* e suas traduções é o de verbo de movimento, mas como podemos observar nos exemplos abaixo, ora assume-o, ora não. O primeiro exemplo e sua tradução são verbos de movimento. No segundo exemplo notamos a ausência de movimento em português.

⁶ Op. cit., p.169.

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil.*

EBDL1 T2(149 4):	So Sally and her pals were well into their first set as we walked, or rather limped , past their court on our way to ours.	Por isso, Sally e suas amigas já estavam bem avançadas no primeiro <i>set</i> quando entramos, ou melhor, nos arrastamos na quadra, passando por aquela ocupada por elas.
PBPC1 (61):	Mesmo que os dias fossem todos iguais, com longas horas se arrastando entre o nascer e o pôr-do-sol; mesmo que elas jamais tivessem lido um só livro em suas curtas vidas, e não conhecessem a língua dos homens que contavam as novidades nas aldeias.	Yes, their days were all the same, with the seemingly endless hours between sunrise and dusk; and they had never read a book in their young lives, and didn't understand when the boy told them about the sights of the cities.

Observamos nas tabelas 2 e 3 que o verbo *to limp* e suas traduções, além de verbo de movimento, assumem outras características.

	Verbo de movimento	Não verbo de movimento	Nº total de ocorrências
Limp (TOI1)	2	0	2
Be lame (TTI2)	0	2	2
Drag (TTI2)	5	0	5
Haul (TTI2)	1	0	1
Crawl (TTI2)	1	0	1
Endless (TTI2)	0	1	1
Lethargic (TTI2)	0	1	1
Lame (TOI3)	0	1	1
Drag (TOI3)	5	0	5
Haul (TOI3)	0	0	0
Crawl (TOI3)	1	0	1
Endless (TOI3)	0	4	4
Lethargic (TOI3)	0	0	0
Nº total	15	9	24

Tabela 2 – Número de ocorrências com sentido de movimento em inglês

	Sentido próprio	Sentido metafórico	Nº total de ocorrências
Ser manco (TTPB1)	1	0	1
Arrastar (TTPB1)	0	1	1
Mancar (TOPB2)	2	0	2
Arrastar (TOPB2)	3	6	9
Sem muito ânimo (TTPB3)	0	1	1
Puxar (TTPB3)	1	0	1
Segurar (TTPB3)	0	1	1
Forçado a ir (TTPB3)	0	1	1
Arrastar (TTPB3)	0	1	1
Subir (TTPB3)	0	1	1
Andar como tartaruga (TTPB3)	0	1	1
Sem parar (TTPB3)	0	1	1
Infinitas (TTPB3)	1	1	2
Infinito (TTPB3)	0	1	1
Nº total	8	16	24

Tabela 3 – Número de ocorrências com sentido de movimento em português

5.2 Do sentido próprio e metafórico

Observamos que os termos pesquisados ocorrem com maior incidência em sentido metafórico. Isso pode ser percebido tanto nos textos originais como em suas traduções em ambas as línguas (Tabelas 4 e 5). Abaixo ilustramos o fenômeno, sentido próprio no primeiro exemplo e metafórico no segundo:

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil.*

EBDL 1 T2(149) 4):	So Sally and her pals were well into their first set as we walked, or rather limped , past their court on our way to ours.	Por isso, Sally e suas amigas já estavam bem avançadas no primeiro <i>set</i> quando entramos, ou melhor, nos arrastamos na quadra, passando por aquela ocupada por elas.
PBPC 1 (61):	Mesmo que os dias fossem todos iguais, com longas horas se arrastando entre o nascer e o pôr-do-sol; mesmo que elas jamais tivessem lido um só livro em suas curtas vidas, e não conhecessem a língua dos homens que contavam as novidades nas aldeias.	Yes, their days were all the same, with the seemingly endless hours between sunrise and dusk; and they had never read a book in their young lives, and didn't understand when the boy told them about the sights of the cities.

	Sentido próprio	Sentido metafórico	Nº total de ocorrências
Limp (TO11)	1	1	2
Be lame (TT12)	2	0	2
Drag (TT12)	3	2	5
Haul (TT12)	0	1	1
Crawl (TT12)	0	1	1
Endless (TT12)	0	1	1
Lethargic (TT12)	0	1	1
Lame (TO13)	0	1	1
Drag (TO13)	1	4	5
Haul (TO13)	0	0	0
Crawl (TO13)	0	1	1
Endless (TO13)	1	3	4
Lethargic (TO13)	0	0	0
Nº total	8	16	24

Tabela 4 – Número de ocorrências em sentido próprio e sentido metafórico em inglês

	Sentido próprio	Sentido metafórico	N° total de ocorrências
Ser manco (TTPB1)	1	0	1
Arrastar (TTPB1)	0	1	1
Mancar (TOPB2)	2	0	2
Arrastar (TOPB2)	3	6	9
Sem muito ânimo (TTPB3)	0	1	1
Puxar (TTPB3)	1	0	1
Segurar (TTPB3)	0	1	1
Forçado a ir (TTPB3)	0	1	1
Arrastar (TTPB3)	0	1	1
Subir (TTPB3)	0	1	1
Andar como tartaruga (TTPB3)	0	1	1
Sem parar (TTPB3)	0	1	1
Infinitas (TTPB3)	1	1	2
Infinito (TTPB3)	0	1	1
N° total	8	16	24

Tabela 5 – Número de ocorrências em sentido próprio e sentido metafórico em português

5.3 Da classe gramatical

Quanto à classe gramatical, observamos que o verbo *to limp* sofreu as alterações explicitadas na tabela abaixo:

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil.*

→ Sentido da pesquisa					
TOI1	TTPB1	TOPB2	TTI2	TOI3	TTPB3
Limp (2 v)	Manco (1 adj) Arrastar (1 v)	Mancar (2 v)	Lame (2 adj)	Lamely (1 adv)	Sem muito ânimo (prep+ adv+ s)
		Arrastar (8 v)	Drag (5 v)	Drag (3 v)	Puxar (1 v) Segurar (1 v) Subir (1 v)
				Drag (2 adj)	Forçado (1 adj) Arrastar (1 v)
			Haul (1 v)	Haul (Ø)	Ø
			Endless (1adj)	Endless (4 adj)	Sem parar (1 prep+ v) Infinitas (1 adj) Infinitas (1 s) Infinito (1 adj)
		Crawl (1 s)	Crawl (1 s)	Andar como uma tartaruga (1 exp v-nom)	
		Arrastado (1 adj)	Lethargic (1 adj)	Lethargic (Ø)	Ø

Tabela 6 – Classe gramatical

Legenda: v – verbo; s – substantivo; adj – adjetivo; adv – advérbio; prep – preposição; exp v-nom – expressão verbo-nominal; Ø – nenhum exemplo encontrado.

Notamos que na primeira tradução do verbo *to limp* ele assume papel de verbo e de adjetivo.

EBDL1T2 (540):	There's Joe, he's got serious back trouble, wears a corset all the time and can barely manage to serve overarm; Rupert, who was in a bad car crash a few years ago and limps with both legs, if that's possible; and Humphrey, who has arthritis in his feet and a plastic hip-joint.	Um é Joe, que tem problemas de coluna muito sérios, veste um colete o tempo todo e quase não consegue sacar com o braço para cima; tem Rupert, que sofreu um sério acidente de carro uns anos atrás e é manco das duas pernas, se isso for possível; e também Humphrey, com artrite nos pés e uma prótese no quadril.
EBDL1T2 (1494):	So Sally and her pals were well into their first set as we walked, or rather limped , past their court on our way to ours.	Por isso, Sally e suas amigas já estavam bem avançadas no primeiro <i>set</i> quando entramos, ou melhor, nos arrastamos na quadra, passando por aquela ocupada por elas.

Em seguida partimos para a análise da manutenção da mesma classe gramatical na tradução. Como citado anteriormente, notamos que a mudança de classe gramatical é diretamente proporcional ao uso da palavra em sentido figurado. Abaixo, um dos exemplos colhido no COMPARA:

EBDL1T2 (1180):	When <i>The People Next Door</i> was new and still finding its feet I used to attend rehearsals practically every day, but now it runs like a train (or like a train should run – this one has suddenly slowed to a crawl for some reason, and we haven't even got to Watford junction) and I just put in an appearance once a week to check that everything's going smoothly, and maybe do a little fine-tuning on the script.	Quando <i>Os vizinhos do lado</i> era coisa nova e ainda tentando, achar o tom, costumava assistir aos ensaios quase todos os dias, mas agora a coisa corre como um trem (ou como um trem deveria correr – este em que estou de repente diminuiu a velocidade e agora anda como uma tartaruga , sem razão nenhuma, e a gente ainda nem chegou à junção de Watford) e só vou dar uma espiada uma vez por semana para verificar se tudo está correndo nos trilhos e talvez afinar uns últimos pontos no <i>script</i> .
--------------------	--	--

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

Na alteração da classe gramatical visando à manutenção da convencionalidade em L2, observamos que a escolha tradutória entre palavra-palavra ou palavra-conjunto de palavras é afetada. A utilização da palavra em seu sentido metafórico dificulta a manutenção da relação palavra-palavra, como vemos no exemplo abaixo:

EBDL1T2 (1180):	When <i>The People Next Door</i> was new and still finding its feet I used to attend rehearsals practically every day, but now it runs like a train (or like a train should run – this one has suddenly slowed to a crawl for some reason, and we haven't even got to Watford junction) and I just put in an appearance once a week to check that everything's going smoothly, and maybe do a little fine-tuning on the script.	Quando <i>Os vizinhos do lado</i> era coisa nova e ainda tentando, achar o tom, costumava assistir aos ensaios quase todos os dias, mas agora a coisa corre como um trem (ou como um trem deveria correr – este em que estou de repente diminuiu a velocidade e agora anda como uma tartaruga , sem razão nenhuma, e a gente ainda nem chegou à junção de Watford) e só vou dar uma espiada uma vez por semana para verificar se tudo está correndo nos trilhos e talvez afinar uns últimos pontos no <i>script</i> .
--------------------	--	--

6. Outros Exemplos Coletados em Textos Originais

Iremos apresentar aqui o resultado da pesquisa no *WebCorp* para os verbos de movimento em inglês e em vernáculo citados neste trabalho com o objetivo de observarmos o comportamento gramatical e semântico que eles assumem em textos originais das duas línguas.

Ilustraremos os resultados com alguns exemplos.

Limp

O *WebCorp* apresentou 24 ocorrências. No entanto, nenhuma delas traz *to limp* como verbo de movimento. Dezesesseis ocorrências associam-se à doença que leva alguém a mancar...

Crop, 10, 2004

<http://www.emedicine.com/emerg/topic387.htm>

Document Dated: Unknown

Plain Text Word List 5240 tokens, 33 types

- Incidence of congenital conditions (eg, limp associated with congenital hip dysplasia
- is limited. Common causes of limp in the toddler are infections
- often the cause of a limp. Common causes of limping in

... e as outras relacionam-se com uma banda chamada *Limp Bizkit*.

<http://members.aol.com/davey713/lb.html>

Document Dated: Tue, 24 Nov 1998 04:39:31 GMT

Plain Text Word List 209 tokens, 11 types

- entire page for the best limp Bizkit info you'll find anywhere

Pesquisamos então com o verbo na terceira pessoa do presente. Obtivemos 73 ocorrências, das quais constatamos serem 16 de verbos de movimento (humano ou animal).

<http://old.smh.com.au/news/0109/23/sport/sport1.html>

Document Dated: Unknown

Plain Text Word List 841 tokens, 24 types

- an ad – LEAGUE TIPPING Joey limps but still runs wild Necked
- Newsletter: In this section Joey limps but still runs wild Why

<http://www.rte.ie/sport/2002/0510/rugby/munster.html>

Document Dated: Unknown

Plain Text Word List 253 tokens, 9 types

- beaten by Ulster as O'Connell limps off Filed at: Friday, May

Os outros exemplos são relacionados com umas das acepções possíveis de *to limp* (descreve o locomover de uma embarcação após um acidente)...

http://quickstart.clari.net/qs_se/webnews/wed/am/Qportugal-spain-ship.RJmX_CDI.html

Document Dated: Thu, 19 Dec 2002 19:05:44 GMT

Plain Text Word List 657 tokens, 20 types

- Unwelcome in Spain, fertiliser ship limps to Portuguese port Thursday, 19-Dec-2002

... ou estava ali em sentido figurado (algo/alguém/alguma situação difícil, que se arrasta).

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

<http://www.wsws.org/articles/2000/oct2000/elec-o21.shtml>

Document Dated: Unknown

Plain Text Word List 1987 tokens, 15 types

- presidential debate and beyond: Gore limps toward the finish line By

Drag

Ao consultarmos o *WebCorp* para *to drag* obtivemos 609 concordâncias; no entanto, nenhuma ocorrência como verbo de movimento. A maioria dos casos relaciona-se a *drag race* (“corrida entre dois carros para determinar quem acelera mais rápido a partir de um determinado ponto”, segundo o *The American Heritage Dictionary of Language* online).

<http://www.dragracing.se/>

Document Dated: Tue, 04 Feb 2003 06:34:30 GMT

Plain Text Word List 1765 tokens, 30 types

- here ! SFI inspection SBF:s drag racing committee announces: The annual
- annual SFI / SBF service for drag racing teams is held saturday
- fax/mail or call the drag racing committee tech director Rolf

... *drag and drop* (jargão da área de informática)...

<http://www.javaworld.com/javaworld/jw-03-1999/jw-03-dragndrop.html>

Document Dated: Unknown

Plain Text Word List 3962 tokens, 81 types

- NEW! Rational Resources How to drag and drop with Java 2
- Explore the Java platform's new drag and drop classes Summary Drag
- you have), you've already used drag and drop to transfer data
- transfer data using the familiar drag and drop (D&D) metaphor

...*drag queens/kings*....

<http://www.draglink.com/>

Document Dated: Tue, 28 Jan 2003 06:26:39 GMT

Plain Text Word List 2488 tokens, 15 types

- bearded women beauty boys Berlin drag kings Berlin drag queens big
- boys Berlin drag kings Berlin drag queens big sissies big sissy
- kings & goddesses drag kings and drag queens drag kings & drag queens
- drag kings and drag queens drag kings & drag queens drag kings
- and drag queens drag kings & drag queens drag kings for drag

Crop, 10, 2004

... e *drag washers* (sistema para limpeza de maquinário).

<http://www.smoothdrag.com/>

Document Dated: Thu, 17 Oct 2002 01:05:34 GMT

[Plain Text](#) [Word List](#) 614 tokens, 21 types

- an extensive selection of replacement drag washers for Penn, Penn International
- one set of 130 international drag washers super smoothies. Guide to
- an extensive selection of replacement drag washers for Penn, Penn International
- drags, offering the finest performing drag washer friction material available today

A seguir pesquisamos o verbo flexionado no *present continuous*. O WebCorp produziu 319 resultados. Localizamos apenas 14 exemplos na acepção de verbo de movimento...

<http://fun2send.com/dragging.php3>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 129 tokens, 12 types

- on a sidewalk. Both are dragging their right foot as they

Cerca de cento e cinqüenta ocorrências relacionam-se com a área de computação (“arrastar” alguma coisa com o vetor).

<http://www.mediamacros.com/item/item-930761938/>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 548 tokens, 22 types

- the sprite by clicking and dragging, magnify by double clicking, and

http://www.macromedia.com/support/fireworks/ts/documents/alt_key_dup_subselect.htm

Document Dated: Thu, 03 Jan 2002 18:17:52 GMT

[Plain Text](#) [Word List](#) 397 tokens, 14 types

- becomes activated. Then, when the dragging motion begins, instead of a
- can duplicate an object by dragging it while pressing down the
- before or after you start dragging the object. However, in one

Também constatamos cerca de 50 ocorrências relacionadas à área criminal e jurídica). Nesses exemplos *drag* converte-se em adjetivo e significa o ato de “arrastar alguém para a morte”; dessa maneira vai qualificando os diversos substantivos que podem ser colocados após ele: *dragging death/trial/defense/prosecutors*.

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

http://www.onlineathens.com/stories/091199/new_0911990037.shtml

Document Dated: Unknown

Plain Text Word List 431 tokens, 14 types

- no blacks will hear second dragging trial Secretary says Abernathy used
- no blacks will hear second dragging trial </MCC HEAD>
- white man charged in the dragging death of a black man

As outras ocorrências dividiram-se entre sentido de “conduzir” (na acepção de “arrastar”)...

<http://news.bbc.co.uk/1/hi/world/asia-pacific/1904654.stm>

Document Dated: Unknown

Plain Text Word List 909 tokens, 25 types

- GMT 10:39 UK Israel ‘dragging Mid-East into war’ Israel has
- Lumpur, have accused Israel of dragging the Middle East towards all-out
- international peace and security, and dragging the region towards an all-out

... fenômeno astrofísico...

<http://www.phys.lsu.edu/mog/mog10/node9.html>

Document Dated: Fri, 18 May 2001 19:05:35 GMT

Plain Text Word List 1150 tokens, 35 types

- a rotating mass. The gravitomagnetic “dragging of inertial frames” by rotating
- denotes the coefficient of frame dragging (1 in GR, in the

.... usos da expressão idiomática *drag one’s feet (or beel)* – ‘ser deliberadamente vagaroso ao fazer alguma coisa ou em tomar uma decisão’, segundo o dicionário inglês Oxford Advanced Learner’s.

http://www.worldnetdaily.com/news/article.asp?ARTICLE_ID=27621

Document Dated: Unknown

Plain Text Word List 1627 tokens, 27 types

Guide Movies Stocks Senate gallery ‘dragging feet’ on WND appeal Attorney

- he gallery appears to be “dragging its feet” and making a

http://www.imakenews.com/bermudacom/e_article000089645.cfm

Document Dated: Unknown

Plain Text Word List 1080 tokens, 19 types

Crop, 10, 2004

- Festival hits the rocks Government dragging feet over new Timesharing Act
- Pat Lacklustre Festival line-up Government dragging feet over new Timesharing Act
- Chris Gibbons Why is Government dragging its feet over much-needed revisions

Haul

Pesquisando por *haul* não encontramos nenhum exemplo de verbo de movimento nas 80 concordâncias produzidas pelo WebCorp. No *present continuous* obtivemos 68 ocorrências, sendo apenas 7 verbos de movimento.

<http://www.dragonfare.com/AL/ff/scribble1a.html>

Document Dated: Wed, 11 Sep 2002 22:06:30 GMT

[Plain Text](#) [Word List](#) 6997 tokens, 60 types

- Oh no.” She muttered, finally hauling herself up. “Not this.” She

Interessante notar que muitos dos exemplos apresentados pelo WebCorp eram a expressão *hauling butt/ass/carcass to somewhere* (mover-se até algum lugar).

<http://the-riotact.com/article.php?sid=1023>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 345 tokens, 12 types

- on Christmas Eve and then hauling my carcass to church at

<http://www.digitalmediatree.com/treehouse/comment/c1528/>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 148 tokens, 22 types

- can just picture the security hauling ass out of there. Bad

Crawl

O *WebCorp* gerou 155 concordâncias, sendo apenas 19 delas exemplos de verbo de movimento.

<http://www.babycenter.com/refcap/baby/babydevelopment/6501.html>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 1479 tokens, 26 types

- her own. In the traditional crawl, she'll first learn to balance
- develops Most babies learn to crawl — or an alternate way
- months. But some children never crawl, instead opting for bottom shuffling
- baby when she started to crawl? Younger than five months Five

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

A grande maioria remetia-se a um evento chamado *pub crawl*.

<http://www.camberleyonthe.net/pubcrawl.html>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 908 tokens, 43 types

- guide to the best pub crawl in camberley.... and trust us
- the rules of the pub crawl with your mates with out
- there has tried this pub crawl then let us know of

Encontramos também dois exemplos do uso figurado de *crawl* sobre as companhias de tabaco.

<http://www.pierrelemieux.org/artcrawl.html>

Document Dated: Sun, 02 Feb 2003 04:44:52 GMT

[Plain Text](#) [Word List](#) 877 tokens, 21 types

- the tobacco pariahs seem to crawl, the more they are trampled
- headquarters. Why do tobacco companies crawl so humbly? And why are

Com o verbo no *present continuous* foram apresentadas 264 concordâncias. Para verbos de movimento, detectamos 35 exemplos, sendo a maior parte deles relacionada ao “engatinhar” dos bebês.

<http://www.babycentre.co.uk/refcap/6501.html>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 1034 tokens, 17 types

- learn to go from a crawling position back into a sitting
- propel herself), or skips the crawling stage in favour of walking
- After your baby has mastered crawling, the only thing standing between

Também encontramos um uso em sentido figurado.

<http://www.gizmology.net/lovecraft/works/chaos.htm>

Document Dated: Sat, 11 May 2002 01:56:13 GMT

[Plain Text](#) [Word List](#) 3028 tokens, 15 types

- my mind. Slowly but inexorably crawling upon my consciousness and rising

Mais da metade das ocorrências remetiam, mais uma vez, à área de informática.

Crop, 10, 2004

<http://www.almaden.ibm.com/almaden/feat/www8/>

Document Dated: Unknown

Plain Text Word List 7205 tokens, 68 types

- crawling: a new approach to topic-specific
- up-to-date. To achieve such goal-directed crawling, we designed two hypertext mining
- different levels of specificity. Focused crawling acquires relevant pages steadily while

Arrastar

O *WebCorp* gerou 221 concordâncias, sendo menos de 10% delas relacionadas a “puxar (algo) atrás de si fazendo com que deslize pelo chão”, e nenhuma com sentido reflexivo.

<http://www.estadao.com.br/agestado/noticias/2001/out/18/95.htm>

Document Dated: Unknown

Plain Text Word List 572 tokens, 6 types

- de 2001 – 13h34 Al-Qaeda ameaça arrastar corpos de soldados dos EUA

http://www.pirunello.com.br/1_preescola.htm

Document Dated: Fri, 27 Dec 2002 03:30:25 GMT

Plain Text Word List 2104 tokens, 18 types

- virar de bruços, de costas, arrastar Tocar objetos de tamanhos e
- criança segurar objetos Brincar de arrastar, passar de deitado para assentado
- cheiros Explorar o meio ambiente: arrastar, rolar, engatinhar, girar, ter contato

Também encontramos quatro exemplos do verbo “arrastar” em sentido figurado, de deixar-se arrastar pelas emoções/dependências.

http://www.slaa.com.br/depoimentos/sindrome_de_abstinencia3.htm

Document Dated: Fri, 31 Aug 2001 04:25:49 GMT

Plain Text Word List 2479 tokens, 30 types

- desejo incontrolável de nos deixarmos arrastar por nossos padrões de comportamento
- Ao invés de deixarmos nos arrastar pela dependência, buscamos atividades que

Quase todo o restante dos exemplos relacionava-se ao termo *drag* na área de informática.

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

<http://www.usandoflash4.hpg.ig.com.br/html/14shortcuts9.html>

Document Dated: Sat, 04 Nov 2000 22:07:17 GMT

[Plain Text](#) [Word List](#) 236 tokens, 4 types

- a Barra de espaços e arrastar Pressionar a Barra de espaços
- a Barra de espaços e arrastar Ferramentas de carregamento rápido Mantenha

<http://www.centrodefotografia.com.br/cdfaq.asp>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 3151 tokens, 21 types

de ferramentas. Você pode também arrastar e soltar as imagens do aplicativos que aceitem imagens. Para arrastar e soltar a imagem em

Mancar

Quando pesquisamos por “mancar” o WebCorp forneceu 81 concordâncias. Cerca de 25% desse total traziam a gíria “se mancar” (aperceber-se)...

<http://www.papodearquibancada.hpg.ig.com.br/flamengo.htm>

Document Dated: Wed, 27 Nov 2002 01:20:43 GMT

[Plain Text](#) [Word List](#) 950 tokens, 9 types

- na hora de você se mancar, encerrar a carreira, cuidar de

... e “mancar” (não comparecer).

<http://www.cruiser.com.br/giria/>

Document Dated: Tuesday, 03-Dec-02 14:29:36 GMT

[Plain Text](#) [Word List](#) 1637 tokens, 28 types

- movimento artístico musical) legal! (ótimo!) mancar (desrespeitar compromisso) mini-saia (saia curta).

Todas as outras amostras eram relacionadas ao ato de claudicar (humano e animal).

<http://www.hipismobrasil.com.br/veterinaria/broca.asp?cat= Veterin%E1ria>

Document Dated: Unknown

[Plain Text](#) [Word List](#) 246 tokens, 19 types

- cavalo deve sentir dor e mancar. Como evitar: Neste caso a

http://www2.bananagames.com.br/pc_xenus-050902_not.asp

Document Dated: Unknown

[Plain Text](#) [Word List](#) 484 tokens, 15 types

- perna, o jogador começará a mancar e atrapalhar na locomoção. Ou

Crop, 10, 2004

Manco

Encontramos 64 concordâncias, em todos os casos como adjetivo ou substantivo masculino; nenhum em sentido figurado, mas muitos de forma pejorativa e como apelido.

<http://www.planetapiada.hpg.ig.com.br/papagaio.htm>

Document Dated: Sat, 11 May 2002 20:34:29 GMT

Plain Text Word List 541 tokens, 7 types

- e pergunta: E aquele papagaio manco ali? O papagaio responde: manco
- manco ali? O papagaio responde: manco é a sua avó, eu

<http://www.tipos.com.br/blog/index.php?itemid=8709>

Document Dated: Unknown

Plain Text Word List 3762 tokens, 32 types

- mais novos, mas eu era manco de uma perna. Poderia ingressar

<http://www.bethynha.com.br/o-manco.htm>

Document Dated: Sat, 04 Aug 2001 02:00:00 GMT

Plain Text Word List 720 tokens, 12 types

- que nunca tinha vez... era manco, e não podia bater na
- gente estava na captura desse manco há muito tempo. Ele é

7. Análise Comparativa entre os termos do COMPARA e os Coletados no *WebCorp*

Cruzando-se os dados colhidos no *WebCorp* com os obtidos no COMPARA notamos que:

a. das duas traduções propostas para *limp*, a que propõe “é manco” é a mais adequada, já que em textos originais em inglês o verbo *to limp* claramente denota “mancar”;

b. nas duas traduções de “mancava” apresentadas no COMPARA, percebe-se que o verbo “mancar” fora traduzido pelo substantivo masculino *lame*; outra solução, encontrada no *WebCorp*, seria recorrer ao verbo *limp* flexionado;

c. segundo nossa pesquisa no *WebCorp*, as cinco traduções de “arrastavam” pelo verbo *to drag* procedem. A tradução por *crawl*, embora não tenhamos encontrado nenhum exemplo no *WebCorp*, também é precisa já que uma das acepções desse verbo relaciona-se diretamente ao movimento lento de veículos;

d. no *WebCorp* constatamos que a utilização de *drag* em textos originais de língua inglesa associa-se a “arrastar”, por isso a tradução “segurar” não nos parece adequada.

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil*.

8. Conclusão

Verificamos assim que em um *corpus* paralelo bidirecional de textos originais e suas respectivas traduções, em língua inglesa e em português do Brasil, a bidirecionalidade dos termos não se mantém: do verbo inglês *to limp* chegamos à expressão “andar como tartaruga”, e até ao adjetivo “infinito”. O *corpus* paralelo mostrou-se uma importante ferramenta de auxílio na orientação da tradução, mas não infalível nem completa. Nosso trabalho corrobora as observações de Teubert⁷ (1999) de que o usuário de dicionários bilíngües não encontra nessa ferramenta um aliado seguro, bem como as pesquisas de Váradi e Kiss⁸ (2001) que apontam para as deficiências do *corpus* unidirecional e sugerem que o tradutor lance mão de seu conhecimento prévio da língua – acrescido ao bom-senso – no momento da tradução. Talvez com um *corpus* ou uma amostragem maiores possa-se afirmar que o conhecimento prévio do usuário não seja necessário. Observamos também que a opção pela manutenção ou não da classe gramatical, assim como a de sentido de movimento, deve-se à tentativa de manutenção da convencionalidade da língua de chegada, ou seja, a manutenção do “jeito de ser” de L2. Evita-se com isso o *translationese*⁹, fenômeno comum em traduções que incorrem na tentativa de manutenção de mesma ordem sintática ou classe gramatical de L1.

A utilização de *corpus* nas línguas originais dos termos pesquisados mostrou-se bastante produtiva. Foi possível observar o comportamento mais comum desses vocábulos, assim como notar as variações que eles podem assumir nos diversos contextos em que ocorrem.

Bibliografia

- EBELING, Jarle. 1998. Contrastive linguistics, translation, and parallel Corpora, *Meta*, XLIII, 4, p. 602-615.
- TAGNIN, Stella Ortweiler. 1989. *Expressões Idiomáticas e Convencionais*. São Paulo: Ática, Série Princípios.
- VÁRADI, Tamás e KISS, Gábor. 2001. Equivalence and Non-equivalence in Parallel Corpora, *International Journal of Corpus Linguistics* vol.6 (Special Issue), pp.167-177.

Endereços eletrônicos

COMPARA: www.linguateca.pt/ COMPARA

WebCorp: www.webcorp.org.uk

⁷ Op. cit., p.168.

⁸ Op. cit., p.168.

⁹ Ebeling (1998), p.606.

Crop, 10, 2004

Anexo I – As pesquisas no COMPARA

Procura inicial: limp.*. Pedido de: concordância em contexto. Corpus: De inglês para português.

Resultados da pesquisa

3 ocorrências.

Concordância

Procura: [word="limp.*" & _texto="E.*" & (_texto="EU.*" | _texto="PB.*" | _texto="EB.*" | _texto="ES.*") & (_variantetrad="americana" | _variantetrad="brasileira" | _variantetrad="britanica" | _variantetrad="sulafriicana")].

EBDLIT 2(540):	There's Joe, he's got serious back trouble, wears a corset all the time and can barely manage to serve overarm; Rupert, who was in a bad car crash a few years ago and limps with both legs, if that's possible; and Humphrey, who has arthritis in his feet and a plastic hip-joint.	Um é Joe, que tem problemas de coluna muito sérios, veste um colete o tempo todo e quase não consegue sacar com o braço para cima; tem Rupert, que sofreu um sério acidente de carro uns anos atrás e é manco das duas pernas, se isso for possível; e também Humphrey, com artrite nos pés e uma prótese no quadril.
EBDLIT 2(1494):	So Sally and her pals were well into their first set as we walked, or rather limped , past their court on our way to ours.	Por isso, Sally e suas amigas já estavam bem avançadas no primeiro <i>set</i> quando entramos, ou melhor, nos arrastamos na quadra, passando por aquela ocupada por elas.

Procura inicial: manc.*. Pedido de : concordância em contexto. Corpus: De português para inglês .

Resultados da pesquisa

15 ocorrências.

Concordância

Procura: [word="manc.*" & _texto="P.*" & (_texto="EU.*" | _texto="PB.*" | _texto="EB.*" | _texto="ES.*") & (_variantetrad="americana" | _variantetrad="brasileira" | _variantetrad="britanica" | _variantetrad="sulafriicana")].

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil.*

PBPC1 (407):	Ele conhecia em detalhes cada uma delas: sabia qual mancava , qual iria dar cria daqui a dois meses, e quais eram as mais preguiçosas.	He knew everything about each member of his flock: he knew which ones were lame , which one was to give birth two months from now, and which were the laziest.
PBPC1 (444):	Depois começou a examinar as seis ovelhas, e descobriu que uma mancava .	Then the old man began to inspect the sheep, and he saw that one was lame .

Procura inicial: arrast.*. Pedido de : concordância em contexto. Corpus: De português para inglês .

Resultados da pesquisa

9 ocorrências.

Concordância

Procura: [word="arrast.*" & _.texto="P.*" & (_.texto="EU.*" | _.texto="PB.*" | _.texto="EB.*" | _.texto="ES.*") & (_.variantetrad=" americana" | _.variantetrad="brasileira" | _.variantetrad="britanica" | _.variantetrad="sulafriicana")].

PBMA1 (659):	Se chegam visitas de alguma cerimônia, manda-o levar para dentro ou para baixo e, resistindo ele sempre, o espanhol toma-o a princípio com muita delicadeza, mas vinga-se daí a pouco, arrastando-o por uma orelha ou por uma perna, atira-o ao longe, e fecha-lhe todas as comunicações com a casa	If visitors of some importance arrive, he has him taken away or brought downstairs, and since he always resists, the Spaniard leads him quite carefully at first, but gets his revenge soon after, dragging him by an ear or a leg, flinging him and cutting him off from all communication with the house.
PBMA2 (277):	E, dizendo isto, travou-lhe dos pulsos e arrastou-a para fora; depois entrou e fechou-se.	And so saying Reason grabbed Folly by the wrists and dragged her outside.
PBMA2 (404):	E tu via isso, porque arrastava os olhos da compota para ele e dele para a compota, como a pedir-lhe que ma servisse; mas fazia-o em vão.	I could see that because I dragged my eyes away from the compote to him and then from him back to the compote, as if begging him to serve me some. But it was in vain.
PBMA2 (433):	Foi um estouro esta minha palavra; a estupefação imobilizou a todos; os olhos espriavam-se a uma e outra banda; trocavam-se sorrisos, segredos, à socapa, as mães arrastavam as filhas, pretextando o sereno.	Those words of mine were an explosion. Stupefaction immobilized everyone. Eyes looked out all over. Smiles were exchanged, furtive whispers. Mothers dragged their daughters off with the pretext of the dew.

PBMA2 (474):	Era boa moça, lépida, sem escrúpulos, um pouco tolhida pela austeridade do tempo, que lhe não permitia arrastar pelas ruas os seus estouvamentos e berlindas; luxuosa, impaciente, amiga de dinheiro e de rapazes.	She was a good girl, cheerful, without scruples, a little hampered by the austerity of the times, which wouldn't allow her to haul her flightiness and her gossip games through the streets, fond of luxury, impatient, a friend of money and young men.
PBMA3 (125):	Levantou-se com o passo vagaroso do costume, não aquele vagar arrastado se era dos preguiçosos, mas um vagar calculado e deduzido, um silogismo completo, a premissa antes da consequência, a consequência antes da conclusão.	He got up with his usual slow step: not the lethargic gait of a lazy man, but a logical, calculated slowness, a complete syllogism, the premise before the consequence, the consequence before the conclusion.
PBPC1 (61):	Mesmo que os dias fossem todos iguais, com longas horas se arrastando entre o nascer e o pôr-do-sol; mesmo que elas jamais tivessem lido um só livro em suas curtas vidas, e não conhecessem a língua dos homens que contavam as novidades nas aldeias.	Yes, their days were all the same, with the seemingly endless hours between sunrise and dusk; and they had never read a book in their young lives, and didn't understand when the boy told them about the sights of the cities.
PBRF1 (1651):	Empilhei todos os móveis que pude arrastar , menos a cadeira de rodas, em vários montes sucessivos.	I stacked in several successive piles all the pieces of furniture I could drag , except the wheelchair.
PBRF1 (1867):	Quando chegamos na rua Barata Ribeiro o trânsito começou a se arrastar lentamente.	When we got to Barata Ribeiro Street, the traffic slowed to a crawl .

Procura inicial: lam.*. Pedido de : concordância em contexto. Corpus: De inglês para português.

Resultados da pesquisa

4 ocorrências.

Concordância

Procura: [word="lam.*" & _.texto="E.*"& (_.texto="EU.*" | _.texto="PB.*" | _.texto="EB.*" | _.texto="ES.*") & (_.variantetrad="americana" | _.variantetrad="brasileira" | _.variantetrad="britanica" | _.variantetrad="sulaficana")].

HANNA, Kátia; HILST, Tereza Cristina. *Uma Análise Contrastiva Preliminar do Verbo Limp em Inglês e suas Traduções em Português do Brasil.*

EBDLIT 2(1314):	"It just came in the post," I said, gesturing lamely with the balloon.	-- Acabou de chegar pelo correio -- respondi, gesticulando sem muito ânimo com o balão nas mãos.
--------------------	---	---

Procura inicial: drag.*. Pedido de : concordância em contexto. Corpus: De inglês para português.

Resultados da pesquisa

5 ocorrências.

Concordância

Procura: [word="drag.*" & _.texto="E.*" & (_.texto="EU.*" | _.texto="PB.*" | _.texto="EB.*" | _.texto="ES.*") & (_.variantetrad="americana" | _.variantetrad="brasileira" | _.variantetrad="britanica" | _.variantetrad="sulaficana")].

EBDLIT 2(537):	I have to sort of drag the right leg like Long John Silver when I hop around the court, but it's better than nothing.	Tenho de puxar a perna como o pirata Long John Silver quando dou viradas pela quadra, mas é melhor do que não fazer nada.
EBDLIT 2(779):	I practically dragged him from his seat by the lapels, like people do in movies.	Praticamente o segurei pelo colarinho e levantei-o da cadeira, como fazem nos filmes.
EBDLIT 2(1962):	Much later, when I was a teenager, I used to attend a Catholic Youth Club, because Maureen Kavanagh was a Catholic and belonged to it; and occasionally I would get trapped or dragged in to some kind of service on Sunday evenings, a recitation of the rosary in the parish hall, or something they called Benediction in the church next door, a funny business with a lot of hymn-singing in Latin and clouds of incense and the priest on the altar holding up something like a gold football trophy.	Muito mais tarde, quando eu era adolescente, costumava freqüentar um grupo jovem da Igreja católica, porque Maureen Kavanagh era católica e pertencia ao grupo; e de vez em quando eu era forçado a ir a algum tipo de missa nos domingos à noite, uma recitação de um rosário na igreja paroquial, ou algo que eles chamavam de Benedictus na própria igreja, um negócio engraçado com muita cantoria de hinos em latim e nuvens de incenso e o padre no altar segurando uma coisa dourada que parecia um troféu de futebol.
EBDLIT 2(1134):	Just when he had decided that he would have to take the initiative by ambushing his British colleagues as they passed his door at coffee-time and dragging them into his office, they began to acknowledge his presence in a way which suggested long but not deep familiarity, tossing him a perfunctory smile as they passed, or nodding their heads, without breaking step or their own conversations.	Justamente quando já havia decidido que teria de tomar a iniciativa e ficar de tocaia junto à porta para abordar seus colegas britânicos no momento em que passassem na hora do café e arrastá- los à sua sala, começaram a notar sua presença de uma maneira que sugeria longa, mas não profunda, familiaridade, abrindo um sorriso ao passar, ou cumprimentando-o com a cabeça, sem diminuir o passo ou interromper a conversa com o colega ao lado.

Crop, 10, 2004

EBDL3T 2(1460):	Through the murk the dull red eye of a sun that had scarcely been able to drag itself above roof level all day was sinking bleakly beneath the horizon, spreading a rusty stain across the snow-covered surfaces.	O esboço avermelhado que bem poderia ser o sol, que quase não tinha se dado ao esforço de subir acima do nível dos telhados o dia inteiro, estava agora afundando no horizonte, espalhando uma mancha enferrujada acima das telhas cobertas de neve.
--------------------	--	--

Procura inicial: haul.*. Pedido de : concordância em contexto. Corpus: De inglês para português.

Resultados da pesquisa

Procura: [word="haul.*" & _texto="E.*" & (_texto="EU.*" | _texto="PB.*" | _texto="EB.*" | _texto="ES.*") & (_variantetrad="americana" | _variantetrad="brasileira" | _variantetrad="britanica" | _variantetrad="sulaficana")].

Lamentamos, mas não foi encontrado nenhum caso!

Procura inicial: "endless". Pedido de : concordância em contexto. Corpus: De inglês para português.

Resultados da pesquisa

4 ocorrências.

Concordância

Procura: [word="endless" & _texto="E.*" & (_texto="EU.*" | _texto="PB.*" | _texto="EB.*" | _texto="ES.*") & (_variantetrad="americana" | _variantetrad="brasileira" | _variantetrad="britanica" | _variantetrad="sulaficana")].

LOFFREDO, Laís; GROSSMAN, Deborah; BITAR, Gladys; GONÇALVES, Janice. *Verbos de Elocução – As Diferenças entre o Inglês e o Português*.

EBDLIT 2(732):	Is it a substitute for sex, I wonder, that endless fondling of the knob on the end of the gear lever, that perpetual pumping of the clutch pedal?	Será um substituto para o sexo, me pergunto, esse jeito de acariciar sem parar a bola da alavanca do câmbio, esse movimento constante de ficar pisando o pedal da embreagem?
EBDLIT 2(1067):	Perhaps it's the eager, puppyish, tail-wagging energy he seems to have such endless reserves of.	Talvez seja sua energia de cachorrinho balançando o rabo, sempre querendo agradar, energia da qual parece ter reservas infinitas.
EBDL3T 2(508):	She continues to gaze out of the window, though there is nothing to see except cloud, stretching to the horizon like an endless roll of roof insulation.	Ela continua encarando a janelinha, embora não haja nada para se ver, a não ser nuvens, que se alongam no horizonte como um lençol infinito.
EBDL3T 2(691):	He felt a particularly pressing need to castigate naïve theories of realism because they threatened his masterwork: obviously, if you applied an open-ended system (life) to a closed one (literature) the possible permutations were endless and the definitive commentary became an impossibility.	Sentia uma necessidade incontrolável de castigar teorias ingênuas de realismo, porque elas ameaçavam sua obra principal: obviamente, se você aplicar um sistema infinito e aberto (a vida) a um fechado (a literatura) as possíveis inter-relações são infinitas e a crítica perfeita se tornaria uma impossibilidade.

Procura inicial: crawl.*. Pedido de : concordância em contexto. Corpus: De inglês para português.
Resultados da pesquisa
Uma ocorrência.

Concordância

Procura: [word="crawl.*" & _texto="E.*" & (_texto="EU.*" | _texto="PB.*" | _texto="EB.*" | _texto="ES.*") & (_variantetrad="americana" | _variantetrad="brasileira" | _variantetrad="britanica" | _variantetrad="sulafricana")].

EBDL IT2(1180):	When <i>The People Next Door</i> was new and still finding its feet I used to attend rehearsals practically every day, but now it runs like a train (or like a train should run – this one has suddenly slowed to a crawl for some reason, and we haven't even got to Watford junction) and I just put in an appearance once a week to check that everything's going smoothly, and maybe do a little fine-tuning on the script.	Quando <i>Os vizinhos do lado</i> era coisa nova e ainda tentando, achar o tom, costumava assistir aos ensaios quase todos os dias, mas agora a coisa corre como um trem (ou como um trem deveria correr – este em que estou de repente diminuiu a velocidade e agora anda como uma tartaruga, sem razão nenhuma, e a gente ainda nem chegou à junção de Watford) e só vou dar uma espiada uma vez por semana para verificar se tudo está correndo nos trilhos e talvez afinar uns últimos pontos no <i>script</i> .
--------------------	--	--

Verbos de Elocução – As Diferenças entre o Inglês e o Português

*Laís Loffredo**
*Deborah Grossman**
*Gladys Bitar**
*Janice Gonçalves**

Resumo: *A tradução de verbos de elocução pode representar uma tarefa árdua, que requer observação atenta. O presente estudo pretende contrastar o uso dos verbos de elocução em diálogos extraídos de textos em inglês e português. Além disso, tem por objetivo analisar a forma como os tradutores ocupam-se deles.*

Palavras-chave: *verbos de elocução; textos originais; textos traduzidos; resultados.*

Abstract: *The translation of verbs of saying may prove to be a difficult task demanding careful observation. The present paper proposes to contrast the use of verbs of saying in dialogues from texts in English and Portuguese. It also aims at analyzing the way translators deal with them.*

Keywords: *verbs of saying; original texts; translations; results.*

* Ex-alunas do Curso de Especialização em Tradução *lato sensu* da Universidade de São Paulo.

I. Introdução

A - Definição do problema:

Após uma análise, durante uma aula de Linguística Contrastiva II, dos tópicos gramaticais e lingüísticos que poderiam oferecer dificuldades para os tradutores foi enumerada uma série de assuntos entre os quais nos coube pesquisar os verbos de elocução utilizados em diálogos. Essa proposta surgiu da percepção da diferença no uso destes verbos entre o inglês e o português. Suspeitávamos que no inglês não havia muita variação na escolha do léxico, enquanto que no português a variação era quase obrigatória, para tornar o texto mais compreensível e natural.

B – Objetivos do Trabalho:

- Determinar as diferenças com base em textos autênticos, escritos originalmente nas duas línguas.
- Sugerir o que se deve usar na variação para a tradução de cada caso.

C - Metodologia do Trabalho:

Apenas com o intuito de confirmar uma hipótese, iniciamos nosso trabalho com um pequeno levantamento em romances em inglês e português, bem como em sites de jornais nacionais e internacionais, dos verbos de elocução mais utilizados em cada idioma. Verificamos que o número de verbos de elocução obtidos em português excedia bastante o número de verbos de elocução em inglês. Também constatamos a grande ocorrência do verbo **say** nos textos originais em inglês. Em seguida, partimos para a pesquisa em si, utilizando como material o corpus eletrônico COMPARA, um corpus paralelo composto por textos literários originais e traduzidos em inglês e português.

A busca inicial foi feita do inglês para o português através da fórmula (“,” “>>” “*he*” ou “*she*”), que resultou numa grande variedade de verbos, sendo que o verbo **say**, quase sempre no passado, apresentou mais uma vez o maior número de ocorrências. Por esta razão, fizemos uma busca por **said**, tanto de textos originais para tradução, quanto de textos traduzidos para originais, para obtermos uma lista de verbos correspondentes em português. O procedimento seguinte foi exatamente o caminho inverso, ou seja, cada resultado obtido foi novamente pesquisado, desta vez do português para o inglês. Esse procedimento foi repetido diversas vezes na tentativa de esgotar o maior número possível dos verbos de elocução pesquisados. Uma dificuldade que encontramos foi que

na busca pelos verbos de elocução isolados, obtivemos um número bastante grande de resultados, sendo que em muitos deles o verbo não apareceu em diálogo. Foi necessário, então, fazer uma espécie de filtro para que selecionássemos apenas os casos que nos interessavam analisar.

II. Resultados da pesquisa

Neste trabalho exporemos apenas a pesquisa com os textos literários do COMPARA, deixando a análise de textos jornalísticos para uma próxima oportunidade.

Convém mencionarmos que nossa busca por, »*he said* e, »*she said* apesar de muito prática, já que nos forneceu apenas as ocorrências em diálogos, restringiu demais os resultados que, por razões óbvias, não apresentaram casos com o nome da personagem (»Melissa said), por exemplo. Iniciaremos esse trabalho, portanto mostrando a busca apenas por *said* nos dois sentidos possíveis, um de texto original para tradução e o outro de texto traduzido para texto original a fim de analisar as traduções possíveis para o verbo *say* e a frequência de cada uma delas, bem como os verbos originalmente em português cuja tradução é *say* e a frequência de cada um deles.

Para melhor visualização, exporemos os resultados dessa primeira etapa do trabalho em uma tabela dividida em duas colunas distintas como segue:

Etapa 1:

Say (Said)

originais para tradução - 174 (879)

tradução para originais - 162 (527)

Dizer (49%)
Responder (14,4%)
Sem verbo (9%)
Perguntar (8,6%)
Comentar
Exclamar
Explicar
Concluir
Declarar
Insistir
Confirmar

Dizer (91%)
Sem verbo (3%)
Falar (2,5%)
Responder
Perguntar
Insistir
Refletir

Na tabela acima, os números entre parênteses (879 e 527) correspondem aos totais de ocorrências para o verbo *say* no passado. Como nos interessam apenas os casos em que o verbo de elocução aparece em diálogos e nossa busca

apresentou um número muito grande de ocorrências, analisamos as primeiras duzentas ocorrências de cada total e obtivemos o número fora dos parênteses (174 e 162) que representam os totais de **said** em diálogos.

A primeira coisa a se notar na primeira coluna da tabela foi que na maior parte das vezes em que **say** foi traduzido para o português optou-se pelo verbo **dizer** e raramente um verbo diferente de **dizer**, **perguntar** ou **responder** ocorreu nas traduções.

Ao observarmos a segunda coluna, onde **say** é a tradução de algum verbo em português, percebemos que quase sempre esse verbo é **dizer**, ou seja, quando o tradutor se depara com outro verbo diferente de **dizer**, praticamente não utiliza **say**, traduzindo-o por um equivalente mais próximo da língua de origem, mas que talvez não soe tão natural na língua de chegada.

Partimos, então, para a segunda etapa de nosso trabalho, que foi buscar cada um dos resultados do verbo **say** a fim de analisar o caminho inverso, do português para o inglês. Começamos pelo verbo **dizer**, campeão de ocorrências na primeira etapa do trabalho. Como os verbos de elocução em diálogos ocorrem geralmente no pretérito perfeito, fizemos uma busca por **disse** a fim de obtermos o maior número de resultados possíveis:

Etapa 2:

2a) Dizer

originais para tradução - 112 (465)

tradução para originais - 176 (226)

Say (86,6%)
Sem verbo (3,6%)
Tell (3,6%)
Ask (2,6%)
Murmur
Mutter

Say (96%)
Add
Remark
Enquire
Sem verbo
Tell

A partir da comparação do número total de ocorrências (entre parênteses) do verbo **dizer** com o número total de ocorrências do verbo **say**, pudemos concluir que **say** é muito mais usado em textos originais em inglês do que **dizer** em português.

Também verificamos que a tradução de **dizer** é, em praticamente todos os casos, **say** e que quando **dizer** é traduzido, em praticamente todos os casos o verbo original em inglês é **say**, o que não nos surpreendeu. Os resultados das buscas pelos verbos **tell**, **ask**, **murmur** e **mutter**, e dos verbos **add**, **remark** e **enquire** encontram-se na etapa 3 de nosso trabalho.

Crop, 10, 2004

Concluída a pesquisa de **dizer**, chegamos ao verbo **responder**, a segunda opção mais freqüente nas traduções de **say**: A pesquisa de **responder**, como de todos os outros verbos em português com exceção de **dizer** foi feita da seguinte forma: **respond.***. Desse modo, tivemos acesso à ocorrências tanto com **respondi** quanto com **respondeu**, além de alguns poucos casos em que o verbo não apareceu no pretérito perfeito:

2b) Responder

originais para tradução - 50 (73)

Reply (54%)
Answer (24%)
Say (12%)
Sem verbo (10%)

tradução para originais - 18 (37)

Say (77,8%)
Reply (11,1%)
Sem verbo (11,1%)

Ao compararmos o número de textos originais com o número de textos traduzidos em que o verbo **responder** é usado foi possível observar que esse verbo é muito mais freqüente em diálogos em português que seus equivalentes em inglês. Já sabemos que em inglês é bastante comum utilizar **say** para casos em que verbos variados, como **responder**, são utilizados em português. Como existe uma grande tendência em se traduzir **say** por **dizer**, já esperávamos um menor número de ocorrências na segunda coluna da tabela, em que **responder** é a tradução de um determinado verbo em inglês.

Além disso, é bastante evidente o fato de **answer** só estar presente na primeira coluna da tabela, o que indica que embora seja uma opção freqüente para a tradução de **responder**, não representa uma escolha ideal, já que, através dos resultados da segunda coluna, levantamos a hipótese de que esse verbo não é muito usado em textos originais em inglês, o que confirmaremos na terceira etapa do trabalho quando chegarmos aos resultados da pesquisa por **answer**.

O verbo **reply**, apesar de ocorrer em textos originais em inglês, não é tão utilizado quanto **say**, que deveria ser a opção mais freqüente para as traduções de **responder** se a naturalidade da língua-alvo fosse levada em consideração.

O terceiro verbo em número de ocorrências que equivale à tradução de **say**, como indica a tabela da etapa 1, é **perguntar**. Realizamos a pesquisa por **pergunt.***, da mesma forma que fizemos com **responder** e chegamos aos seguintes resultados:

2c) Perguntar

originais para tradução - 96 (141)

Ask (93%)
Sem verbo (5%)
Say (2%)

tradução para originais - 49 (117)

Say (37%)
Ask (31%)
Demand (8%)
Enquire (6%)
Wonder
Reply
Conclude
Gape

É interessante notar, através dos resultados da primeira coluna, que embora pouco utilizado em textos originais em inglês, como confirmaremos mais adiante (ver tabela de **ask** na etapa 3), o verbo **ask** é a escolha quase que unânime nas traduções de **perguntar**. **Say** é raramente lembrado, enquanto que outros verbos como **enquire** e **demand**, que também aparecem em textos originais, são totalmente ignorados pela tradução.

Se compararmos as 96 ocorrências de textos originais com as 49 ocorrências de textos traduzidos, podemos levantar a hipótese de que outros verbos, entre os quais provavelmente **dizer**, estão sendo usados nas traduções de casos em que **perguntar** seria a opção mais acertada. Já os 37% de ocorrências para **say** na segunda coluna parecem indicar que, no caso de **perguntar** e de seus equivalentes em inglês, as traduções do inglês para o português são menos literais e, portanto, a naturalidade da língua-alvo se mantém um pouco mais, enquanto que as traduções do português para o inglês são mais literais e, portanto, a naturalidade da língua alvo se mantém menos.

Os verbos **comentar**, **exclamar**, **explicar**, **concluir**, **declarar**, **insistir** e **confirmar** apareceram poucas vezes como tradução de **say**. Porém, seguimos a mesma orientação de **dizer**, **responder** e **perguntar** também com esses verbos e exporemos agora seus resultados:

2d) Comentar

originais para tradução - 1 (7)

Explain

tradução para originais - 7 (37)

Say (57%)
Comment (43%)

Crop, 10, 2004

2e) Exclamar

originais para tradução - 7 (7)

tradução para originais - 0

Exclaim (72%)

Bellow (14%)

Shout (14%)

2f) Explicar

originais para tradução - 4 (46)

tradução para originais - 3 (31)

Explain

Explain (66,6%)

Say (33,4%)

2g) Concluir

originais para tradução - 5 (13)

tradução para originais - 3 (11)

Conclude (60%)

End (20%)

Add (20%)

Conclude

2h) Declarar

originais para tradução - 1 (7)

tradução para originais - zero (3)

Declare

2i) Insistir

originais para tradução - 4 (13)

tradução para originais - 0 (4)

Insist (50%)

Say (25%)

Ask (25%)

2j) Confirmar

originais para tradução - 1 (4)

tradução para originais - 0 (12)

Agree

É possível observar que não há muitas ocorrências para os verbos acima em diálogos e que nas ocorrências existentes suas traduções são bastante diretas. Em raras exceções, o texto traduzido não utilizou um equivalente exato ao texto original, ou seja, a tradução de **exclamar** é quase sempre **exclaim**, a de **explicar** é sempre **explain** etc.

Percebemos também que a busca por textos traduzidos com esses verbos quase não apresentou resultados. Acreditamos que isso ocorra porque os tradutores somente utilizam **explicar** ou **declarar** em suas traduções quando o texto original é **explain** ou **declare** e esses verbos de elocução não são muito comuns em diálogos de textos em inglês.

Com isso, encerramos a pesquisa dos verbos que apareceram como tradução de **say** (primeira coluna). O verbo **falar** apareceu em nossa pesquisa como original da tradução **say** (segunda coluna), juntamente com **dizer**, **responder**, **perguntar** e **insistir**, que já foram pesquisados, e com **refletir**, que possivelmente representa os originais dos casos em que a tradução é **say to oneself**. Fizemos o caminho inverso de nossa pesquisa apenas com **falar** e **refletir**, por serem essas as únicas ocorrências inéditas até o momento:

2k) Falar

originais para tradução - 17 (42)

tradução para originais - 2 (151)

Say (65%)

Say

Speak (35%)

Come out

2l) Refletir

originais para tradução - 3 (7)

tradução para originais - 0 (3)

Think (to oneself) (66,6%)

Say (to oneself) (33,4%)

Embora seja um verbo bastante comum, **falar** é mais usado no discurso indireto, em que a fala da personagem não é reproduzida tal como ela a teria organizado e emitido. Nos casos em que **falar** ocorreu em discurso direto, apenas os verbos **say** e **speak** foram usados na tradução. **Refletir** também apresentou poucas ocorrências provando ser um verbo pouco usado em textos literários, principalmente em comparação a **dizer**, **responder** e **perguntar**.

Depois de pesquisar cada uma das ocorrências que representam os equivalentes ao verbo **say** em português, podemos iniciar agora a terceira etapa do trabalho que consiste na busca de cada um dos resultados de **dizer**, **responder**,

Crop, 10, 2004

perguntar e outros que sejam diferentes de **say**. O primeiro verbo, diferente de **say** na lista das traduções de **dizer** (indicado pela letra **a** na tabela 2 foi **tell**, seguido por **ask**, **murmur** e **mutter** (verbos que também indicaremos por **a**):

Etapa 3:

3a) Tell (Told)

originais para tradução - 2 (121)

tradução para originais - 12 (137)

Dizer

Dizer

3a) Ask

originais para tradução - 9 (44)

tradução para originais - 95 (154)

Perguntar (89%)

Protestar

Perguntar (90,5%)

Dizer (3,2%)

Sem verbo

Indagar

Insistir

Pedir

Continuar

Fazer uma pergunta

Assim como **falar** no português, os verbos **tell** e **ask**, apesar de muito usados no inglês, não aparecem com muita frequência em diálogos de textos originais. No caso de **tell**, fizemos a busca por **told** e, nos únicos dois casos em que foi usado em diálogo, sua tradução foi **dizer (a alguém)**. Também nos textos traduzidos com **dizer + objeto indireto (a alguém)**, **tell** é a escolha feita na tradução.

A busca por **ask** também foi feita no passado (**asked**) e apresentou apenas nove ocorrências em diálogos, confirmando o que já havíamos deduzido com a pesquisa do verbo **perguntar** na etapa 2. Se em textos originais, **ask** é pouco usado, em textos traduzidos corresponde à quase todas as traduções do verbo **perguntar** que, como já vimos, é bastante frequente em diálogos de textos literários em português. Curiosamente, o verbo **dizer** também apareceu como original de **ask**. É bastante raro **dizer** ser usado após perguntas em português, porém alguns autores como Machado de Assis e Rubem Fonseca utilizaram essa 0de **perguntar** em algumas obras.

LOFFREDO, Laís; GROSSMAN, Deborah; BITAR, Gladys; GONÇALVES, Janice. *Verbos de Elocução – As Diferenças entre o Inglês e o Português*.

3a) Murmur

originais para tradução - 1 (2)

Murmurar

tradução para originais - 7 (9)

Murmurar (85,7%)

Dizer (14,3%)

3a) Mutter

originais para tradução – 1 (1)

Murmurar

tradução para originais – 2 (3)

Murmurar

Murmur e **mutter**, como podemos ver, são verbos de pouquíssimo uso e o correspondente de ambos em português é **murmurar**.

Os verbos **add**, **remark** e **enquire** abaixo foram tirados da segunda coluna da pesquisa por **dizer** onde estão representando, juntamente com **say** e **tell** os originais para essa tradução:

3a) Add

originais para tradução – 25 (38)

Acrescentou

tradução para originais – 6 (24)

Acrescentou (50%)

Emendar

Acudir

Concluir

3a) Remark

originais para tradução – 5 (6)

Dizer (60%)

Comentar (20%)

Observar (20%)

tradução para originais - 0

3a) Enquire

originais para tradução – 7 (10)

Perguntar (57%)

Inquirir (28%)

Querer saber (15%)

tradução para originais - 0

Como suspeitávamos **remark** e **enquire** não aparecem em textos traduzidos para o inglês. No caso de **enquire**, já sabemos que quando o português utiliza o verbo **perguntar**, a tradução para o inglês opta quase sempre por **ask** e quando o português utiliza **dizer**, a tradução para o inglês opta quase sempre por **say**.

Add também é mais comum em textos originais do que em textos traduzidos e sua tradução é invariavelmente **acrescentar**. Os verbos em português que traduzidos para o inglês resultam em **add** (segunda coluna) variam muito pouco, sendo na maior parte dos casos também **acrescentar**.

Analisaremos em seguida a pesquisa por **reply** e **answer** (que indicaremos pela letra b), resultados de **responder** (tabela 2b) que não se repetem na pesquisa de **dizer**.

3b) Reply

originais para tradução - 4 (8)

Replicar (50%)
Responder (25%)
Perguntar (25%)

tradução para originais - 35 (58)

Responder (63%)
Replicar (26%)
Retorquir (5%)
Retrucar
Acudir

3b) Answer

originais para tradução - 8 (19)

Responder (87%)
Concordar (13%)

tradução para originais - 20 (46)

Responder (90%)
Replicar (5%)
Redargüir (5%)

Os verbos **reply** e **answer**, como já esperávamos, quase não aparecem em textos originais, o que nos leva a crer que nos casos em que poderiam ser usados, **say** aparece em seus lugares. Em textos traduzidos, **reply** e **answer** são mais freqüentes pois representam uma tradução literal de **responder** e outros verbos como **replicar** em português.

Os resultados inéditos de **perguntar**, segundo a tabela 2c, são: **demand**, **wonder**, **conclude** e **gape**. A seguir, exporemos o resultado da pesquisa por esses verbos.

LOFFREDO, Laís; GROSSMAN, Deborah; BITAR, Gladys; GONÇALVES, Janice. *Verbos de Elocução – As Diferenças entre o Inglês e o Português*.

3c) Demand

originais para tradução - 4 (5)

tradução para originais - 0 (4)

Perguntar

3c) Wonder

originais para tradução 1 (19)

tradução para originais - 0 (1)

Perguntar (a si mesmo)

3c) Conclude

originais para tradução - 3 (3)

tradução para originais - 3 (3)

Concluir (66,7%)

Perguntar (33,3%)

Concluir

3c) Gape

originais para tradução - 1 (1)

tradução para originais - 0

Perguntar (boquiaberto)

Como podemos perceber o número de ocorrências para os verbos acima não é muito significativo e, com exceção de **conclude**, eles só aparecem em textos originais cujas traduções são invariavelmente **perguntar**. No caso de **conclude**, há três ocorrências de textos traduzidos cujos originais em português utilizaram **concluir** e dos três textos em que **conclude** foi traduzido, **concluir** foi utilizado em dois.

O verbo **explain** foi utilizado na tradução da única ocorrência resultante da busca por **comentar** (2d). Os verbos em textos originais cuja tradução foi **comentar** foram **say** e **comment**. Vejamos agora os resultados de **explain** e **comment**, verbos obtidos a partir da pesquisa de **comentar**:

3d) Explain

originais para tradução - 4 (24)

tradução para originais - 5 (33)

Explicar

Explicar (80%)

Retrucar (20%)

Crop, 10, 2004

3d) Comment

originais para tradução - 4 (5)

tradução para originais - 0 (1)

Comentar

Explain apareceu apenas quatro vezes em diálogos de textos originais e em todas elas sua tradução para o português foi **explicar**. Embora **explicar** tenha sido a principal ocorrência para os cinco textos encontrados cuja tradução foi **explain**, também obtivemos ocorrências com **retrucar**, verbo que, acreditamos, dificilmente apareceria em um texto traduzido.

No caso de **comment**, apenas textos originais foram encontrados e a tradução de **comment** foi **comentar** em todos os casos.

As três traduções encontradas para o verbo **exclaimar** segundo a tabela 2e foram **exclaim**, **bellow** e **shout**. Esses verbos apresentaram os seguintes resultados:

3e) Exclaim

originais para tradução - 9 (10)

tradução para originais - 7 (7)

Exclamar

Exclamar (86%)

Comentar (14%)

3e) Bellow

originais para tradução - 0

Tradução para originais - 1 (3)

Exclamar

3e) Shout

originais para tradução - 5 (9)

tradução para originais - 13 (15)

Gritar

Gritar (58%)

Bradar (34%)

Exclamar (8%)

Fica claro, através da análise dos resultados, que os verbos acima são mais usados em textos traduzidos que em textos originais e que seus equivalentes em português, apesar de variarem, são quase sempre um verbo com sentido bastante próximo do inglês.

O verbo **explicar** (tabela 2f) não apresentou ocorrências inéditas, portanto, passaremos para a análise dos resultados obtidos na pesquisa do verbo **end**, resultado inédito de **concluir** (tabela 2g).

LOFFREDO, Laís; GROSSMAN, Deborah; BITAR, Gladys; GONÇALVES, Janice. *Verbos de Elocução – As Diferenças entre o Inglês e o Português*.

3g) End

originais para tradução - 0 (14)

tradução para originais - 1 (20)

Concluir

O verbo **end**, como podemos ver, não apresentou ocorrências em diálogos de textos originais em inglês e na única ocorrência de texto traduzido, sua original era **concluir**. Apresentaremos agora o resultado da busca por **declare**, única ocorrência encontrada na pesquisa de **declarar**, representada na tabela 2h:

3h) Declare

originais para tradução - 2 (10)

tradução para originais - 1 (3)

Declarar

Declarar

Apesar de ambos apresentarem poucas ocorrências em diálogos de textos literários, é possível concluir que assim como a tradução de **declarar** é **declare** (tabela 2h), também a tradução de **declare** é **declarar**.

O verbo seguinte a **declarar** na etapa 2 do trabalho, **insistir**, apresentou como tradução os verbos **insist**, **say** e **ask** dos quais apenas **insist** ainda necessitaria ser pesquisado. Na tabela a seguir encontra-se o resultado dessa pesquisa:

3i) Insist

originais para tradução - 1 (1)

tradução para originais - 4 (15)

Insistir

Insistir (75%)
Tornar a falar (25%)

Tanto **insistir** (tabela 2i) quanto **insist** também são verbos pouco utilizados nos textos literários pesquisados e a exemplo de **declare**, **end** e outros dispensa comentários muito extensos.

O verbo **agree** resultante da pesquisa por **confirmar** (tabela 2j) também apresentou poucas ocorrências como podemos ver na tabela abaixo:

3j) Agree

originais para tradução - 1 (14)

tradução para originais - 4 (12)

Concordar

Concordar (50%)
Afirmar (25%)
Confirmar (25%)

Crop, 10, 2004

O equivalente a **agree** em português é quase sempre **concordar**, mesmo nos textos em que **agree** é tradução. Nesses casos, **confirmar** apresentou apenas a metade das ocorrências de **concordar**.

Os resultados das pesquisas por **speak**, que apareceu na tabela 2k como equivalente de **falar**, e de **think**, extraído da tabela 2l, onde é resultado de **refletir**, encontram-se nas tabelas 3k e 3l respectivamente:

3k) Speak

originais para tradução - 0 (7)

tradução para originais - 7 (28)

Falar

3l) Think (to oneself)

originais para tradução - 1 (4)

tradução para originais - 2 (5)

Pensar (consigo mesmo)

Desafiar (de si para si)

Pensar

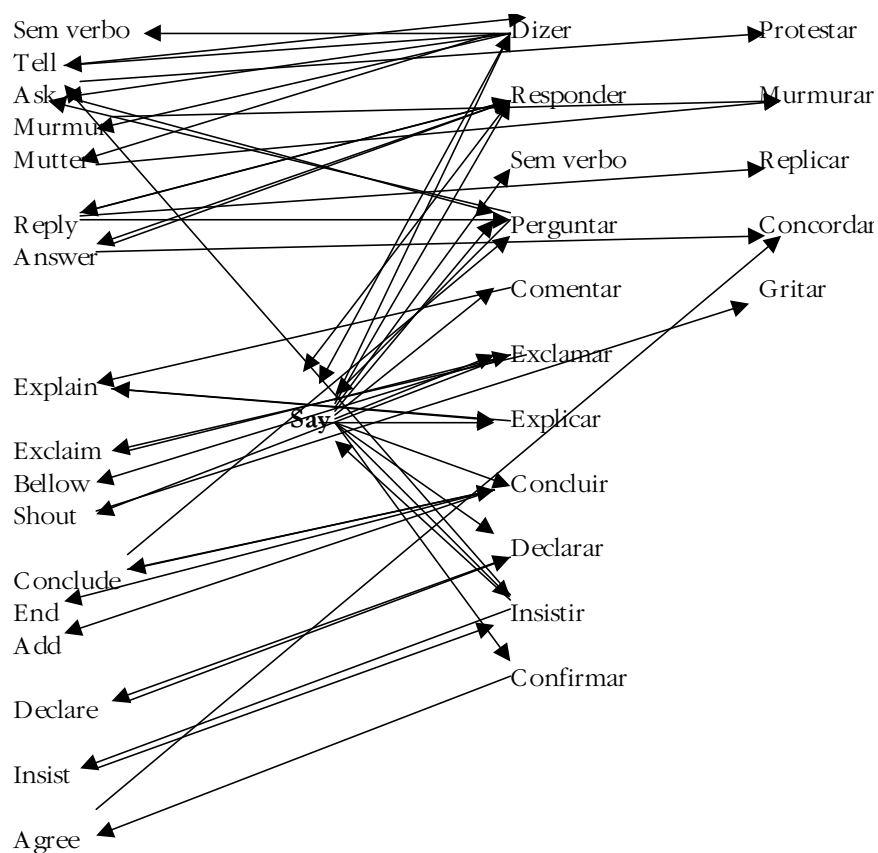
O verbo **falar**, como já vimos na etapa 2 do trabalho, é bastante usado em português, porém quase nunca em diálogos. Já o verbo **speak**, equivalente a **falar** é comparativamente menos usado em inglês e não ocorre em diálogos, já que apresenta um significado bastante próximo de **say**, que é o verbo mais usado em diálogos. Nos casos em que **speak** apareceu como tradução, seu original foi **falar**.

Os resultados do verbo **think** foram obtidos a partir da busca por “thought” “to”, pois importavam apenas os casos de **thought to myself**, **thought to himself** etc. A não ser quando a ação de pensar apresenta uma carga semântica diferente e mais forte do que a do próprio verbo que a representa, como foi o caso do texto original com **desafiar**, **think to oneself** é geralmente o equivalente a **pensar** seguido de um complemento como **consigo mesmo** ou **de si para si**.

Concluída essa terceira etapa do trabalho, é possível fazer um levantamento dos verbos em português que não haviam aparecido na busca por **say**, mas que foram surgindo a medida que um novo verbo em inglês era pesquisado. Uma quarta etapa da pesquisa poderia ser iniciada a partir da busca por cada um desses verbos, e, depois dela, talvez uma quinta e sexta etapas até que nenhuma ocorrência inédita fosse obtida, e, assim, a pesquisa pudesse ser considerada encerrada. Porém, acreditamos que as três etapas já realizadas foram suficientes para tirarmos conclusões importantes acerca da tradução dos verbos de elocução em diálogos.

LOFFREDO, Laís; GROSSMAN, Deborah; BITAR, Gladys; GONÇALVES, Janice. *Verbos de Elocução – As Diferenças entre o Inglês e o Português*.

Por fim, expomos no gráfico abaixo toda a pesquisa de textos originais para textos traduzidos de forma resumida procurando assim oferecer uma visão mais clara dos passos seguidos e dos resultados obtidos:



Posição dos verbos de elocução

Antes de chegarmos à conclusão final de nosso trabalho, que nada mais é que um apanhado geral das conclusões que conseguimos chegar ao longo da

Crop, 10, 2004

pesquisa, gostaríamos de fazer algumas considerações a respeito da posição dos verbos de elocução nos diálogos. Enquanto que no inglês o pronome pessoal e geralmente o substantivo precedem o verbo de elocução posposto à fala da personagem, como neste exemplo: “*I really ought to go back inside,*” **he said**. (EBDL3T1), no português ocorre o contrário: nos casos em que o verbo de elocução está posposto à fala da personagem, ele é que sempre inicia a frase seguido do pronome pessoal ou do substantivo, como no exemplo a seguir: “Esqueceu alguma coisa? **perguntou Marcela** no pé do patamar.” (PBMA2) Imaginávamos encontrar problemas em relação a posição dos verbos em traduções do inglês para o português, ou seja, casos em que o verbo de elocução aparecesse depois do pronome pessoal ou substantivo e por isso analisamos as traduções de textos com os verbos **say** e **ask**. Porém esse problema não se apresentou. Encontramos, no entanto, alguns exemplos de textos originais em português, que talvez por influência do inglês, seguiram seu padrão. Um exemplo foi: “Quer vender todas?”, **ele perguntou?**(PBRF1).

III. Conclusão

A idéia dessa pesquisa surgiu da percepção de que o inglês e o português apresentam certas diferenças estilísticas no que se refere à escolha lexical que poderiam passar despercebidas na tradução. Através dela, pudemos confirmar nossa hipótese de que os verbos de elocução não são usados da mesma forma nos dois idiomas. Enquanto que o inglês não se preocupa com a variedade, o português tende a utilizar diferentes verbos de elocução dependendo do valor significativo que pretende atribuir à fala da personagem. É bastante comum encontramos capítulos inteiros de obras originais em inglês em que o único verbo de elocução utilizado é **say**, algumas vezes acompanhado de um advérbio de modo que lhe acrescenta um sentido mais específico. Porém, em obras originais em português, **dizer** é apenas mais um dos muitos verbos de elocução que acompanham seus diálogos.

Seria bastante coerente, portanto, que em traduções do inglês para o português, o verbo **say** em expressões como **he said** se transformasse em um verbo diferente de **dizer**, dependendo da intenção comunicativa da personagem (e do autor) ou de seu tom de voz, no entanto, foi comprovado com esse trabalho que, talvez por se preocupar muito com a fidelidade da tradução, por falta de tempo para analisar cada contexto, ou simplesmente por pouca observação da parte dos tradutores, os verbos de elocução são geralmente traduzidos de forma literal. O mesmo ocorre em traduções do português para o inglês, em que verbos como **perguntar**, **responder** e **continuar** em expressões do tipo **pergun-**

LOFFREDO, Laís; GROSSMAN, Deborah; BITAR, Gladys; GONÇALVES, Janice. *Verbos de Elocução – As Diferenças entre o Inglês e o Português*.

tou ela acabam traduzidos por **ask**, **answer** e **continue** quando o mais natural seria **say**. Pretendemos com esse trabalho chamar a atenção de pessoas envolvidas em atividades relacionadas à tradução para a questão da preservação da naturalidade da língua de destino na tradução de verbos de elocução e, com isso, contribuir para que surjam trabalhos cada vez mais precisos.

Bibliografia

COLLINS, Glenn. Objects at hand, objects of veneration [on line]. 2002. Disponível: <http://www.nytimes.com/>.

GRAHAM, Janice. *Firebird*. New York: Putnam, 1998.

LABAKI, Amir. Salles defende documentário de autor [on line]. 2002. Disponível: <http://www.folha.uol.com.br/>

RIBEIRO, João Ubaldo. *O sorriso do lagarto*: 3 ed. Rio de Janeiro: Nova Fronteira, 1991.

COMPARA: <http://www.linguateca.pt/COMPARA.html>

The Case of *laugh*: a Parallel Corpus-based Research

Josimeire Cristina Martins*

Abstract: *Through a contrastive analysis of a parallel corpus, it is possible to verify the interlingual relations from an original to a translated text. This paper shows the results of a study using parallel corpora to identify how the word laugh is translated into Portuguese and to analyze these relations.*

Keywords: *parallel corpora; corpus linguistics; translation.*

Resumo: *Com a análise contrastiva de um corpus paralelo, podemos verificar as relações interlinguais de um texto original para um texto traduzido. O presente artigo mostra os resultados de um estudo utilizando corpora paralelos para identificar as traduções da palavra laugh para o português e analisar essas relações.*

Palavras-chave: *corpora paralelos; lingüística de corpus; tradução.*

Introduction

The aim of this article is to report the results of a study using parallel *corpora* constituted of original and translated English and Portuguese texts. Our purpose is to identify how the word *laugh* is translated into Portuguese and try to build the interlingual relations through a bi-directional research, from an original text to its translation equivalent.

* Mestranda do Programa de Estudos Lingüísticos e Literários em Inglês da FFLCH/USP.

MARTINS, Josimeire Cristina. *The Case of laugh: a Parallel Corpus-based Research*.

For this research we used COMPARA, an electronic parallel corpus, constituted of original texts in Portuguese and English and its translations into these same languages. This corpus is freely accessed through the Web and can be used to investigate the process of translation, helping not only researchers, but also translators, learners or anyone who works with these languages.

Besides COMPARA, we also made a research in the British National Corpus (BNC), a corpus constituted of original texts in British English. This kind of research is important since it lets us identify and analyze translations and their problems and difficulties, as well as the preference translators have for certain words.

We based our research on the word *laugh* and its translations into Portuguese. The method used is based on the contrastive analysis of Schmieid (2000) and Váradi (2001).

Methodology

According to Baker “translated language on the whole (within a particular historical and social context) may have certain features that distinguish it from original language” (1999: 291).

Our question is whether the variety of words occurs mainly in the original or the translated text. Baker says that

“the choice of a suitable equivalent will always depend not only on the linguistic system or systems being handled by the translator, but also on the way both the writer of the source text and the producer of the target text, i.e. the translator, choose to manipulate the linguistic systems in question.” (Baker, 1992:18)

We hypothesize that the author tends to use a greater variety of equivalents since he is dealing with his own language, while the translator would tend to avoid such variety, using the same lexical item more frequently.

The following keys were used in the process: OE – original English; TE – translated English; OP – original Portuguese and TP – translated Portuguese.

COMPARA lets us choose the variety of English and Portuguese we want to investigate in our search. We used Brazilian Portuguese and American and British English. We began our search with the word *laugh* used as a noun, in both singular and plural forms and its translations into Portuguese. Then, we made another search, with all the words found in the translation, this time

Crop, 10, 2004

using the original Portuguese corpus, looking for the results in English. Figure 1 shows the procedure.



Figure 1 – The structure of the search

Through this methodology we could verify the number of occurrences in original and translated English and Portuguese and analyze the words that were used most.

Besides these searches, we also looked the words up in monolingual and bilingual dictionaries, in a thesaurus as well as in the BNC corpus.

laugh [la:f; læf]
n 1 riso, risada. 2 escárnio. 3 piada, coisa engraçada. •vt+vi 1 rir, gargalhar. 2 escarnecer. to have the last laugh rir por último. to laugh at rir-se de.
Source: Michaelis online dictionary

Main Entry: ²laugh
Function: *nom*
1 : the act of laughing
2 a : a cause for derision or merriment : JOKE b : an expression of scorn or mockery : JEER
3 plural : DIVERSION, SPORT <play baseball just for laugh>
Source: Thesaurus online dictionary

risada
[De *riso* + *-ada*.]
S. f.
1. **Riso (1).**
2. **Riso franco e estrepitoso; gargalhada.**
3. **Riso conjunto de muitas pessoas**

gargalhada
[De *gargalhar* + *-ada*.]
S. f.
1. **Risada franca e mais ou menos ruidosa e prolongada; casquinada, gaitada.**

riso
[Do lat. *risu*.]
S. m.
1. **Ato ou efeito de rir; risada.**
2. **Alegria, contentamento, satisfação.**
3. **Coisa ridícula.** [Cf. *riço*, do v. *riçar*.]

Riso amarelo. 1. Riso forçado, contrafeito; sorriso amarelo.
Riso sardônico. 1. Med. Expressão facial indicativa de forma de arrebato espasmódico ou tetânico de dentes ou riso involuntário. 2. P. ext. Riso forçado, ou sarcástico, que lembra essa expressão.
Perdido de riso. 1. Que não pode conter o riso.

Source: Houaiss online dictionary

Figure 2: Some examples of the words *laugh*, *riso*, *risada* and *gargalhada* as nouns in mono and bilingual dictionaries

We started our search with the word *laugh*, from original English texts into translated Portuguese. We got 6 examples as a noun out of the 11 found in COMPARA.

EBDLIT2 (104):	She gave a brief, barking laugh , which unsettled me.	Ela deu uma risada curta, como um latido, que me deixou apreensivo.
EBDLIT2 (1107):	It was a resonant, full-bodied laugh , rather a surprise given her petite figure and general style of poised professional career woman; and when she laughed she threw back her head and shook it from side to side, making her hair fan out.	Era uma risada risonante, cheia de vontade, um pouco fora do comum, dada sua forma <i>petite</i> e seu estilo geral de mulher profissional de carreira; e quando ela ria, jogava a cabeça para trás balançava-a de um lado para o outro, fazendo seus cabelos se abrirem como um abano.
EBDLIT2 (1673):	That's always bad news: some idiot with a very loud, inane laugh , who goes on baying or cackling or shrieking at something long after everybody else has stopped, or starts up when nobody else is laughing, in the lull between two gags.	Isso é sempre má notícia: uma idiota qualquer (por alguma razão, é quase sempre uma mulher) com uma risada alta, num ritmo monótono, que começa com uma espécie de balido ou cacarejado ou grito, rindo de algo que todos já riram há muito tempo, ou começa a rir quando ninguém está rindo, naquele vão entre as linhas engraçadas.

Figure 3: Some examples of the search in COMPARA

From the 6 examples, 4 were translated as *risada*, 1 as *fazer mais engraçado* and 1 as *cair na gargalhada*. The next step was to search for the head words we got in original Portuguese texts and compare the words used in the translation into English. The noun *risada(s)* returned 9 examples:

PBAA2(233):	E soltou uma risada .	and she broke into giggles .
PBCB1(213):	Recrudescem as risadas .	The laughter grows louder.
PBMAA1 (186):	Cada frase que proferia era interrompida com uma risada prolongada e sonora, e com um certo caído de cabeça para trás, talvez gracioso se não tivesse muito de afetado.	Every sentence she uttered was punctuated with a long, sonorous giggle and a certain backward toss of her head that would perhaps have been characterful if it had not been so highly affected.
PBMAA1 (212):	-- Qual... respondeu Vidinha acompanhando este <i>qual</i> da sua costumada risada ; estou já tão cansada... que nem posso!	«Aww,» Vidinha replied, accompanying that «aww» with her customary giggle . «I'm so tired now... that I just can't.»

Figure 5: Search results for *risada(s)* in COMPARA

PBAA2 (1264):	As senhoras espantaram-se, mas abriram logo em gargalhadas .	The ladies, startled at first, immediately broke into giggles.
PBAA2 (1613):	E, após uma gargalhada seca, voltou-se para Raimundo e ofereceu-lhe com ar pretensioso «um talher na sua parca mesa».	And after a dry chuckle, he turned toward Raimundo and pretentiously offered him 'a place at my humble table.'
PBAD1(5):	Se fosse Tuim por exemplo, João tinha dado uma boa gargalhada .	If it'd been Tuim, for example, João would have laughed outright.
PBMA2(207):	A figura soltou uma gargalhada , que produziu em torno de nós o efeito de um tufão; as plantas torceram-se e um longo gemido quebrou a mudez das coisas externas.	The figure let out a guffaw, which produced the effect of a typhoon around us; plants twisted and a long moan broke the silence of external things.

Figure 6: Some examples of the search results for *gargalhada(s)* in the COMPARA

We also looked up the same words in the BNC corpus to verify the number of occurrences and then analyzed the results.

Results

The process mentioned in the Methodology section was applied to all the words found. Table 1 shows the results from this process and Table 2 shows the occurrences found in the BNC corpus:

Original Texts	Word	Occurrences	Translated texts	Words found and occurrences
OE	laugh(s)	06 (as nouns) (out of 11)	TP	risada (04) fazer mais engraçado (01) cair na gargalhada (1)
OP	risada(s)	09 (03 sing. e 06 plural)	TE	giggle(s) (3) giggle (v) (1) laugh (1) laughter (4)
OE	giggle(s)	-	TP	-
OE	laughter	06	TP	risada(s) (4) rir (v) (2) gargalhada (1)
OP	gargalhada(s)	14	TE	giggles (2) chuckle (1) laugh (v) (7) guffaw (1) laughter (1) laughing (n) (1) chuckles of laughter (1) laugh (n) (1)
OE	guffaw(s)	-	-	-

Table 1: Occurrences of the words laugh, laughter, giggle, guffaw and chuckle (original texts) and their translations.

Word	Results	Occurrences
laugh (n)	50	891
laughs (n)	50	148
giggle (n)	50	130
giggles (n)	50	119
guffaw (n)	26	26
guffaws (n)	20	20
laughter	50	2065
chuckle (n)	50	139
chuckles (n)	21	21

Table 2: Examples found in the BNC (British National Corpus)

Looking at the results in Table 1, we can notice that the noun *laugh* was translated as *risada* 66% of the time and as *gargalhada* 16%. The same did not occur when we got the results translated into English. The noun *laugh* occurred only in 11% of the examples, while *giggle* (as a noun) occurred 33% and *laughter* 44%, showing that the noun *laugh* was not the main choice in translations.

Analyzing all the returns we got from the original English texts, we noticed that there was a greater variety of equivalents chosen by the translators, not by authors, as the nouns *giggle*, *guffaw* and *chuckle* did not occur in original English. On the other hand, we did not have great differences from original and translated Portuguese texts on the variety of equivalents, although we found more occurrences of *gargalhada* (14 occurrences) in original Portuguese than in translated texts.

The BNC results showed a considerable difference from the number of occurrences between the noun *laugh* (1039) that represented only 11% of the occurrences in translated English texts and *giggle(s)* (249) that occurred 33%, reinforcing the numbers that show the preference translators have for equivalents.

Conclusion

Toury says that “in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of more habitual options offered by a target repertoire” (Toury *apud* Baker:

Crop, 10, 2004

289), that is, translations avoid innovations, something that would not happen to the author, who tends to use a greater variety of words, as a stylistics resource or because the language itself offers these varieties.

This was not what we found in our research, which showed a greater variety of words in the translated English texts, not in the original ones, although it is important to bear in mind the limitation of the corpus.

However, we hope this paper demonstrated some of the researchs that can be done using a parallel corpus in the investigation of the translation process and word choice.

References

- BAKER, Mona. *In Other Words. A coursebook on translation*. Routledge, London, 1992
- BAKER, Mona. The role of corpora in investigating the linguistic behaviour of professional translators. In: *International Journal of Corpus Linguistics*, 1999, vol. 4, pp. 281-298
- SCHMIED, Josef and FINK, Barbara. Corpus-based contrastive lexicology: the case of English with and its German translation equivalents. In: Botley, McEnery & Wilson. 2000, p. 157-176.
- SCHMIED, Josef. Translation corpora in contrastive research, translation and language learning. In: *TradTerm 10 (Revista do Centro Interdepartamental de Tradução e Terminologia FFLCH/USP)*. 2004, São Paulo: Humanitas, p. 83-115)
- VÁRADI T.; KISS, G. *Equivalence and Non-equivalence in Parallel Corpora*. In: *International Journal of Corpus Linguistics*. 2002, vol. 6, pp. 167-177)
- VARANTOLA, Krista. Disposable corpora as intelligent tools in translation. In: *Cadernos de Tradução*. 2002, n. 9, pp. 171-189)

Electronic references

On-line Dictionaries (pages accessed on August 5, 2004)

www.thesaurus.com

<http://www.m-w.com/home.htm>

<http://houaiss.uol.com.br/busca.jhtm?>

http://cf6.uol.com.br/michaelis/dicionar.cfm?dicion_id=8



Absolutamente e *Absolutely* – São ou Não São Absolutamente Intercambiáveis?

*Alvamar Helena de Campos Andrade Lamparelli**

Resumo: *Todo item lexical numa língua tem seu padrão de comportamento único e exclusivo, sendo a opção por um equivalente pelo tradutor extremamente complexa quando os itens são semanti-camente relacionados. É preciso conhecer os hábitos colocacionais dos itens nas duas línguas para obter uma equivalência que considere os padrões regulares de ocorrência. Essa regularidade é de suma importância, pois o uso mais corrente de um padrão indica uma produção mais natural. Os lingüistas de corpus se ocupam essencial-mente não só com quais palavras, estruturas e usos são possíveis numa língua mas também com o que é provável, a diferença entre o que os falantes podem dizer e o que na verdade dizem. Esses itens estão envolvidos em uma série de padrões fraseológicos que tendem a corresponder a uma função específica, distinguindo-os de outros quase-sinônimos. Esta descrição de padrões leva em conta as associações contextuais e sua função pragmática e se baseia na linguagem autêntica, sem mediação, de corpora monolíngües, numa abordagem direcionada pelo corpus.*

Palavras-chave: *hábito colocacional; co-ocorrência; fraseologia; quase-sinônimo; abordagem direcionada pelo corpus; evidência do corpus.*

* Mestranda do Programa de Estudos Lingüísticos e Literários em Inglês da FFLCH/USP.

Abstract: *Every Lexical item has its own pattern of behaviour, which means that choosing an equivalent is extremely complicated for the translator when the items are semantically related. Knowing their collocational habits is crucial in order to attain an equivalence which considers the regular patterns of occurrence. This regularity is highly relevant because the most frequent use of a pattern indicates a more natural production. Corpus linguistics are typically concerned not only with what words, structures or uses are possible in a language but also with what is probable, likely to occur, the difference between what the speakers can say and what they actually say. These items are involved in a series of phraseological patterns which tend to correspond to a specific function, distinguishing them from their near-synonyms. This description of patterns takes into account their contextual associations and pragmatic function and is based in an un-mediated, naturally-occurring authentic language in a corpus-driven approach.*

Keywords: *collocational habit; co-occurrence; phraseology; near-synonym; corpus-driven approach; corpus evidence.*

1. Introdução

O conceito de equivalência na tradução vem merecendo um interesse renovado por pesquisadores, pois as evidências fornecidas por corpora eletrônicos têm sugerido que esse conceito seja revisitado. O advento das pesquisas com corpora deu ênfase àquilo que pode ser descoberto a partir do estudo de um grande número de dados e que não é revelado a partir de outros meios como dicionários e gramáticas. Os dicionários bilíngües, que por razões de espaço pouco contexto fornecem, constituem ferramenta limitada para tradutores, podendo até mesmo fazê-los desviar do caminho. Em muitos casos a evidência do uso da língua obtida a partir de corpora é capaz de refinar, ou mesmo corrigir a informação fornecida por esses meios.

O presente trabalho se baseia numa abordagem “direcionada pelo corpus” (*corpus-driven*), pois privilegia a evidência fornecida pelo corpus, que constitui mais do que um repositório de exemplos que sustentam teorias pré-existentes.

Indeed, many of the statements are of a kind that are not usually accessible by any other means than the inspection of corpus evidence. [De fato, não se poderia chegar a muitas dessas constatações senão ao se inspecionar as evidências fornecidas pelo corpus] (Tognini-Bonelli 2001; 85)

Este estudo começa por discutir uma área problemática, a das palavras cognatas, em que o tradutor precisa conhecer as circunstâncias exatas em que um item lexical pode ser substituído por outro de forma semelhante. Como pode ser empregada a análise de corpus para determinar quando equivalentes potenciais de tradução são falsos cognatos e quando são mais confiáveis?

Na verdade, a opção por um item lexical é extremamente complexa. Como a seleção é feita entre itens semanticamente relacionados, a substituição de um pelo outro pode não mudar o sentido da frase de modo evidente, mas um deles pode ser mais apropriado do que o outro. Isto quer dizer que o tradutor precisa conhecer os hábitos colocacionais dos itens nas duas línguas em questão a fim de atingir um grau de equivalência que leve em conta os padrões regulares de ocorrência.

Cabe aqui mencionar que a análise lingüística realizada por meio de corpora computadorizados baseia-se na busca de padrões recorrentes. Se houver a constatação de que um padrão ocorre com freqüência na língua, essa regularidade é relevante para o lingüista ou tradutor, pois indica uma *probabilidade* maior de ocorrência daquele padrão. Em geral estes padrões não são imediatamente observáveis através da mera introspecção, mas se tornam aparentes por meio de concordâncias. Partington explica a importância dessa busca ao comentar:

In other words, if something is seen to happen frequently in a language, then it is significant. It is significant precisely because this frequent occurrence or, regularity, can be used as the basis for predicting how other, as yet unanalysed, chunks of language will behave, and in the end, for hypothesising a description of how the entire universe of discourse under study is constructed. Very often these language patterns are not immediately obvious in the course of simple introspection, but they can become more apparent through the medium of the concordance. [Em outras palavras, se algo ocorrer com freqüência numa língua, esse algo é significativo. Isso precisamente porque essa ocorrência freqüente ou regularidade pode ser utilizada na previsão do comportamento de outros segmentos ainda não analisados de linguagem, e em última análise, para supor uma descrição de como todo o universo de discurso em estudo é construído. Geralmente esses padrões de linguagem não são imediatamente óbvios através da simples introspecção, mas podem se tornar mais evidentes por meio da concordância.] (Partington, 1998: 9)

A observação dessa regularidade é de suma importância para o tradutor, pois o uso mais corrente de um padrão indica uma produção mais natural. Tagnin, ao descrever o **tradutor ingênuo**, levanta a questão desse “desapercebimento”.

LAMPARELLI, Alvamar Helena de Campos Andrade. *Absolutamente e Absolutely – São ou Não São Absolutamente Intercambiáveis?*

Por estranho que pareça, mesmo como falante nativo da língua alvo, o tradutor pode ter problemas no nível da produção para conseguir soluções naturais, caso se atenha tanto ao texto de partida a ponto de não perceber que, entre formas igualmente gramaticais, uma delas é de uso mais corrente. Em outras palavras, pode não se dar conta de que, dentro de uma gama de formas gramaticalmente possíveis, há certas formas que têm uma probabilidade maior de ocorrerem. Caso o tradutor selecione uma dessas formas *possíveis*, em detrimento da mais *provável*, produzirá uma tradução não natural, não fluente. Esse problema certamente se agravará quando o tradutor não estiver traduzindo para sua língua materna (Tagnin, 2002/1: 193).

Como a mera introspecção não lhe será de suficiente ajuda, uma análise das concordâncias o munirá com evidências dessas recorrências. Uma concordância KWIC (KeyWord In Context) constitui uma lista de linhas de textos não relacionadas entre si que foram extraídas através de concordanciadores, de um corpus eletrônico, isto é, uma coletânea de textos mantidos em uma forma que é acessível ao computador. No centro de cada linha está o item em estudo (palavra de busca, ou nóculo). O resto de cada linha contém o co-texto imediato à esquerda e à direita da palavra de busca. Essa lista possibilita ao analista procurar por eventuais padrões nos co-textos, o que oferece pistas quanto ao uso do item procurado. A busca por essas pistas constitui a base da metodologia descrita no estudo de caso apresentado.

Os itens estudados estão envolvidos em uma série de padrões fraseológicos, que tendem a corresponder a uma função específica, distinguindo-os de outros quase-sinônimos. Partington levanta essa dificuldade de distinção:

Every lexical item in the language has its own and unique pattern of behaviour. It is no surprise that translators and language learners experience great difficulties in the search for equivalents in texts. [Todo item lexical numa língua tem seu padrão de comportamento único e exclusivo. Não é surpresa que tradutores e aprendizes de língua experimentem grandes dificuldades na busca de equivalentes nos textos.] (Partington, 1998; 46)

2. Metodologia

Este trabalho teve como ponto de partida os estudos de Alan Partington e Elena Tognini-Bonelli que revisitam o conceito de equivalência, com base no conceito de Halliday:

Crop, 10, 2004

If meaning is function in context, as Firth used to put it, then equivalence of meaning is equivalence of function in context. [Se o sentido é a função no contexto, como Firth costumava considerar, então a equivalência de sentido é a equivalência de função no contexto.] (Halliday 1992:16; in Tognini-Bonelli, 2001:131)

Bonelli busca definir uma unidade de análise: a “unidade funcionalmente completa” que na leitura de Sardinha é “*uma unidade de sentido de caráter fraseológico, a respeito da qual há evidências no Corpus*” (Berber-Sardinha, 2001/2002:35).

Alan Partington parte para uma investigação detalhada do comportamento de pares de “*look-alike items*” em inglês e italiano (*correct/corretto; absolutely/assolutamente; completely/completamente; entirely/interamente*), num estudo de equivalência funcional no nível fraseológico. Em termos de Halliday, um estudo de como o tradutor pode lançar mão do conhecimento do contexto da frase, proporcionado pelas concordâncias, para refinar a escolha de equivalentes potenciais de tradução.

O que se procura é descrever os padrões regulares de um item em questão, segundo os passos propostos por Sinclair: “*identifying and defining the meaning of words which takes into account on the one hand their contextual associations and on the other their pragmatic function.*” [“Identificação e definição do sentido das palavras que leva em conta por um lado suas associações contextuais e por outro sua função pragmática.”] (Bonelli, 2001:19; adaptado de Sinclair 1996)

É importante ressaltar que o ponto de partida para se identificar a configuração e as realizações de duas unidades de sentido comparáveis tem de ser a linguagem autêntica, sem mediação (“*naturally-occurring*”, “*un-mediated language*”), daí as observações serem baseadas em dois corpora monolíngües.

A informação será reunida a partir de corpora monolíngües em inglês, uma parte dos dados já identificados por Partington, utilizando um corpus de textos jornalísticos extraídos do jornal *The Independent*, de 4 milhões de palavras; complementados por dados de um corpus jornalístico do CETRAD (Curso de Especialização em Tradução) constituído pelos jornais *Washington Post*, *Financial Times*, *Los Angeles Times*, *The Guardian*, *The New York Times*, *The Times*, *U.S.A.Today*, de aproximadamente 2 milhões de palavras, e do BNC on-line (*British National Corpus*), de 100 milhões de palavras, fornecendo sempre 50 resultados (<http://sara.natcorp.ox.ac.uk>) e em português (um corpus de textos jornalísticos da Revista Veja de aproximadamente 10 milhões de palavras e a *Web* por meio do formatador de resultados do buscador *google*. “*Kwic Google*” oferecido pelo CEPRL (<http://lael.pucsp.br/corpora>), escolhendo no menu das preferências a opção para se obter mais resultados(100) na tela do *Google*. Um corpus

paralelo, COMPARA, composto de originais e traduções em inglês e português, será utilizado para buscar outros candidatos a equivalentes, ampliando o número de possibilidades já identificadas e utilizadas por tradutores e confirmadas pelo uso tradutório real.

A ferramenta *WordSmith Tools* de Mike Scott foi utilizada para analisar o corpus da Veja e do CETRAD, para fazer as concordâncias e analisar os colocados.

Cabe aqui mencionar que o acesso a um corpus de língua geral em português constitui ainda uma dificuldade para o tradutor. Temos o Lacio-Web, de 10 milhões de palavras, que permite a compilação de vários sub-corpora, uma amostra do Banco de Português elaborado pela PUC com cerca de 1,1 milhão de palavras, e a ferramenta *Kwic-google* de acesso à *Web*, que é oferecida pelo Cepril, aliás extremamente útil, compensando esse acesso limitado a um maior número de dados.

3. A pesquisa

Como já mencionado, a presente pesquisa teve como ponto de partida um estudo realizado por Alan Partington, ao trabalhar com palavras de forma semelhante e sentido similar no inglês e italiano, ou seja, dois itens que parecem ter a mesma significação, ser “*true friends*”, e que na verdade são utilizados de modos sutilmente diferentes e em contextos diversos em cada uma das línguas. Essa percepção é de extrema valia para o tradutor, pois às vezes sua expectativa de que estes constituam “excelentes amigos” não é confirmada. A diferença no comportamento colocacional de duas palavras comparadas nas duas línguas pode ser significativa, pois cada uma interage de um modo com seus respectivos contextos lingüísticos, às vezes preferindo um campo léxico-semântico em uma língua e não na outra.

A título de ilustração, o autor mostra que as palavras “*correct*” e “*corretto*” (em italiano) co-ocorrem com uma série de itens de campos léxico-semânticos variados, a primeira com bastante frequência com o de “*response*”; “*answer*”. Quando se procura no corpus italiano os colocados para a palavra “*risposta*”, constata-se que nunca ocorre com “*corretta*”, e sim com “*adeguata*”, “*concreta*”, “*precisa*”. “*Correct*” co-ocorre com uma série de itens do campo léxico-semântico de “*time*” (*time, date, moment*). A frase “*correct time*” é funcionalmente equivalente a “*l’ora esatta*”. Isso não quer dizer que os falantes nativos classificariam “*risposta corretta*” ou “*ora corretta*” como inaceitáveis, porém a evidência do corpus sugere outras colocações que são preferidas. Não se trata do que é **possível**, mas do que é **provável**, a diferença entre o que os falantes **podem** dizer e o que na verdade **dizem**.

Corpus linguists are concerned typically not only with what words, structures or uses are possible in a language but also with what is probable – what is likely to occur in language use. [Os lingüistas de corpus se ocupam essencialmente não só com quais palavras, estruturas ou usos são possíveis numa língua mas também com aquilo que é provável – o que apresenta probabilidade de ocorrer no uso da língua.](Kennedy, 1998:9)

Em outros casos, não se trata apenas de preferência, mas a correspondência entre os sentidos e usos de um item nas duas línguas não constitui uma justaposição perfeita (“*perfect overlap*”). É o caso de “*absolutely*” e “absolutamente”, cuja pesquisa em corpora pouco a pouco revela aspectos inusitados. Esse caminho “direcionado pelo corpus” é descrito por Partington:

A researcher has an intuition about language, checks this against the data the corpus provides, and this checking process frequently suggests other avenues of research to be taken, often entirely unsuspected at the start of the process (the so-called: “serendipity” principle, Higgins 1991). Intuition and data collection work hand in hand. [Um pesquisador tem uma intuição sobre a língua, a investiga nos dados fornecidos pelo corpus e esse processo de investigação geralmente sugere outros caminhos de pesquisa a serem tomados, com frequência completamente impensados no início do processo (o chamado: “princípio da descoberta por acaso”). A intuição e a coleta de dados trabalham de mãos dadas.] (Partington, 1998:1)

Ao investigar os resultados obtidos pelas concordâncias de “absolutamente” fornecidas pelo *Kwic Google*, um primeiro exemplo em que “*absolutely*” não constituiria um equivalente de tradução para “absolutamente” é revelado:

“Os partidos políticos ditos de direita têm uma coisa em comum com os partidos políticos ditos de esquerda: São partidos políticos. **Não estão absolutamente interessados** nas suas ou nas minhas idéias, mas apenas no seu e no meu voto...”

Uma versão que contivesse “*absolutely*” como: “*They are not absolutely interested in your opinions or in mine...*” não passaria o sentido pretendido que é “de jeito algum”, “de modo nenhum”. Para que isso ocorresse, teríamos de nos valer de outros itens em inglês que se associassem com essa construção negativa. Algo como *They are not at all interested*, *They are completely unconcerned* seriam mais “fiéis” ao que se quis dizer. É o que a análise de corpora nos permitirá dizer.

A fim de verificar até que ponto “absolutamente” e “*absolutely*” são correspondentes foram feitas concordâncias nas duas línguas, a fim de identificar seu ambiente colocacional, coligacional, de preferência e de prosódia semânticas.

Uma **primeira função** de “*absolutely*” observada é **intensificação de adjetivos hiperbólicos** (aqueles que expressam uma atitude ou opinião enfáticas, com um sentido superlativo inerente). Essa função de “*absolutely*” já foi observada por vários pesquisadores (Altenberg, 1991:137, Partington 1998:58). Partington constata a co-ocorrência regular no seu corpus de estudo com adjetivos como *appaling, gutted, delighted, outstanding, hilarious, disgraceful* e *fantastic*. Acrescenta também que há um equilíbrio razoável entre itens favoráveis e desfavoráveis, ou seja, não há uma prosódia semântica preponderante.

Cabe aqui relembrar o termo “prosódia semântica”, cunhado por Sinclair (1991) e desenvolvido por Louw (1993), Partington(1998) e, mais recentemente, por Hoey(2000), Hunston(2000) e Stubbs (2001). Segundo Hunston (2000:104):

Briefly, a word may be said to have a particular semantic prosody if it can be shown to co-occur typically with other words that belong to a particular semantic set. [Em suma, pode-se dizer que uma palavra tem uma determinada prosódia semântica quando se demonstra sua ocorrência típica com outras palavras que pertencem a um determinado campo semântico.]

Berber Sardinha enfoca esse fenômeno em alguns trabalhos, e o define como:

“Associação recorrente entre itens lexicais e um campo semântico, indicando uma certa conotação (positiva, negativa ou neutra) ou instância avaliativa” (Berber Sardinha, 2004: 236). E segue dizendo:

“A prosódia semântica é importante para o entendimento da tradução porque, embora carregue significado importante, não é indicada nos dicionários, ou manuais de tradução, de modo sistemático quando são apontados os vocabulários equivalentes. Assim um tradutor pode utilizar a prosódia semântica inadequada sem saber, ao empregar termos que são tidos como equivalentes”. (Berber Sardinha, 2004: 236)

É o caso de *impressionante*, que no italiano possui uma prosódia semântica negativa, colocando-se com itens relacionados a aumento de preços e atentados. Já seu equivalente no inglês (*impressive*), possui uma prosódia semântica positiva, colocando-se com itens como *achievement, talent e dignity*. Louw (Louw

1993 *apud* Stubbs, 1998:51) também analisa as implicações negativas do advérbio “utterly”, como em “utterly confused” e “utterly ridiculous”. Ao cometer uma contravenção involuntária desse padrão, por desconhecer que há uma divergência na preferência e/ou prosódia semântica, o tradutor estará “tingindo” o sentido com uma conotação não pretendida, pois há uma quebra não intencional da expectativa daquilo que vem a seguir. Já a intencionalidade nessa quebra, ou seja, a contravenção deliberada para atingir um efeito irônico, pode constituir um recurso, e não um “desconhecimento” do tradutor, mas isso está fora do escopo dessa pesquisa.

Temos então que “*absolutely*” se coloca regularmente com adjetivos hiperbólicos, tanto favoráveis como desfavoráveis, como observou Partington, e aqui esse perfil é corroborado por uma investigação no corpus jornalístico do CETRAD, em que as concordâncias obtidas pelo *WordSmith Tools* revelam: *monstrous; massive, devastated, crackers, enormous, amazed, terrified, pathetic e stupid*. O BNC (British National Corpus) on-line apresenta também: *horrific, outstanding, phenomenal, fascinating, wonderful, devastated, abysmal, delighted e minimal*.

Já seu equivalente “absolutamente”, embora ocorra com adjetivos hiperbólicos: fantástico(2); ridícula(2); revoltante; chocante (resultados de 59 ocorrências *Kwic-Google*) e admirável, assustador, aviltante, fantástico, fascinados(2), horrível, incrível; magnífico; lamentável; pavorosas (resultados de 313 concordâncias extraídas do corpus Veja), também se coloca regularmente com adjetivos não hiperbólicos: acessível, convergente, moderno, estável, presente, (*Kwic-Google* acima); normal(14), legais(5), comprovada(2), original(2), compreensível, comum, casual, convencional, flexível, saudável, simples (Veja, acima).

Com base nessa evidência, poderíamos afirmar que “absolutamente” não tem a função primordial de intensificar itens hiperbólicos como “*absolutely*”. A grande co-ocorrência de “absolutamente” com o campo semântico da “normalidade”; “legalidade” no Corpus da Veja, instiga uma busca adicional no *Kwic-Google*, com “absolutamente normal”, revelando 108 ocorrências e “absolutamente legal”, 98 ocorrências. Isso nos leva a outra pergunta: será que usamos com frequência em inglês “*absolutely legal*”, “*absolutely normal*”? Uma busca no BNC on-line revela 8 instâncias com *normal* e somente 1 com *legal*, dos 50 resultados mostrados. Será que a normalidade em inglês é “*taken for granted*”, não precisando ser intensificada?

Uma **segunda função** de *absolutely* é se **colocar com um outro grupo de itens lexicais**, apontado por Partington, **pertencente ao campo semântico de “*highly important*”**, incluindo *essential*(5), *crucial*(3), *vital*(2) e *necessary* (19). Novamente esse perfil é reiterado pela busca no Corpus jornalístico do

CETRAD, que revela *necessary*(2), *crucial*(2), *imperative*, *essential*. No BNC online, há 139 ocorrências de “*absolutely necessary*”, 122 de “*absolutely essential*”, 30 de “*absolutely crucial*”, 53 de “*absolutely vital*” e 7 de “*absolutely imperative*”.

Esse mesmo campo semântico também é muito bem representado nos colocados de “absolutamente”. Os resultados do *Kwic-Google* para “absolutamente essencial” apresentam 102 ocorrências, e “absolutamente necessário” 110 ocorrências. O Corpus da Veja revela também “necessário” (9), “indispensável”(3), essencial(2), imprescindível, fundamental e vital.

Uma análise das concordâncias de “absolutamente necessário” e “absolutamente essencial” nos chama a atenção para os casos em que aparece uma estrutura de negação, o que aliás nos remete ao ponto inicial de uma possível ambigüidade. Tomemos alguns exemplos extraídos do *Kwic-Google*:

“Encorajamos você a minimizar o tamanho dos exemplos, mas isso **não é absolutamente necessário**. Se o erro for reproduzível, nós o encontraremos de qualquer forma.”; “O atributo “*Namespace*” **não é absolutamente necessário** para o *WebService* funcionar, mas faz parte da norma dar um nome único ao *WebService*”; “Embora **não seja absolutamente necessário**, convém que o texto esteja sempre...”; “É desejável, mas **não absolutamente necessário** espelhar todos os arquivos...” “Se desejar, faça um segundo círculo para acrescentar decoração extra ao *soundhole*, mas não **é absolutamente necessário**...”

“Absolutamente”, em português, além de “totalmente”, “inteiramente”, “de modo absoluto”, também tem a acepção “de modo nenhum”, “de jeito algum”, já mencionada pelo dicionário Houaiss como regionalismo no Brasil, e pelo Aurélio como sinônimo de “em absoluto”. O que também causa confusão em sua tradução para o inglês, onde “*absolutely*” traz a idéia de “*positively*”, “*surely*”, “*definitely*”. Essa é uma armadilha a ser evitada pelo tradutor, pois a resposta “Absolutamente” no português do Brasil à pergunta: “Foi isso que você disse?” quer dizer, “de modo nenhum”, “de jeito algum” e no português de Portugal, à mesma pergunta, quer dizer, “É claro que sim!” (Mário Prata, *Schifaizfavoire*), equivalente à resposta “*absolutely*” em inglês, à mesma pergunta “*Is that what you meant?*” ou seja, “sim”, “sem dúvida”, “claro”. “*Absolutely not*” corresponderia ao nosso “Absolutamente”.

Percebemos claramente nos exemplos “**não é absolutamente necessário**” que o sentido pretendido está longe de “não é em absoluto necessário”, “é totalmente desnecessário”, “não é necessário de jeito algum”. Observando várias linhas expandidas de concordância, constatamos que há uma idéia inerente de: “não é absolutamente necessário, mas convém, é útil, é desejável”.

Quando essa construção negativa aparece em inglês também tem essa idéia explícita pela conjunção: “*Not absolutely necessary, you can do without it but it is desirable, advisable, helpful*”:

*“This item is **not absolutely necessary but** it is very helpful for moving the cake around while you are involved in delicate icing work./ This is not an important distinction because it reminds us that a high level of teamwork, **whilst** perfectly desirable, is **not absolutely essential** for many routine activities./ Wellington boots are advisable, **but** they were **not absolutely necessary** even in this exposed situation/ **While not absolutely essential**, a helicopter radio will make things much easier (BNC)*

Uma incursão ao corpus também revela outras associações regulares: “Somente/Apenas quando absolutamente necessário/a menos que seja absolutamente necessário” e “Restringir-se/reduzir ao absolutamente necessário” com o sentido de “estritamente necessário” com um padrão equivalente em inglês:

*“I would really recommend every body to think twice before having any form of plastic surgery, particularly **unless it’s, if it’s not absolutely necessary**./” Not to have a nail or any ornament **that is not absolutely necessary**./ Put straight into the waste-paper basket everything **which is not absolutely necessary** (BNC).*

Uma terceira função apontada por Partington é sua **co-ocorrência com um sub-grupo importante de itens que expressam “certainty/clarity”**, incluindo *certain*(2), *sure*(1) e *clear*(9). Uma investigação em português mostra que esse campo também é amplamente representado. Os resultados *Kwic-Google* apontam 97 ocorrências de “certo” em duas acepções: no sentido de “certeza” e no sentido de “correto”. Em inglês, essa última acepção (“*right*”/“*correct*”) se coloca com “*response/answer*”. O corpus da Veja revela que “absolutamente” co-ocorre com “certa” (1), “certas” (1), “certo”(4), “certos”(2) na acepção de “correto” e somente 1 vez na de “certeza” em: “É absolutamente certo que a agenda internacional já é outra”. “Absolutamente correta” (feminino e singular) aparece também com uma frequência alta de 108 no *Kwic-Google*.

Quanto à “claro” também está bem representado no corpus em português, com 99 ocorrências, no *Kwic-Google*. Há algumas frases contendo uma construção negativa, “não é/está absolutamente claro”:

- (1) “Não é absolutamente claro que isso seja uma boa idéia”
- (2) “Não é absolutamente claro de que modelo se tratava”
- (3) “Mas não está absolutamente claro para onde quer ir”

O interessante, é que as frases contendo “*clear*” em inglês mais “*absolutely*” pouquíssimo aparecem numa estrutura negativa (3 ocorrências no BNC). O autor observa que o advérbio “*entirely*”, é encontrado com grande frequência na fraseologia “*not/never + intensifier + adjective* (He was already drunk, and *not*

entirely pleased to be visited by journalists) (66 vezes com essas partículas negativas em 250 ocorrências de “*entirely*”). No corpus do CETRAD esse comportamento é corroborado pelos exemplos:

(4) *Not entirely ready to face life on the outside.*

(5) *The evening I arrived was not entirely peaceful.*

(6) *It is not entirely clear that Buchanan could get the Reform Party's nomination.*

Partington vai mais longe ainda e constata que “*entirely*” raramente intensifica qualidades “boas”, mas geralmente as nega. É provável que uma tradução mais adequada das frases (1)(2)(3) acima contivesse “*not entirely*”: “*It is not entirely clear*”.

Um quarto ponto é que, em geral, há mais exemplos de “**absolutamente**” sendo usado em expressões negativas que de “**absolutely**” (no estudo de Partington, somente 5 das 155 ocorrências de “*absolutely*” colocavam-se com adjetivos contendo um prefixo negativo, *im, in, ir* ou *un*. Em contrapartida, quando o autor examinou “*completely*”, uma proporção bem maior desses itens foi revelada, 31 de 200). O Corpus do CETRAD revela ainda quanto às 86 ocorrências de “*completely*”: *unstuck, unsleazy, unrelated, unnecessary, unjustified, unafraid* e *unacceptable*. No Corpus da Veja de 313 ocorrências de “absolutamente” temos uma série de adjetivos que co-ocorrem, contendo um prefixo negativo: inútil(3), inverídicos(2), improcedentes, atípico, anormal, incompatíveis, inabitável, incomum, incompreensíveis, inofensivos, insignificantes, inexistente, inepto, ineficaz, incompetente, desacreditado, desconfortável(2) e desnecessária.

Com base na evidência de que “*completely*” freqüentemente co-ocorre com adjetivos que contêm prefixo negativo, seria apropriado considerar que a versão para o inglês de uma frase que contivesse “absolutamente” junto aos adjetivos enumerados acima utilizasse “*completely*”, pois, nesse caso, seu ambiente colocacional é semelhante nesse aspecto: *completely untrue/incomprehensible/uncomfortable*”.

É justamente nesse ponto, o das construções negativas, que retomamos o exemplo mencionado no início deste estudo, qual seja: “Não estão absolutamente interessados na minha ou na sua opinião”. Percebemos que “*absolutely*” não é o candidato mais apropriado quando se trata de uma fraseologia negativa. Constatou-se que o sinônimo “*completely*” se associa com adjetivos modificados por partículas negativas. Talvez uma versão: “*They are completely unconcerned/uninterested/indifferent*” fosse mais adequada.

* Somente como modificador de adjetivo, pois outra função de “*absolutely*” é como intensificador de “*nothing*”; “*no*” e “*not*”, palavras negativas.

Em outras palavras, “*absolutely*”, como modificador de adjetivo⁶, não tem uma prosódia semântica negativa, ou seja, a expectativa do que vem a seguir não é de uma qualificação negativa. Por outro lado, “*completely*” parece se associar mais a esse campo semântico negativo, sugerindo o seu uso ao invés de “*absolutely*”, para dar conta da fraseologia negativa do exemplo citado. Poderíamos então sugerir que essa fraseologia negativa em que “absolutamente” ocorre traz uma prosódia semântica negativa, que encontrará seu equivalente funcional, em inglês, em “*completely*”.

Nessa altura, o tradutor poderá recorrer à pesquisa no COMPARA, corpus paralelo contendo originais e traduções, a fim de verificar como essa fraseologia “não” + “é” + “absolutamente” foi solucionada. Qual a surpresa ao deparar com:

Original:

(a) “*Catherine is not unmarriageable, but she is absolutely unattractive*”(EUHJ1)
James, Henry

tradução:

(b) A Catherine não é “incasável”, mas **não é absolutamente nada** atraente.

A princípio parece tratar-se de uma “*mistranslation*” (erro de tradução), porém, ao verificar que a tradução é feita para o português de Portugal, faz uma pesquisa no Kwic Google com a estrutura “não estão absolutamente” e encontra outros exemplos de “não estão absolutamente nada”, todos de sites portugueses: “Nós não temos dúvidas acerca de que estão homens sentados nesta Assembléia que **não estão absolutamente nada** interessados na construção do socialismo;... constituído por alunos que **não estão absolutamente nada** preocupados com...; A existência de notória semelhança **não é absolutamente nada** grave; ...**não estão absolutamente nada** fixas...”

Seria pertinente aqui mencionar outra observação feita por Partington, de que “*completely*” e “*entirely*”, em total contraste com “*absolutely*”, se colocam freqüentemente com palavras de dois campos semânticos: o de “**absence**” expresso por palavras como: *devoid, drained, painless, symptomless, valueless, empty, stripped, free of*; e o de “**change of state**”: *different, changed, new, restored, rebuilt, revamped*. Uma análise das concordâncias de “*completely*” e “*entirely*” extraídas do corpus do CETRAD corrobora essa tendência no que se refere a “*absence*”: *powerless, empty, absent, free from*; e no que se refere a “*change*”: *new (10), different(5) revised, revamped, redone, replaced*. Uma análise de “absolutamente” com esses dois campos semânticos, representados pelas palavras “sem” e “diferente”, no *Kwic-Google*, revela uma freqüência de 97 e 102 ocorrências respectivamente. Aqui uma boa tradução para o português brasileiro do exemplo citado

do COMPARA onde consta “*absolutely unattractive*”, seria “absolutamente sem graça/sem sal”. Os itens “completamente” e “totalmente” também mostraram uma significativa co-ocorrência com esse campo semântico de ausência, expresso por “sem”, 105 e 94 ocorrências respectivamente. Esse campo semântico em inglês, representado por “*free of*”, foi investigado no BNC para a expressão “*absolutely free of*”, que apontou somente 7 ocorrências, 4 delas na colocação “*free of charge*”.

Por fim, outra característica particular verificada por Partington do intensificador “*absolutely*” é o fato de, diferentemente de “*completely*” e “*entirely*”, intensificar as palavras *nothing*, *no* e *not*, já identificada por Altenberg (Altenberg, 1991:137 apud Partington, 1998:58). Essas palavras corresponderiam em português a “nada”, “nenhum”, “nenhuma”. O *Kwic-Google* revela a mesma tendência, com 145 ocorrências para **nada**, 114 para **nenhum**, 109 para **nenhuma**. O Corpus da Veja revela 33 ocorrências de **nada**, 8 de **nenhum(a)**.

Uma última investigação no COMPARA amplia o quadro de possibilidades nas duas direções:

Original Inglês: *Better say nothing at all* (EBLC1)

Tradução Português: O melhor é não dizer absolutamente nada.

Original Inglês: *It is a most provoking thing*. (EBLC1)

Tradução Português: É absolutamente irritante. (A associação de “absolutamente” à idéia de superlativo inerente é refletida na tradução)

Original Inglês: *He felt perfectly happy* (EBOW1)/ *He is a perfect stranger* (EUHJ1)

Tradução Português: Sentia-se absolutamente feliz/ Ele é absolutamente um estranho.

Original Português: Nós não sabemos absolutamente nada dos textos que roemos. (PBMA3)

Tradução Inglês: *We know absolutely nothing about the texts we chew*.

Original Português: Não quero absolutamente nada da casa de meus pais. (PPCC1)

Tradução Inglês: *I don't want anything at all from my parents' house*.

Percebemos que as observações de que “*absolutely*” coloca-se com “*nothing*” e pouco ocorre em contexto negativo é de extrema valia ao tradutor; sem essa informação, produziria algo que não é apropriado e que causaria um certo ruído por não ser natural, sendo fraseologicamente pouco convencional: “*We don't know absolutely anything*”. Uma busca no BNC revela 19 ocorrências de *absolutely anything*, porém somente num contexto positivo: “*This drawback is offset by the chance to create absolutely anything words can conjure up.*”

4. Tabela: "Resumo"

Inglês- Funções: <i>absolutely</i>	COLOCADOS	Português- Funções absolutamente hiperbólicos	COLOCADOS
<p>A. adjetivos hiperbólicos (que expressam opinião ou atitude incisivas, sentido inerente de superlativo). E equilíbrio razoável entre itens favoráveis e desfavoráveis.</p>	<p>COLOCADOS</p> <p><i>appalling</i>; <i>grateak</i>; <i>delightak</i>; <i>oustanding</i> - They are absolutely <i>amazed</i> at the response they have had. (CETRAD) - That's absolutely <i>horrific</i> that is! (BNC)</p>	<p>-adjetivos hiperbólicos</p> <p>-adjetivos não hiperbólicos</p>	<p>COLOCADOS</p> <p>-admirável: ("É absolutamente admirável a maneira como vem reorganizando as finanças...") (Veja) -fantástico: ("Mas o que já foi feito, do ponto de vista científico, é absolutamente fantástico".) (Veja) -horível: ("Sua mulher, Cherie, é horível, absolutamente horrível...") (Veja) -assustador: ("O número é absolutamente assustador para um país...") (Veja) -comum: ("William Thacker, um londrino absolutamente comum".) (Veja) -normal: ("A gestação é um estado absolutamente normal, porém o mecanismo feminino...") (Google); -legal: ("Comprar via internet é algo absolutamente legal".) (Google) -simples: ("...chapéus esquecíveis - tudo absolutamente simples, seco, sem frutos.) -flexível: ("Há ainda algumas discussões, mas a Anfavea está absolutamente flexível") -essencial: ("A escovação regular é absolutamente essencial para os cães jovens...") (Google) -indispensável: ("E, estes agentes se tornam absolutamente indispensáveis para qualquer comprador do Ocidente".) (Veja) -fundamental: ("As exportações foram absolutamente fundamentais para a recuperação da economia mexicana".) (Veja) -imprescindível: ("Clitos positivos, sombra cintilante, e absolutamente imprescindível, salto plataforma de no mínimo 12...") (Veja) -vital: ("Se o sono não desempenha uma função absolutamente vital, então ele é a maior falha do processo evolutivo...") (Veja) -necessário: ("Não remova a vítima salvo se for absolutamente necessário;" "Em educação é absolutamente necessário contar com um tempo...") (Google)</p>
<p>B. campo léxico-semântico de "highly important"</p>	<p>COLOCADOS</p> <p><i>essential</i>: "It is absolutely <i>essential</i> for Scotland that the new terminal..."; (CETRAD) <i>crucial</i>: "Getting the same lunch period is absolutely <i>crucial</i>." (CETRAD) <i>vital</i>: "Does the Minister accept that it is absolutely vital that disposal of such waste should... (BNC) <i>necessary</i>: "A corresponding religion is absolutely necessary to mankind" (BNC)</p>	<p>B. campo léxico-semântico de "highly important"</p>	<p>B. campo léxico-semântico de "highly important"</p>

<p>C. campo léxico-semântico de "certainty/clarity"</p>	<p>-certain: "You have to be absolutely certain they're going to achieve their objectives." (BNC) -sure "After making absolutely sure it was the right one" (CEFRAD) -clear: "The sexual preferences of the cast are made absolutely clear." (CEFRAD)</p>	<p>C. campo léxico-semântico de "certainty/clarity"</p>	<p>-certo (certeza): ("O segundo turno é absolutamente certo"; "Esteja absolutamente certo de que está usando o servidor certo".) (Google) -certo (correto): "Era o início da época do bordão. Absolutamente certo!"; "O absolutamente certo do J. Silvestre estava absolutamente correto!" (Google) -claro: ("Mas gostaria de deixar absolutamente claro que...") (Google)</p>
<p>D. intensificador de determinadas palavras negativas</p>	<p>-no ("The event had absolutely no emotional effect on me") (BNC) -nothing ("David said absolutely nothing, not a word") (BNC) -not: ("I'm positively, absolutely not pregnant") (BNC)</p>	<p>D. intensificador de determinadas palavras negativas</p>	<p>-nenhum: ("Não tínhamos absolutamente nenhum interesse em assinar contratos.") (Google) -nenhuma: ("Não há absolutamente nenhuma necessidade de criar um mundo virtual") (Google) -nada: ("Os ateuas não esperam provar absolutamente nada".) (Google)</p>
<p>E. raramente ocorre em fraseologia negativa Obs: "entirely" aparece com frequência em construções negativas (<i>not entirely pleased</i>)</p>		<p>E. ocorre em fraseologia negativa.</p>	<p>não + absolutamente+ adjetivo: ("Quando o evento não é agendado como o Carnaval ou a Copa do Mundo, elas não estão absolutamente preparadas para uma cobertura total.") (Google)</p>

5. Conclusão

Este estudo procurou num primeiro momento, identificar a padronização formal de um item e as funções correspondentes, “por função entendendo-se o resultado da descrição dos padrões de um item de interesse na L1 segundo seu ambiente colocacional, coligacional, de preferência e de prosódia semântica” (Berber-Sardinha, 2002: 35); num segundo momento identificar um equivalente tradutório *prima facie* (aparente, sem maiores análises) para cada função; num terceiro momento seguir os mesmos passos do primeiro, só que no corpus da L2, da segunda língua presente na pesquisa.

O que percebemos é que embora à primeira vista alguns itens pareçam equivalentes, são diferenciados bem claramente pelos padrões de co-seleção revelados nas linhas de concordância, ou seja, em seus perfis colocacionais respectivos. Embora o tradutor conte com a intuição, a evidência do corpus invariavelmente o levará a identificar os padrões de uso real; e ao permitir uma descrição dos mesmos possibilitará que proceda a uma escolha bem-informada, que não causa estranheza nem ruído e que leva em conta as diferenças de uso.

Referências Bibliográficas

- BERBER-SARDINHA, A.P. (2001/2002). Corpora eletrônicos na pesquisa em tradução. *Cadernos de Tradução*, vol.9.(2001/2002) *Tradução e Corpora*, Stella Tagnin (Org.). NUT, Universidade Federal de Santa Catarina, 35.
- BERBER-SARDINHA, A.P. (2004). Lingüística de Corpus e tradução: prosódia semântica. *Lingüística de Corpus*. Manole, 236.
- HUNSTON, S.; FRANCIS, G.(2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdã/Filadélfia, John Benjamins, 104-108.
- KENNEDY, G. (1998). *An Introduction to Corpus Linguistics*. Nova York, Longman, 9.
- PARTINGTON, Alan.(1998). Introduction, (2) Collocation and Synonymy, (3) True and False Friends. *Patterns and Meanings: using corpora for English language research and teaching*. Amsterdã/Filadélfia, John Benjamins,1-14,29-47, 48-64.
- PRATA, Mário. *Schifaiçfavoire* – Verbetes – Nova página 3.: www.marioprataonline.com.br/obra/literatura/adulto/dicionário
- STUBBS, Michael.(2001). *Words and Phrases: corpus-based studies of lexical semantics*. Oxford, Routledge.
- TAGNIN, Stella E.O.(2001/2002). Os Corpora: Instrumentos de Auto-ajuda para o tradutor. *Cadernos de Tradução*, vol.9. (2001/2002) *Tradução e Corpora*, Stella Tagnin (Org.). NUT, Universidade Federal de Santa Catarina, 193-194.
- TOGNINI-BONELLI, Elena. (2001). (2) Language teaching, (7) Working with corpora across languages. *Corpus Linguistics at Work*. Amsterdã/Atlanta, John Benjamins, 14-46, 131-156.

LAMPARELLI, Alvamar Helena de Campos Andrade. *Absolutamente e Absolutely – São ou Não São Absolutamente Intercambiáveis?*

Dicionário Eletrônico Houaiss da Língua Portuguesa. Versão 1.1 – Dezembro de 2001.
Editora Objetiva.

Dicionário Eletrônico Novo Aurélio 3.0. Editora Nova Fronteira.

Lingüística de Córpus e Lexicografia Bilíngüe: O Caso Experimental de *Como* e suas Traduções para o Francês

*Adriana Zavaglia**

Resumo: *Este artigo tem como escopo apresentar um verbete bilíngüe experimental português – francês para o marcador como utilizando, como fundamentos e suportes principais para a sua elaboração, expedientes provindos da lingüística enunciativa de Culioli (1999a, 1999b, 2000) e da lingüística de corpus, tentando abordar a lexicografia bilíngüe de forma diferencial, dando especial atenção à definição e à contextualização bilíngüe do marcador.*

Palavras-chave: *lexicografia; corpus; tradução; definição; contextualização.*

Abstract: *This paper aims at presenting a bilingual experimental Portuguese-French dictionary entry of the marker "como". The main theoretical foundations of this new proposal are based on Culioli's enunciative linguistics as well as on Corpus Linguistics. This work represents an attempt to approach bilingual lexicography in a differential manner, thus paying special attention to the definition and to the bilingual contextualization of the marker.*

Keywords: *lexicography; corpus; translation; definition; contextualization.*

* USP/CITRAT – Bolsita Pós-Doc. FAPESP.

ZAVAGLIA, Adriana. *Linguística de Corpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

Introdução

Ao serem confrontados um verbete de dicionário monolíngüe com um verbete de dicionário bilíngüe (ou multilíngüe), percebe-se de imediato que cada um guarda características próprias que determinam a sua singularidade: enquanto a microestrutura do verbete monolíngüe conta com definições e contextualizações para as diversas acepções da entrada, o bilíngüe a preenche basicamente com opções na segunda língua, as quais são tratadas como correspondentes, equivalentes ou sinônimos da entrada, que não é definida nem contextualizada. Segundo Rey-Debove, aliás, a definição só se mostra necessária num dicionário bilíngüe “quando aquilo que tem nome numa língua não o tem na outra” (1984:67). Entretanto, essa característica parece não satisfazer por completo o consulente do dicionário bilíngüe. Como dissemos alhures (Zavaglia, 2000:844):

Os consulentes que fazem uso desses dicionários, sejam eles estudantes ou professores de língua, tradutores ou simples curiosos, muitas vezes passam do dicionário bilíngüe para o dicionário monolíngüe por não compreenderem em que contextos podem utilizar determinada palavra, ou seja, faltam-lhes uma definição cuidadosa, exemplos mais acurados e bem contextualizados; ou então percorrem o caminho inverso, indo do dicionário monolíngüe para o bilíngüe, porque entenderam o sentido da palavra procurada em contexto mas faltam-lhes correspondentes tradutórios. Assim, o trabalho de passar de uma língua para outra torna-se fastidioso.

Para solucionar esse tipo de problema, já em 1959, Jakobson apontava, com dizeres precursores, para “a necessidade urgente, a importância teórica e prática de dicionários bilíngües diferenciais, que definam cuidadosa e comparativamente todas as unidades correspondentes, em sua extensão e profundidade” (1995:66).

Seguindo a linha diferencial sugerida por Jakobson e inspirando-nos nos procedimentos lexicográficos utilizados em dicionários monolíngües brasileiros recentes (cf. Borba 1991, 2002), estamos buscando, desde 2000 (cf. Zavaglia 2000, Zavaglia & Zavaglia 2000), trazer à baila, no universo da lexicografia bilíngüe ou trilingüe, uma discussão em torno de duas dimensões até então características apenas da lexicografia monolíngüe ou da terminografia: a definição e a contextualização. Para levar a cabo tal projeto, incorporamos a ele recentemente mais três dimensões: a da tradutologia, com as Modalidades Tradutórias de Aubert (1998), a da lingüística enunciativa, com a Teoria das Operações Enun-

ciativas de Antoine Culioli (1999a, 1999b, 2000), e a da lingüística de córpus, com uma ferramenta computacional chamada WordSmithTools.

Para trabalhar esses dois aspectos em lexicografia bilíngüe, há três caminhos possíveis: partir de dicionários monolíngües, bilíngües ou multilíngües, de córpus comparáveis ou de córpus paralelos. De imediato, não nos pareceu interessante realizar a pesquisa a partir de outros dicionários, uma vez que não seria possível cercar as lexias do ponto de vista contextual ou da definição, já que não teríamos material suficiente para servir de base às nossas análises que nos fornecesse uma relação direta entre ocorrências em contexto. Esse tipo de trabalho, muito comum em lexicografia bilíngüe, dá origem ao que Desmet denomina de *falsos bilíngües* (2004). Tampouco mostrou-se pertinente aos nossos propósitos – uma vez que o nosso experimento visa à contextualização autenticamente bilíngüe (paralela) e a uma definição do marcador totalmente dedutiva (dirigida pelo córpus) – a realização da pesquisa a partir de textos escritos em cada uma das línguas ou, em outras palavras, a partir de córpus comparáveis. Confrontado com a primeira opção, esse é um caminho mais interessante, porém sujeito, a nosso ver, a muitas intervenções por parte do lexicógrafo baseadas tanto em sua experiência pessoal como também em consultas a dicionários. Todavia, a abordagem lexicográfica a partir de córpus comparáveis pode dar bons resultados dependendo do ponto de vista adotado (cf. em particular Tognini-Bonelli, 2001). No nosso entender, somente o córpus paralelo pode nos fornecer uma relação contextual autenticamente bilíngüe entre agenciamentos de duas línguas distintas, já que a tradução é, de certa forma, uma resposta espontânea aos problemas lingüístico-culturais que advêm do contato entre duas línguas que queremos observar, mapear e solucionar. Além disso, a nosso conhecimento, não há, dentre os trabalhos atuais em lexicografia bilíngüe, pesquisas – com resultados já publicados – que tenham sido feitas exclusivamente a partir de córpus paralelos. Por este último motivo, qualificamos o verbete que apresentaremos posteriormente de *experimental*. Note-se, igualmente, que já aplicamos a metodologia e os procedimentos descritos e utilizados neste artigo em outros trabalhos cujos resultados revelaram-se promissores (cf. Zavaglia 2004a, 2004b, 2004c). Faz-se importante observar que, pela natureza do córpus utilizado, o verbete apresentado tem como público hipotético lusófonos brasileiros aprendizes de francês ou francófonos franceses aprendizes de português e profissionais (tradutores e professores) das duas línguas.

Dados os aspectos anteriormente discutidos, temos como escopo apresentar neste artigo um verbete bilíngüe experimental português – francês para o marcador *como* utilizando como fundamentos e suportes principais para a sua

ZWAGLIA, Adriana. *Linguística de Corpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

realização expedientes provindos da linguística enunciativa de Culioli (1999a, 1999b, 2000) e da linguística de corpus, tentando abordar a lexicografia bilíngüe de forma diferencial, dando especial atenção à definição e à contextualização bilíngüe do marcador.

O corpus e o seu tratamento inicial

De uma pequena base de dados paralelos (29.653 palavras, das quais 13.611 em português – conto integral “Minha Gente” de *Sagarana* – e 16.042 em francês – tradução integral do conto “Minha Gente” para o francês, intitulada “Les Miens”), extraímos um corpus paralelo constituído de enunciados que continham o marcador *como* em relação de tradução, ou seja, enunciados em português e em francês. Para manusear esse corpus eletronicamente, utilizamos o programa WordSmithTools.

A primeira etapa de nossa análise consistiu em fazer a concordância do marcador *como*, nesse caso apenas no texto em português, com cinco co-ocorrências à direita e cinco à esquerda. Obtivemos como resposta 36 concordâncias de *como*. Num primeiro momento, já se observaram no corpus algumas das particularidades do marcador, como o seu perfil e o seu padrão colocacional (com os seus colocados mais freqüentes), que se mostraram ainda mais evidentes com o auxílio da lista de palavras oferecida pelo programa, e o seu perfil semântico-funcional. Arriscando uma leitura rápida dos dados colhidos, a qual poderá ou não se confirmar, poderíamos dizer, por exemplo, que *como* tem grandes chances de funcionar como um elemento lógico de ligação entre duas entidades, sejam elas palavras ou orações. Com essas informações prévias, demos início a uma primeira análise enunciativa de *como* que conduziu ao agrupamento dos enunciados de acordo com o seu funcionamento.

A segunda etapa do tratamento do corpus foi o seu alinhamento: feitas as análises de *como* exclusivamente em português – observe-se que a direção de tradução almejada parte do português para o francês – voltamos para o marcador em contexto de tradução com o intuito de verificar as relações, as suas variações, os seus comportamentos.

Somente após essas duas etapas pudemos construir um verbete bilíngüe diferencial para *como*. Contudo, antes de passarmos à descrição e explanação dos procedimentos e da metodologia empregada, apresentaremos resumidamente a seguir os expedientes provindos da Teoria das Operações Enunciativas (cf. Culioli 2000, 1999a, 1999b) que utilizamos para levar a cabo nossas primeiras observações analíticas.

Fundamentos teórico-lingüísticos

A partir de alguns dos conceitos culiolianos, dentre os quais o de *noção*, cremos ser possível analisar o corpus escolhido, interpretando-o qualitativamente em contexto. Essa primeira fase é essencial para definir coerentemente o marcador, de forma que sua definição possa ser aplicada à microestrutura do verbete bilíngüe que pretendemos elaborar.

As noções, que são sistemas complexos de propriedades físico-culturais, não coincidem com as palavras; *elas são captadas* pelas palavras e com elas não se correspondem termo a termo. Culioli elaborou esse conceito com vistas a dar conta da plasticidade dos fenômenos lingüísticos, cujas formas empíricas (palavras e agenciamento de palavras) não têm relação direta com formas semânticas. Isso quer dizer que qualquer palavra (*criança*, por exemplo) ou conjunto de palavras só assume determinado sentido quando de sua efetiva enunciação. Desse modo, *criança* não tem apenas um, mas vários sentidos, dependendo do contexto e do contexto em que ocorre.

Todo indivíduo constrói noções; cada indivíduo tem, portanto, uma representação mental para *criança* elaborada a partir de ocorrências *fenomenológicas* (crianças que viu, que ouviu, que tocou, etc.) e *abstratas* (enunciados contendo a palavra *criança* que ouviu) das quais extrai simbolicamente propriedades. A partir dessas propriedades, o sujeito tipifica a sua representação mental de *criança*. Durante o processo de tipificação, que não é finito, o indivíduo constrói uma listagem de ocorrências abstratas de *criança* dentre as quais uma terá uma posição privilegiada e será chamada de pólo de referência ou P (ou o predicado <ser-criança>). Em torno de P vão circular as outras ocorrências abstratas, que se organizarão de forma a constituir o que Culioli chamou de domínio nocional, representado metalingüisticamente da seguinte maneira: $p_i p_j p_n$. As ocorrências $p_i p_j p_n$ remetem às mesmas propriedades, já que são ocorrências de P. Se quisermos contrapô-las a outras ocorrências para diferenciá-las quanto às propriedades em questão, podemos lançar mão do complementar P', ou de suas ocorrências, $p'_i p'_j p'_n$, que estão no exterior do domínio de P, ou seja, não remetem a P mas a P'. Se quisermos identificar ocorrências, acrescentando a elas outras propriedades, podemos utilizar um outro símbolo na escritura, Q, por exemplo, que pode ou não ser identificado a P.

A partir dessa escritura de base, da qual derivam muitas outras, afinaremos nossa análise incorporando a ela aspectos ligados à intersubjetividade.

ZWAGLIA, Adriana. *Linguística de Córpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

Grupos funcionais do marcador *como*

Como já dissemos, os resultados da análise do córpus pelas ferramentas oferecidas pelo WordsmithTools, as quais permitiram a concordância e o levantamento dos colocados do marcador, facilitaram e simplificaram a análise enunciativa, feita enunciado a enunciado. Por essa análise, constituímos os seguintes grupos funcionais (em ordem decrescente considerando o seu valor de frequência de sua ocorrência nos dados analisados):

Valor referencial A: *como* marca uma relação de proporção comparativa entre dois termos, conduzindo uma propriedade Q, de domínio desconhecido, ao domínio de P, conhecido, aproximando as propriedades de Q às propriedades de P. (O enunciador conhece Q e P; o enunciatário conhece apenas P.)

Exemplo: *A fruta mal madura da cagaiteira, comida com sol quente, tonteia COMO cachaça.*

O tontear-da-fruta-madura-da-cagaiteira-comida-com-sol-quente (informação que o enunciatário não compartilha com o enunciador) é feito o tontear-da-cachaça (informação que o enunciatário compartilha com o enunciador).

Q tonteia da mesma maneira que P.

Valor referencial B: *como* marca uma relação de proporção por inferência entre dois termos, entre um domínio P, evidente, e um domínio Q, esperado, inferindo Q através de P. (O enunciador e enunciatário conhecem P e o enunciador espera Q.)

Exemplo: – *Mas você, casado COMO é, não tem vergonha de andar com outra mulher?*

Já que você é casado (informação que enunciador e enunciatário compartilham), você não pode andar com outra mulher (inferência do enunciador).

Já que P, então Q.

Valor referencial C: *como* marca uma relação de proporção sem estabilização entre dois termos, entre dois domínios, P e Q, com varredura apenas em P. (O enunciador não conhece Q, conhece apenas P e tenta estabilizar a relação entre Q e P com o recurso ao enunciatário.)

Exemplo: – *COMO é ela?*

Crop, 10, 2004

Não sei de que maneira ela é (O enunciador não consegue estabelecer uma relação entre P e Q por si só; recorre, então, ao enunciatário.)
De que Q é P?

Os 36 enunciados – note-se mais uma vez que até então trabalhamos apenas com dados extraídos do texto em português – distribuem-se nos três grupos acima apresentados. Vejamos qual a relação de frequência das ocorrências dos valores referenciais assumidos por *como*:

<i>COMO</i>		
Valor Referencial	Número de Ocorrências	Frequência das Ocorrências (%)
A - comparação	22	61,11
B - inferência	7	19,44
C - varredura	7	19,44
Total	36	100

Do quadro acima podemos retirar os valores de frequência de ocorrência do marcador com relação aos valores referenciais que ele assume em contexto e co-texto específico. O resultado dessa observação nos conduzirá a uma ordem em que aparecerão os valores referenciais no verbete. No caso em questão, o valor A deverá aparecer em primeiro lugar e com mais abonações, e os valores B e C, por terem apresentado um mesmo número de ocorrências no corpus analisado, em segundo ou terceiro lugar e com menos abonações.

Traduções de *como* no corpus analisado e sua relação com os valores referenciais

Feita a análise do marcador *como* em contexto monolíngüe, passamos a trabalhar verdadeiramente com o corpus paralelo, alinhando os *paralelismos* (para mais detalhes sobre esse conceito, cf. Zavaglia 2004c). Obtivemos as seguintes relações do marcador *como* com as suas traduções para o francês:

ZWAGLIA, Adriana. *Linguística de Corpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

Relação entre valores referenciais e traduções			
Valor Referencial	Como traduzido para o francês	Número de Ocorrências	Freqüência das Ocorrências (%)
comparação	comme	18	50
	tel, telle	2	5,55
	tout pareil que	1	2,77
	que	1	2,77
inferência	comme	6	5,66
	en	1	2,77
varredura	comment	7	19,44
Total	comme, tel/telle, tout pareil que, (aussi) que; comme, en; comment	36	100

A elaboração desse quadro de freqüência nos foi essencial, conjuntamente com a do quadro anterior sobre a relação de freqüência das ocorrências dos valores referenciais assumidos por *como*, para a elaboração do verbete: os valores de freqüência das traduções de *como* nos indicou o número de abonações bilíngües que deveria constar para cada caso e também a sua ordem de inserção; respeitamos, assim, um princípio de proporcionalidade. Nesse caso, o Valor A contará com 10 abonações (7 paralelismos *como-comme* e 1 dos demais), o valor B com 4 (3 paralelismos *como-comme* e 1 *como-en*) e o Valor C também com 4 (todas apresentando o paralelismo *como-comment*).

Apresentação do verbete bilíngüe diferencial de *como*

Transpondo de forma experimental os resultados da pesquisa anteriormente apresentados para a estrutura de um verbete diferencial português-francês, com a consideração do valor de freqüência e com a apresentação da entrada, no caso *como*, definida e contextualizada de maneira bilíngüe, temos a seguinte arquitetura:

COMO (invariável) prop. 1. COMO marca uma comparação entre propriedades de dois conjuntos semânticos distintos // COMO marque une comparaison entre les propriétés de deux ensembles sémantiques distincts: **comme, tel(le), (tout) pareil que, (aussi) que** **1.1** [...] *A fruta mal madura da cagaiteira, comida com sol quente, tonteia COMO cachaça // [...] le fruit à peine mur de la colicaire, mangé sous un soleil brûlant, enivre COMME de la cachaça* **1.2** *Santana, ledor de Homero e seguidor de Aibókhin, também, COMO um e outro, cochilou. // Santana, lecteur d'Homère et disciple d'Aillokhine, également, COMME l'un et l'autre, s'est assoupi.* **1.3** *Os mangues da outra margem jogam folhas vermelhas na corrente. Descem COMO canoibas. // Les manguiers de l'autre rive jettent des feuilles rouges dans le courant. Elles descendent COMME de petites barques.* **1.4** *É boa tática...Um "gambito do peão da Dama", COMO Santana diria... // C'est une bonne tactique... « Un gambit du pion de la reine », COMME dirait Santana...* **1.5** *Estava tosando ar alto, mas nós olhávamos o vôo COMO quem se inclina para espionar um peixe num aquário. // Il tondait l'air des hauteurs, mais nous regardions son vol COMME qui se penche pour observer un poisson dans un aquarium.* **1.6** *E graças aos encontros inesperados dos velhos amigos que eu fico reconhecendo que o mundo é pequeno e, COMO sala-de-espera, ótimo, facilimo de se aturar... // Et c'est grâce aux retrouvailles inattendues de vieux amis que je reconnais que le monde est petit et, COMME salle d'attente, très agréable, très facile à supporter...* **1.7** - *Pororoca! Será que ninguém aqui pensa COMO eu?!... // -Quel méli-mélo ! Il n'y a donc personne ici pour penser COMME moi?...* **1.8** [...] *as árvores ficavam tão quietas, que aquele campo parecia correr, COMO um vau de riacho raso, de transparência movente. // [...] les arbres se tenaient si tranquilles que cette campagne semblait courir, TEL un gué de ruisseau lisse, à la transparence mouvante.* **1.9** *E ambos corriam do assunto e voltavam ao assunto, e era bem COMO na estória da onça e do veado, que, alternadamente e com muita confiança em Deus, construíram uma casa, ignorando-se mutuamente a colaboração. // Et tous deux éludaient le sujet et revenaient au sujet, et c'était tout PAREIL QUE dans l'histoire du jaguar et du cerf qui, à tour de rôle et avec moult confiance en Dieu, avaient construit une maison, chacun ignorant leur mutuelle collaboration.* **1.10** *E Santana estende-me a carteirinha, porque há também a carteirinha, o xadrezinho de bolso, que eu me esquecia de mencionar; tão permanente na algibeira do meu amigo COMO os óculos de um míope na cara de um míope. // Et Santana m'a tendu le modèle réduit, car il y a aussi le jeu en miniature, le petit échiquier de poche, que j'allais oublier de mentionner; aussi permanent dans la poche de mon ami QUE les lunettes d'un myope sur le visage d'un myope.* **2.** COMO marca uma relação dedutiva entre dois termos // COMO marque une relation déductive entre deux termes: **comme; en.** **2.1** *E, COMO foi em honra deles dois, que são meus amigos, faço*

ZWAGLIA, Adriana. *Linguística de Corpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

questão de que eles sejam os padrinhos!... // Et COMME c'est en leur bonheur à tous les deux, qui sont mes amis, je tiens à ce qu'ils en soient les parrains!... 2.2 ... porque, COMO diz o capiau conterrâneo, "a minha parte de histórico eu prefiro em dinbeiro!"... // ...parce que, COMME dit le paysan de chez nous, «ma part de gloire, je la préfère en argent comptant!»... 2.3 [...] e, COMO as eleições estão próximas, tudo vai muito intenso e muito alegre, a maravilhas mil. // [...] et, COMME les élections sont proches, ça y va à toute vapeur, à coeur joie et à merveille. 2.4 COMO bom capiau, Bento Porfírio acha que ainda é cedo para me avisar. // EN bon péquenot, Bento Porfírio estime qu'il est encore trop tôt pour m'en informer. 3. COMO marca uma tentativa de qualificação de um termo por outro // COMO marque une tentative de qualification d'un terme moyennant un autre: comment. 3.1 – Mas, COMO é que você pode saber isso tudo, José? – indagueu Santana, surpreso. // -Mais COMMENT est-ce que tu peux savoir tout ça, José? a demandé Santana, étonné. 3.2 COMO é ela? É alta? // COMMENT est-elle? Grande? 3.3 Pergunta COMO é que vai o Juquinba... // Demande COMMENT va le Juquinba... 3.4 Não sei COMO foi: um grito de raiva, uma pancada, o t'bum n'água de uma queda pesada, como um pulo de anta. // Je ne sais pas COMMENT ça s'est passé: un cri de rage, un coup, le blouf dans l'eau d'une chute lourde, comme un plongeon de tapiro.

A entrada do verbete vem seguida de uma informação importante: o marcador em questão, *como*, é invariável. Após essa informação, aparece a abreviação da operação enunciativa (ou funcional) geral marcada por *como*; se o consulente tiver curiosidade, ele poderá ler a introdução do dicionário do qual faz parte o verbete e consultar a lista de abreviaturas. Cada valor referencial do marcador é definido de forma mais simples em português e em francês, e contextualizado também nas duas línguas. As abonações permitem ao consulente o contato imediato com os padrões colocacionais do marcador em português (co-texto e co-ocorrências) e com os padrões paralelos de *como* (o par *como / comme* para comparar é mais freqüente; o par *como / comment* é utilizado para interrogar, etc.). As abonações, que vão funcionar como enunciados prototípicos desses funcionamentos, não são *escolhidas* pelo lexicógrafo, mas sim dadas pelos valores de proporção e freqüência retirados do próprio corpus de estudo. Desse modo, os funcionamentos específicos do marcador que apareceram no corpus são registrados mediante critérios previamente estabelecidos.

O acréscimo da definição e da contextualização à microestrutura faz desse verbete um elemento radicalmente diferente daquilo que se pode encontrar em outros dicionários de mesmo teor. Para levar a cabo essa comparação, elencamos abaixo verbetes de *como* extraídos de três dicionários bilíngües brasi-

leiros¹ português/francês – observe-se que nos dois primeiros, em ordem de apresentação, nada se diz, nos textos que introduzem, apresentam ou prefaciam as obras, sobre os fundamentos teóricos que nortearam a sua elaboração ou sobre os critérios utilizados na sua construção:

Burtin-Vignoles (2003:576):

como, adv. Comme, de même que; vu que; suivant que; tant que, autant que. BRANCO _ A NEVE, blanc comme la neige. VEJA _ ELE TRABALHA, voyez, comme il travaille. TANTO DE UM _ DE OUTRO, autant de l'un comme de l'autre. ASSIM _, comme. _ COMO QUER QUE SEJA, quoi qu'il en soit. PROCEDER _ MÃE, agir en mère. NÃO SEI _ É, je ne sais pas comment cela est. ASSIM _ EU, tout comme moi. _ É LINDA ESTA CRIANÇA! Que cet enfant est beau! _ FOI QUE ACONTECEU? Comment cela est-il arrivé?

Rónai (1989:343):

como adv. Comment; (*em comparações*) comme; en; *falo-lhe _ amigo* je vous parle en ami; *como?* (*para ouvir melhor uma resposta*) vous dites?; *plait-il?*; *interj.* comment!

Signer (1998:286):

como conj + adv comme; puisque; que. *adv* combien; tel. **nada como** = rien de tel. *adv + nm* comment. **como vai?** = comment allez-vous?. *interj* comment!

Embora possam ser de alguma serventia para o consultante, principalmente quando este busca complementar a pesquisa feita anteriormente em dicionário monolíngüe, percebe-se que as opções que lhe são oferecidas nos dicionários acima, descontextualizadas (com alguma exceção, talvez, para as locuções e exemplos – visivelmente forjados – que aparecem nas microestruturas) e sem definição, podem não solucionar o problema que procura resolver. Descontextualizada, a palavra permanece sem sentido, já que sem contexto, não há sentidos, e sem sentidos, não há definição.

À diferença desses verbetes, os paralelismos que apresentamos na microestrutura de nosso verbete experimental estão de acordo, a nosso ver, com aqui-

¹ Embora o dicionário português de Domingos de Azevedo, *Grande Dicionário Português-Francês*, registre variantes brasileiras, preferimos permanecer, até mesmo por falta de espaço, apenas no contexto lingüístico do Brasil.

ZWAGLIA, Adriana. *Linguística de Córpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

lo que Tognini-Bonelli (2001:19) chamou de *unidades de sentido ampliadas*, ou seja, uma palavra tem sentido na medida em que se relaciona com um co-texto particular. Os elementos desse co-texto acabam, ademais, formando com a palavra um padrão, já que *costumam* acompanhá-la, aparecendo ao seu lado ou nas proximidades. Por esse motivo, a definição e a contextualização, objeto principal de nossa pesquisa, cumprem, a nosso ver, um papel essencial na explicitação da variação semântica do marcador e de suas possíveis traduções num verbete bilíngüe.

Considerações finais

Pelo verbete bilíngüe português – francês de *como*, resultado de um procedimento metodológico particular fundamentado em dois ramos da lingüística também singulares, um em franco desenvolvimento – a lingüística de córpus – e outro ainda mal descoberto – a lingüística enunciativa de Culioli, apresentamos neste artigo a possibilidade real de co-textualização, contextualização e definição das lexias-entrada de dicionários bilíngües através da pesquisa direta em córpus paralelos, com opções oferecidas ao consulente na microestrutura do verbete algumas vezes inusitadas (o par *como* e *pareil que*, por exemplo), que podem não aparecer em dicionários contruídos a partir de outros dicionários ou a partir de córpus comparáveis. Segundo o nosso ponto de vista, a explicitação da variação semântica da lexia-entrada e a sua relação com outras lexias em outras línguas só é possível pela definição e pela contextualização, recursos esses imprescindíveis para resolver o problema da polissemia em contexto autenticamente bilíngüe. No entanto, o procedimento adotado também apresenta desvantagens: i) demanda da parte do lexicógrafo uma observação detalhada dos dados que lhe toma muito tempo e ii) permite-lhe registrar apenas os paralelismos que ocorreram no córpus (outras possibilidades podem não constar). Para tentar contornar esses problemas em busca de representatividade e confiabilidade dos dados, com a intenção de construir um dicionário bilíngüe português – francês de marcadores léxico-gramaticais, necessitamos: ampliar o córpus tanto em número, com mais ocorrências, quanto em qualidade, com textos tipologicamente distintos (literários, jornalísticos, técnicos etc.). Embora ainda em fase experimental e apesar das falhas e obstáculos até então detectados, tal procedimento tem demonstrado que, ao lado de outras abordagens, o futuro da lexicografia bilíngüe diferencial dirigida por córpus paralelos mostra-se bastante promissor.

Referências Bibliográficas

- AUBERT, Francis Henrik. Modalidades de tradução: teoria e resultados. *TradTerm*, São Paulo, n.1 (ano 5), p.99-128, 1sem. 1998.
- BORBA, Francisco da Silva (Coord.). *Dicionário de usos do português do Brasil*. São Paulo: Ática, 2002.
- _____. *Dicionário gramatical de verbos do português contemporâneo do Brasil*. São Paulo: Edit. da Unesp, 1991.
- BURTIN-VINHOLES, Suzanne. Dicionário francês – português, português – francês. 40. Ed. São Paulo: Globo, 2003.
- CULIOLI, A. *Pour une linguistique de l'énonciation: opérations et représentations*. 2. ed. rev. Paris: Ophrys, v.1, 2000.
- _____. *Pour une linguistique de l'énonciation: formalisation et opérations de repérage*. Paris: Ophrys, v.2, 1999a.
- _____. *Pour une linguistique de l'énonciation: domaine notionnel*. Paris: Ophrys, v.3, 1999b.
- DESMET, Isabel. Comunicação pessoal – Palestra e seminários sobre terminologia e terminografia ministrados na FFLCH – USP, 1sem. 2004.
- JAKOBSON, Roman. Aspectos lingüísticos da tradução. In:_____. *Lingüística e comunicação*. Tradução de Izidoro Blikstein e José Paulo Paes. São Paulo: Cultrix, 1995. p.63-72.
- REY-DEBOVE, Josette. Léxico e dicionário. Tradução de Clóvis Barleta de Morais. *Alfa*, Araraquara, v. 28, 1984, p.45-69.
- RÓNAI, Paulo. Dicionário francês / português – português / francês. 2.ed. Rio de Janeiro: Nova Fronteira, 1989.
- ROSA, G. *Sagarana*. 10 ed. Rio de Janeiro: Nova Fronteira, 2001.
- ROSA, G. *Sagarana*. Trad. de Jacques Thiériot. Paris: Albin Michel, 1997. (Les Grandes Traductions)
- SIGNER, Rena. Dicionário brasileiro: francês – português, português – francês. São Paulo: Oficina de Textos, 1998.
- TEIXEIRA, Elisa. *Apostila explicativa sobre o WordSmithTools*. (inédito)
- TOGNINI-BONELLI, Elena. *Corpus linguistics at work*. Amsterdam/Philadelphia: John Benjamins, 2001.
- ZAVAGLIA, Adriana. Por uma lexicografia bilíngüe diferencial. In: DURÃO, A.B.A.B. *Lingüística contrastiva: teoria e prática*. Londrina: Moriá, 2004a. p.169-177.
- _____. Lexicografia bilíngüe e tradução: por uma abordagem diferencial. *Anais do III Congresso Ibero-americano de Tradução e Interpretação – CIATI*, São paulo, p.1-11, 2004b.
- _____. A relação entre original e tradução no contexto da lexicografia bilíngüe dirigida por corpus: o conceito de paralelismo. (texto integral da comunicação apresentada no 52 Seminário do GEL, Campinas, 2004c)

ZAVAGLIA, Adriana. *Linguística de Corpus e Lexicografia Bilíngüe: O Caso Experimental de Como e suas Traduções para o Francês*.

_____. A elaboração de um verbete-modelo para a construção de um dicionário temático bilíngüe de cores francês – português. *Estudos Linguísticos* (Bauru), p.843-848, 2000.

ZAVAGLIA, Adriana; ZAVAGLIA, Claudia. A elaboração de um dicionário trilíngüe de cromônimos italiano-português-francês / francês-português-italiano: reflexões e considerações. *Linguística* (Alfal), São Paulo, v. 12, p. 235-247, 2000.

A Construção do Sentido em Vocabulários Técnicos: o Uso de Corpora e Outros procedimentos

Guilherme Fromm*

Resumo: *A construção do sentido, em vocabulários técnicos, pode ser alcançada através de vários procedimentos de análise semântica, computadorizados ou não. Mostraremos, aqui, alguns desses procedimentos e quais são as tendências atuais por parte dos terminólogos.*

Palavras-chave: *Terminologia; terminografia; semântica; lingüística de corpus.*

Abstract: *Construction of meaning in technical dictionaries can be achieved through a number of semantic analysis procedures, whether with or without the aid of computers. This paper aims at presenting some of these procedures and some of the new trends in the field of terminology.*

Keywords: *terminology; terminography; semantics; corpus linguistics.*

Apresentação

A construção do sentido, em qualquer perspectiva, é sempre uma tarefa inglória. Parece nunca haver consenso quanto à metodologia de pesquisa, ao levantamento de dados e ao resultado obtido. O objetivo deste trabalho é mos-

* Doutorando do Programa de Estudos Lingüísticos e Literários da FFLCH/USP.

FROMM, Guilherme. *A Construção do Sentido em Vocabulários Técnicos*.

trar, através de uma proposta de construção de vocabulários técnicos, algumas teorias de como construir o sentido. Essa construção será discutida em alguns aspectos e teorias abordadas pela Semântica e pela Lingüística de *Corpus*.

O objetivo inicial

Nossa pesquisa de mestrado (Fromm, 2002) procurou montar uma proposta de *glossário* técnico de informática para tradutores. Partindo da abordagem metodológica de Barbosa (2001), renomearemos, a partir deste instante, essa proposta para *vocabulário* técnico. Para o nosso doutorado, continuamos com a idéia de montar apenas uma proposta, porém pretendemos aumentar a quantidade de áreas estudadas (Informática, Jurídica, Administração, Médica e Técnica-Industrial), tornar a base de dados bilíngüe (inglês/português) e desenvolver um programa que administre o banco de dados dos *corpora* e a construção para a estrutura dos verbetes (macroestrutura, microestrutura e o sistema de remissivas).

Antes de entrarmos nas questões da definição em si, porém, é necessário que entendamos como é feita a construção de um verbete em uma obra lexicográfica ou terminográfica.

A construção do verbete

A construção de um verbete segue basicamente três partes: a seleção da macroestrutura (todas as palavras/vocábulo que serão selecionados a partir da nomenclatura¹ da língua), a microestrutura (a definição e outros componentes) e o sistema de remissivas (que fornece as relações semânticas entre as palavras/vocábulo apresentados na obra).

A macroestrutura é levantada através de uma pesquisa ao *corpus* de especialidade, em que os critérios de freqüência e de palavras-chave² são os adotados para a seleção dessa listagem de palavras.

A construção da microestrutura (a qual denominamos artigo) pode ser esquematizada através de:

¹ O termo é aqui adotado como sendo a listagem de todas as palavras existentes no sistema de uma língua determinada.

² A determinação de palavras-chave pode ser alcançada através de um *corpus* de exclusão. Quando usamos a ferramenta de análise lexical Wordsmith Tools, por exemplo, devemos confrontar o *corpus* estudado com um *corpus* geral da língua ou com outro *corpus* de especialidade (ambos dez vezes maiores que o *corpus* de análise)

Crop, 10, 2004

Artigo={+ entrada (ou verbete) + enunciado lexicográfico (± PI + PD ± PP)³}

Em que :

- o Paradigma Informacional (PI): constituído de abreviaturas, categoria gramatical, gênero, número, pronúncia, conjugação, homônimos, etc. Ainda segundo Haensch (1982, pp. 480-501), teríamos aqui também as diferenças ortográficas, cronológicas e geográficas, a etimologia, níveis de estilo e conotações, atribuição a uma matéria ou especialidade, marcas registradas, denominações oficiais;
- o Paradigma Definicional (PD): descrevem-se os semas ou unidades de significação;
- o Paradigma pragmático (PP): contém informações contextuais como exemplos e abonações. Haensch (1982, p. 470) subdivide esse conceito entre parte sintagmática (colocações e fraseologia) e/ou parte paradigmática (sinônimos, antônimos, parônimos e hipônimos).

Existiria ainda um paradigma comum em dicionários bilíngües:

- o Paradigma de Formas Equivalentes (PFE): fornece a tradução do verbete.

Fromm, 2002, p.27.

O sistema de remissivas apresenta as relações de significado, dentro da obra lexicográfica/terminológica, entre os seus verbetes. De um modo geral, apresenta-se como um sistema informativo, em que o consulente pode estender a compreensão do verbete através da consulta a verbetes que tenham uma relação (ou semas) dentro do mesmo campo semântico ou na intersecção desse com outros campos semânticos.

Dentre os itens apresentados, trabalharemos aqui somente com aqueles que requerem um levantamento semântico para a definição (paradigma definicional).

³ Em que o sinal positivo indica a obrigatoriedade e os sinais positivo e negativo, juntos, representam a possibilidade de opção.

A construção do Paradigma Definicional

Como em qualquer obra lexicográfica ou terminográfica, a definição sempre é o ponto-chave para a construção dos verbetes⁴. A grande questão que se faz presente é: como construir esse paradigma, ou seja, como definir algo? Apresentamos, a seguir, uma relação de algumas das possibilidades de construção do Paradigma Definicional, tendo sempre em vista o objetivo a ser atingido, que é a construção de termos técnico em diferentes áreas de especialidade.

Busca em obras já publicadas

Um dos processos mais antigos para a construção de uma definição é buscá-la em outras obras lexicográficas/terminográficas. Através do levantamento e análise contrastiva entre vários dicionários/vocabulários/glossários da área específica, especificamos os semas comuns e construímos uma nova definição.

O grande problema, nesse tipo de levantamento, é saber quais obras foram realizadas sob uma perspectiva científica (ou seja, bem estruturadas teoricamente) e quais não passam de uma listagem de palavras seguidas de um sinônimo (ou antônimo). O resultado pode apresentar-se insatisfatório devido à grande mistura de estilos na criação da definição e à grande probabilidade de imprecisões nas mesmas.

Uma análise desse tipo, porém, pode ser interessante sob uma perspectiva diacrônica. Sob outros vieses, ela pode balizar os parâmetros na criação dos outros paradigmas, possibilitando que o pesquisador acrescente ou diminua a quantidade de itens nos mesmos.

Sistema de Ontologias

As ontologias, as quais Zavaglia (2002) considera “uma especificação de uma conceptualização”, são desenvolvidas, quase exclusivamente, para serem trabalhadas em processamento de linguagem natural (PLN). O PLN, teoricamente, poderia ser trabalhado de forma independente em relação ao lingüista e construiria automaticamente uma definição. As ontologias forneceriam a base para que um programa de PLN construísse essa definição.

⁴ Embora muitas obras contemporâneas, especialmente as bilíngües, no intuito de serem práticas, apresentam somente o Paradigma de Forma Equivalente (ou seja, uma possível tradução) na construção do enunciado lexicográfico.

Crop, 10, 2004

A autora nos oferece um modelo de ontologia em língua portuguesa, baseada no modelo inglês, com duas grandes divisões: Classes Fundamentais e Domínios. Ela defende, em sua tese, que uma ontologia pode ajudar a desambiguar homônimos; dentro do lexema *salto*, por exemplo, poderíamos encontrar duas acepções:

1. Salto: movimento em que se eleva bruscamente o corpo do chão para cair em outro lugar

Tipo: Movimento
Supertipo: Ato
Domínio: Geral

2. Salto: peça saliente e dura de sapato (geralmente feminino)

Tipo: Vestuário
Supertipo: Manufaturado
Domínio: Indústria de confecção

Embora, nos exemplos acima, não tenhamos uma definição criada exclusivamente pela ontologia, acreditamos que um programa desenvolvido especialmente para isso seria capaz de tanto.

Ontologias, porém, não existem na natureza, devem ser construídas e aí reside o seu grande problema. Não é difícil para um programa de PLN acessar e trabalhar uma ontologia, mas a construção da mesma requer uma grande quantidade de especialistas em diversas áreas do conhecimento, o que inviabiliza o procedimento temporal e financeiramente.

Primitivos Semânticos e Vocabulários Básicos

A idéia de primitivos semânticos, que é trabalhada exaustivamente por Wierzbicka (1996), apresenta dificuldades em uma análise terminológica. Os primitivos são:

As expressões nas línguas naturais, que, sendo em si mesmas impossíveis de definir, servem para explicar todas as expressões (realizações da língua). Esta lista de definidores deverá ser o mais limitada possível, contendo, por um lado, apenas os elementos absolutamente indispensáveis, e, por outro lado, os elementos que expliquem adequadamente todas as realizações frásicas.

Wierzbicka (1972, apud Vilela, 1983)

Dentre essas idéias básicas⁵, consideradas definidoras, temos:

⁵ Aqui considerada somente a sua primeira listagem.

FROMM, Guilherme. *A Construção do Sentido em Vocabulários Técnicos*.

Conjuntos	Palavras
Substantivos	eu, você, alguém, alguma coisa, pessoas
Determinadores	este, o mesmo, outro
Quantificadores	um, dois, muitos, todos
Predicados mentais	pensar, saber, querer, sentir
Discurso	dizer
Ações e eventos	fazer, acontecer
Avaliadores	bom, ruim
Descritores	pequeno, grande
Tempo	quando, antes, depois
Espaço	onde, cima, abaixo
Taxonomia	parte (de), tipo (de)
Metapredicados	não, poder, muito (intensidade)
Conectores Interclausais	se, porque, como

Embora sejam comuns (e até esperadas) em dicionários gerais de língua (Vilela, 1983, p. 18), definições terminológicas criadas a partir de primitivos não seriam precisas. Consideremos, na relação entre os planos da expressão e do conteúdo, definições monossêmicas possíveis somente em um glossário, são as definições polissêmicas as mais comuns em vocabulários e dicionários. Os primitivos que comporiam a definição não conseguiriam delimitar o campo do conteúdo de uma forma clara. Embora esses semas sejam de fácil compreensão, seu uso exclusivo acaba dificultando o entendimento por parte do consulente. Acreditamos, ainda, que esse fator se agravaria em um vocabulário, já que parte dos consulentes já trabalham na área e a precisão na definição é um traço esperado na obra.

Os vocabulários básicos (ou fundamentais ou de explicação) seriam as palavras mais frequentes da língua. Eles deveriam constar em todas as obras lexicográficas e terminográficas e as definições de todos os outros verbetes deveriam ser baseadas exclusivamente nesse vocabulário. Vilela cita o dicionário Procter (1981), por exemplo, que usa aproximadamente duas mil e duzentas palavras básicas (obviamente nele incluídas e definidas) para a construção do

Crop, 10, 2004

paradigma definicional de todas as outras. Novamente encontraríamos dificuldades na construção de um vocabulário, já que essas palavras básicas, provavelmente, não fariam parte da área em questão e apenas aumentariam a quantidade de verbetes do mesmo, dificultando a sua consulta.

A questão de primitivos e vocabulários básicos, acreditamos, costuma ser deixada de lado pelos terminólogos quando da confecção de suas obras. A especificidade que essas obras pedem, em virtude do público-alvo e do tamanho (geralmente reduzido), pressupõe, por parte do consulente, uma certa familiaridade com os temas e com o vocabulário básico da língua.

Definições a partir de um *corpus*

O procedimento mais adotado por parte dos terminólogos, atualmente, é trabalhar a partir de um *corpus*. Há várias abordagens que podem ser trabalhadas, mas alguns critérios, quando do trabalho terminográfico, parecem ser consensuais:

- quanto maior o *corpus*, melhor⁶;
- deve haver um balanceamento entre os textos utilizados (ou haverá a tendência para alguma das áreas que o compõem e isso influenciaria o resultado final);
- o tratamento a ser dado tem de ser computadorizado (dada a grande quantidade de dados).

A partir da coleta do *corpus*, existem possibilidades diferentes para trabalhá-lo. Apresentamos a seguir algumas delas.

FrameNet

A idéia, proposta por Atkins, Rundell & Sato (2003), toma como base as idéias de quadros e outros esquemas cognitivos (como esquemas, planos, scripts e cenários, bem resumidos em Fávero (1991)). A *FrameNet*, segundo os autores, trabalharia com os papéis semânticos que uma análise aprofundada, a partir de *corpora*, pode apresentar. Essa análise ligaria os significados de uma palavra no texto aos contextos sintáticos nos quais essa mesma palavra poderia ser manifestada.

⁶ BERBER-SARDINHA (1999) os divide entre pequenos (até 80.000 palavras), pequeno-médio (de 80 a 250 mil), médio (de 250 mil a um milhão), médio grande (de 1 a 10 milhões) e grande (mais de 10 milhões de palavras)

Temos, como exemplo, o verbo⁷ *quarrel*. A FrameNet forneceria o seguinte quadro:

QUARREL Frame		
Definition: Two (or more) people engage in a verbal disagreement. This frame is a blend of the Conversation frame and the Fighting frame.		
Core FEs:		
FE	Description	Examples
Interlocutor_1	Interlocutor_1 is the more prominent party in a verbal disagreement (when the parties are expressed disjointly).	The President DISAGREED with his top advisors.
Interlocutor_2	Interlocutor_2 is the less prominent party in a verbal disagreement (when the parties are expressed disjointly).	The President DISAGREED with his top advisors .
Interlocutors	Interlocutors are both (or all) parties in a verbal disagreement, when the parties are expressed jointly. The expression that receives this role may be a conjunction of noun phrases or a notionally plural noun phrase, or a prepositional phrase whose object is a conjunction of noun phrases or a plural noun phrase.	The President and his advisors ARGUED briefly before the summit. The lawyers BICKERED before the trial. There was a brief ALTERCATION between the lawyers .
Topic	Topic is the subject matter of the verbal interaction (this role is expressed by many predicators associated with the Communication frame).	They had a QUARREL about the seating order . We always ARGUE over money .
Lexical Units altercation.n, argue.v, argument.n, badinage.n, bicker.v, bickering.n, debate.n, debate.v, disagreement.n, disputation.n, disputatious.a, dispute.n, fight.v, parley.n, parley.v, quarrel.n, quarrel.v, quibble.v, row.n, row.v, spat.n, squabble.v, tiff.n, wrangle.n, wrangle.v.		

⁷ Apesar de apresentar-se como possível para todas as classes de palavras, notamos que a FrameNet trabalha basicamente com os verbos.

Podemos observar que essa prática fornece muito mais que uma simples definição do termo. As FE (Frame Elements) fornecem os possíveis quadros relacionados ao verbo, com respectivos exemplos (que não, necessariamente, usam o mesmo verbo proposto), a partir dos quais levantamos sinônimos.

A grande problemática, a nosso ver, seria o árduo trabalho de montar toda essa rede de relações semânticas. Pela idéia proposta, cada palavra analisada apresentaria todas as possibilidades de combinação, com todas as classes gramaticais, em todos os contextos possíveis. Teríamos definições terminológicas altamente precisas do ponto de vista semântico, porém precisaríamos gastar muito tempo de análise para traçar a rede de cada termo, o que inviabilizaria o trabalho do terminólogo como o conhecemos hoje.

Constelações Lexicais

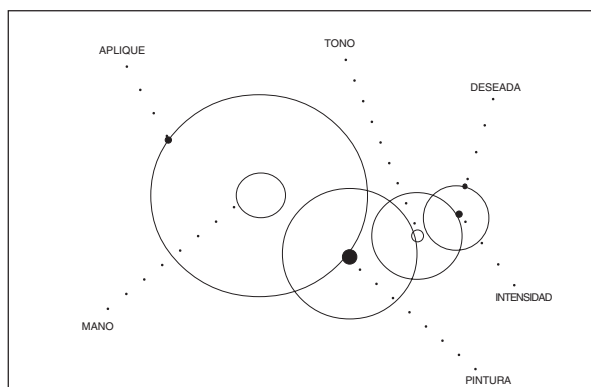
A idéia proposta por Cantos & Sánchez (2001) pretende ir além no tipo de análise com a qual os terminólogos estão acostumados: o KWIC (KeyWords in Context). Uma análise feita a partir de uma tela KWIC trabalha, geralmente, com as colocações próximas à palavra centralizada⁸. Veja um exemplo, retirado de uma monografia (Fromm, 2003):

N	Concordance	Set	Tag	ord	No	File	%
1	TEXTO FONTE Título: Manutenção de Impressoras a Jato de Tinta				17	m\po01im_p.txt	1
2	ral: Irformática Assunto espec fco: impressoras DISTRIBUIÇÃO Tipo:				65	m\po01im_p.txt	3
3	A maioria dos problemas encontrados em impressoras a jato de tinta podem ser				85	m\po01im_p.txt	4
4	ressora não esteja ligando - no caso de impressoras que possuam fonte externa,				388	m\po01im_p.txt	18
5	oras que possuam fonte externa, como as impressoras da HP. O simples teste da				395	m\po01im_p.txt	19
6	vendidos como se fossem ncvos). Das impressoras que passam pelo teste				553	m\po01im_p.txt	26
7	não se mcimentar corretamente. As impressoras normalmente utilizam uma				859	m\po01im_p.txt	41
8	axa poventura existentes. Nota: As impressoras Epson não usa graxa nos				905	m\po01im_p.txt	43
9	rica feita com anéis grifitados. Nessas impressoras não devemos usar graxa,				924	m\po01im_p.txt	44
10	namento do carro. A lubrificação dessas impressoras deve ser feita, portanto,				946	m\po01im_p.txt	45
11	selhos em relação ao papel utilizado em impressoras a jato de tinta.				1.081	m\po01im_p.txt	51
12	utilize somente papel recomendado para impressoras a jato de tinta. Esse tipo				1.092	m\po01im_p.txt	52
13	pecificado que o uso é recomendado para impressoras laser e a jato de tinta.				1.121	m\po01im_p.txt	53
14	e ser causado por dois motivos: Em impressoras da HP, a posição do carro				1.321	m\po01im_p.txt	63
15	s técnicos é fazer estoque de sucata de impressoras, ou seja, guardar				1.564	m\po01im_p.txt	74
16	sucata de impressoras, ou seja, guardar impressoras que foram condenadas para o				1.568	m\po01im_p.txt	75
17	a o reaproveitamento de peças em outras impressoras. Em muitos casos, você				1.579	m\po01im_p.txt	75
18	ras TEXTO FONTE Título: Família de Impressoras Irloprint 1000Brasil				20	em\po01imb.txt	3
19	ral: Irformática Assunto espec fco: Impressoras DISTRIBUIÇÃO Tipo:				63	em\po01imb.txt	10
20	ção: n. a. Família de Impressoras Irloprint 1000Brasil				81	em\po01imb.txt	13
21	0Brasil Família de Impressoras Irloprint 1000				86	em\po01imb.txt	14
22	mato mais amplo, são impressoras departamentais de alto				133	em\po01imb.txt	22
23	ção para luador. As impressoras da família Irloprint 1000				558	em\po01imb.txt	91
24	ral: Irformática Assunto espec fco: Impressoras DISTRIBUIÇÃO Tipo:				63	em\po02imb.txt	31
25	pressora térmica. As impressoras térmicas IBM 4400 clerecem				158	em\po02imb.txt	78
26	ral: Irformática Assunto espec fco: Impressoras DISTRIBUIÇÃO Tipo:				68	em\po03imb.txt	9
27	linha de impressoras para você. Corlira a linha				116	em\po03imb.txt	15
28	Gerenciamento remoto de várias impressoras a				161	em\po03imb.txt	21
29	lgumas impressoras a laser.				240	em\po03imb.txt	31
30	Corligura, gerencia e monlora as impressoras de				677	em\po03imb.txt	87
31	Entrega os jobs para as impressoras, fax e				687	em\po03imb.txt	88
32	Suporta uma grande variedade de impressoras.				710	em\po03imb.txt	91
33	ção de impressoras de rede.				719	em\po03imb.txt	93

⁸ Normalmente um espaço de cinco palavras para a esquerda e cinco palavras para a direita.

FROMM, Guilherme. *A Construção do Sentido em Vocabulários Técnicos*.

Os autores levantam, contudo, a questão de como elaborar um campo semântico baseado apenas nas palavras próximas ao termo estudado. Limitando a definição do termo às palavras que giram à sua volta acaba por diminuir ou apagar muitos traços distintivos deste termo. Eles propõem, então, que estude-mos as palavras-chave como uma estrela (núcleo) que atrai várias outras estrelas (colocações próximas) que atraem luas (colocações distantes). Veja o exemplo apresentado para o termo espanhol *mano*:



Uma simples consulta à uma listagem KWIC nos forneceria os semas *aplique* e *pintura* como comuns à idéia de *mano*. Essa análise detalhada, no entanto, indica que *pintura*, nos contextos estudados, também atrai as idéias de *tono*, que por sua vez atrai *intensidad* e *deseada*. Essa constelação representa um conceito ou idéia que pode aumentar o campo semântico de *mano* e melhorar sua definição.

Ficha Terminológica

Aubert (1996) nos coloca que as pistas para encontrar a definição de um termo estão nos contextos em que ele se encontra no *corpus*. Enquanto um contexto associativo nos mostra apenas a existência do termo (como pertinente ao objeto da pesquisa), contextos explicativos (em que aparecem alguns traços conceptuais) e definitórios (conjunto completo de traços conceptuais) são os elementos-chave para a elaboração de uma ficha terminológica, o instrumento mais comum com o qual o terminólogo levanta seus dados para a construção da definição.

Crop, 10, 2004

Segundo Cabré (1993, p. 281-282), as fichas terminológicas são:

... materiales estructurados que deben conter toda la información relevante sobre cada término. Las informaciones que presentan se extraen de las fichas de vaciado o de la documentación de referencia, y se representan siguiendo unos criterios fijados previamente. Hay muchos modelos de fichas teminológicas, de acuerdo com los objetivos de cada trabajo y las necesidades de cada organismo.

Uma ficha, no que tange a construção da definição, normalmente reúne os contextos explicativos e definitórios de um termo, retirados de um *corpus*, e busca, através de uma análise contrastiva, levantar os semas básicos para a confecção dessa definição.

A constituição da ficha, como já apontou Cabré (1993), depende muito do objetivo final do pesquisador. Se um simples contexto definitório for suficiente para ele, esse contexto pode ser a própria definição do termo. Se contextos definitórios não constarem no *corpus*, há a necessidade de levantamento dos semas a partir de um contexto explicativo. Vários outros elementos podem ser a ela agregados. Vejamos um exemplo fornecido por Aubert (1996) para uma ficha terminológica bilíngüe:

01 pt	02 cessão	03 cessão	04	
05 MAME	06	07	08 Sfs	09
10 Transferência de todos os diretos sobre a coisa para uma outra pessoa.				
11		12 transferência		
13 en	14 assignment	15		
16 Direito		17 Direito Civil		
18 Direito Contratual				
19 14/11/95		20 FRAU	21	
22				

Em que :

CAMPO	NOME	EXTENSÃO
Campo 01:	LP	02
Campo 02:	OCORRÊNCIAS-LP	36
Campo 03:	TERMO LP	36
Campo 04:	PADRONIZADO	01
Campo 05:	FONTE	04
Campo 06:	ANO	02
Campo 07:	USO	02
Campo 08:	MORFOLOGIA	10
Campo 09:	SINTAXE	10
Campo 10:	CONTEXTO1	250
Campo 10a:	CONTEXTO2	250
...Campo 10n	CONTEXTO n	250
Campo 11:	UNITERMOS	72
Campo 12:	SINÓNIMOS	72
Campo 13:	LC	02
Campo 14:	TERMO LC	36
Campo 15:	EQUIVALÊNCIA	1
Campo 16:	AREA	50
Campo 17:	SUBÁREA	50
Campo 18:	TEMA	110
Campo 19:	DATA	8
Campo 20:	DOCUMENTADOR	4
Campo 21:	REVISOR	4
Campo 22:	DEFINIÇÃO	250
	TOTAL	≥ 1.012

Percebemos aqui que uma única fonte foi o suficiente na construção do contexto, que serve como definição (já que o campo definidor, o de número 22, não foi preenchido). O objetivo básico da ficha é mostrar a relação entre o termo na LP (língua de partida, português no caso) e o termo equivalente na LC (língua de chegada, inglês). Para tanto, haveria a necessidade de construção de dois *corpora*, paralelos (com textos originais e suas traduções) ou comparáveis (textos originais e traduções na mesma área). Uma obra baseada nesse modelo de ficha terminológica, portanto, destacaria o Paradigma de Forma Equivalente e não o Paradigma Definicional.

Para exemplificar a construção do Paradigma Definicional, usaremos um exemplo da nossa dissertação (Fromm, 2002; ver Anexo 1). Nesse caso, o *corpus* era monolíngüe e buscava explicar os empréstimos na área de informática. Para cada termo foram levantados quatro exemplos que forneciam contextos explicativos ou definitórios. Dali foram tirados os conceitos básicos e, em análise contrastiva dos traços distintivos desses conceitos, foi elaborado um conceito final, do qual tiramos uma definição.

Considerações Finais

O trabalho de um terminólogo, na elaboração da definição de um termo para um vocabulário, transita entre dois pólos: da análise totalmente informatizada (representado pelo uso de ontologias em PLN) até a análise totalmente manual (como, por exemplo, a construção através de universais semânticos). As tendências atuais, no entanto, trabalham com uma mescla entre esses pólos e procuram dar uma ênfase no *corpus* de estudo.

Parte-se do *corpus*, bem planejado e constituído, para se chegar à definição. Ela é constituída ora através de levantamentos manuais (como a construção da definição final em uma ficha terminológica), ora através de análises computadorizadas (como as levantadas na FrameNet ou em Constelações Lexicais).

Embora as novas tecnologias que surgem diariamente facilitem o trabalho do terminólogo, acreditamos que há um longo caminho (talvez impossível) a ser percorrido até o computador conseguir reproduzir o modelo cognitivo do cérebro humano e tomar o lugar desse profissional.

Referências Bibliográficas

- ATKINS, S.B.T., FILLMORE, C.J., JOHNSON, C.R. *Lexicographic relevance: selecting information from corpus*. *International Journal of Lexicography*, 16.3, p. 251-280. Londres: 2003.
- ATKINS, S.B.T. & RUNDELL, M., SATO, H. The contribution of Framenet to practical lexicography. In: *International Journal of Lexicography*, 16.3, p. 333-357. Londres: 2003.
- AUBERT, F. H. *Introdução à metodologia da pesquisa terminológica bilingüe*. São Paulo: Humanitas Publicações-FFLCH-USP, 1996.
- BARBOSA, M. A. Dicionário, vocabulário, glossário: concepções. In: ALVES, I. M. (Org.). *A constituição da normalização terminológica no Brasil*. 2 ed. São Paulo: FFLCH/CITRAT, 2001.
- CABRÉ, M. T. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Editorial Empúries, 1993.
- CANTOS, P. & SÁNCHEZ, A. Lexical constellations: what collocates fail to tell. In: *International Journal of Corpus Linguistics*. Vol. 6(2), 2001.
- FÁVERO, L.L. *Coesão e coerência textuais*. São Paulo: Ática, 1991.
- FROMM, G. *Ferramentas de análise lexical computadorizadas: uma aplicação prática. Monografia de conclusão de curso*. São Paulo: FFLCH/USP, 2003.
- FROMM, G. *Proposta para um modelo de glossário de informática para tradutores*. Dissertação de Mestrado. São Paulo: FFLCH/USP, 2002.
- SARDINHA, T. B. *O que é um corpus representativo?* Inédito, 1999.

FROMM, Guilherme. *A Construção do Sentido em Vocabulários Técnicos*.

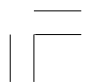
VILELA, M. . *Definição nos dicionários de português*. Porto: ASA, 1983.

WIERZBICKA, A. *Semantics: primes and universals*. Oxford: OUP, 1996.

ZAVAGLIA, C. *Análise da homonímia no português: tratamento semântico com vistas a procedimentos computacionais*. Tese de Doutorado. Araraquara: FCL/UNESP, 2002.

Anexo 1

Entrada: wireless	Formas Equivalentes: Wireless, sem fio	Cat. Gram. s	Nº s	Sing./Plural s/pl	Sigla/ Acrônimo	Entrada por extenso	Var. Morfossintáticas	Área rede	Aceção nº única	Cópus 1063
						Conceito1: rede local sem fio			Fonte OESP 08.01.2001	
						Conceito2: tecnologia de comunicação sem fio.			Fonte OESP 10.12.2001	
						Conceito3: tecnologia de conexão sem fio.			Fonte FSP 03.01.2001	
						Conceito4: sistema de telecomunicações baseado em ondas eletromagnéticas.			Fonte INFO 02.2001	
Con- ceito										
Traços Distintivos										
1	A rede	B local	C	D	E	F	G	H	I	J
2			tecnologia	comunicação	sem fio					
3			tecnologia	conexão	sem fio					
4			sistema	telecomunicações	ondas eletromagnéticas					
Conceito final: rede local, baseada na tecnologia de comunicação sem fio por ondas eletromagnéticas						Definição Dicionarizada:				
Termo Dicionarizado? () sim (x) não										
Definições coincidentes? () sim () não () parcial										
Fonte (s):										
Hiperônimo de:										
Hipônimo de: rede										
Co-hipônimo de:										
Definição: tecnologia de comunicação sem fios por ondas eletromagnéticas.						Sinônimo (s): sem fio				
						Antônimo (s):				
						Conceitos Relacionados: redes				
Notas:										Ficha nº 4



Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos¹

Diana Costa Fortier Silva*
Renato Barros da Costa**

Resumo: *Quando de sua utilização por falantes nativos de um determinado idioma, palavras e expressões de origem estrangeira freqüentemente sofrem modificações em relação aos seus usos originais, dentro dos novos contextos de utilização. O objetivo deste trabalho foi investigar as diferenças de usos apresentadas por anglicismos incorporados ao português do Brasil. A partir do levantamento de uma amostra de 15 anglicismos de uso bastante freqüente em língua portuguesa no Brasil, realizada através de busca em corpora lingüísticos de língua inglesa e portuguesa, comparamos os usos dos termos na língua original (inglês) e as acepções a eles atribuídas após o empréstimo. Nove itens apresentaram modificação importante de sentido, enquanto os seis restantes pareceram conservar seus usos originais praticamente inalterados. O percentual de variação encontrado corresponde a 60% da amostra, o que parece apontar para um comportamento bastante heterogêneo dos anglicismos incorporados ao português do Brasil no que tange aos seus usos, embora aponte para uma possível tendência à alteração nos sentidos originais após o empréstimo.*

Palavras-chave: *empréstimos lingüísticos/ estrangeirismos; anglicismos em língua portuguesa; lingüística de corpus.*

¹ Agradecemos às professoras Dra. Stella Tagnin (USP) e Dra. Paula Lenz (UECE) pela indispensável orientação durante a elaboração deste trabalho.

* UECE / CMLA.

** CAPES – UECE / CMLA.

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

Abstract: *Foreign language words and expressions, as used by native speakers of a certain language, frequently change their original usage in new contexts. This paper aims at investigating the differences in usage that occur when words are borrowed from English into Brazilian Portuguese. After collecting a sample of 15 such words, we compared the usage of the original terms and the modifications they undergo after the borrowing. The sample was a result of our search in English and Brazilian Portuguese corpora. Nine items showed important variations in their usage while the other six seemed to maintain their original usage. The variation found corresponds to 60% of the sample, which suggests a heterogeneous behavior of words incorporated into Brazilian Portuguese in what concerns the conservation or modification of their original use, though it may also point to alterations in the original use after the borrowing has occurred.*

Keywords: *linguistic borrowing; anglicisms in Brazilian Portuguese; corpus linguistics.*

Introdução

Recentemente, o termo *estrangeirismo*, provavelmente desconhecido de grande parte da população brasileira até então, ganhou notoriedade e alcançou posição de maior destaque dentro do vocabulário do dia-a-dia do cidadão comum. Isso se deu graças à discussão política gerada em torno do uso de estrangeirismos por falantes de português brasileiro deflagrada pelo já célebre projeto de Lei 1.676/1999 do deputado Aldo Rebelo (PCdoB/São Paulo), que pretendia dispor sobre “a promoção, a proteção, a defesa e o uso da língua portuguesa”. De acordo com o texto do referido projeto, a batalha pela promoção, defesa, etc., etc. de nossa língua teria como principal alvo justamente o combate aos estrangeirismos, vistos como elementos daninhos ao idioma nacional, capazes de descaracterizá-lo e empobrecê-lo.²

Definido como o “emprego, na língua de uma comunidade, de elementos oriundos de outras línguas” (Garcez & Zilles, 2001), o estrangeirismo surge sempre que diferentes comunidades lingüísticas entram em contato. Dentro de uma concepção estritamente científica e não-preconceituosa, o uso de estrangei-

² A respeito da discussão em torno do projeto de lei anti-estrangeirismos do deputado Aldo Rebelo, conferir os excelentes artigos reunidos por Carlos Alberto Faraco no livro *Estrangeirismos – Guerras em Torno da Língua*. (São Paulo: Parábola, 2001).

rismos é um fenômeno natural e constante na história das línguas, contribuindo para o constante processo de renovação que permite aos idiomas acompanhar a mudança social dos povos que os falam. No português do Brasil, uma grande influência estrangeira no domínio lexical, mensurada a partir da enorme quantidade de palavras e expressões incorporada ao uso cotidiano, é de origem inglesa. Os anglicismos, portanto, constituem um grupo significativo dentre os estrangeirismos atualmente empregados no Brasil, juntamente com palavras de outras e várias origens.

Ao serem utilizados pelos falantes de português, alguns anglicismos (bem como estrangeirismos de várias outras origens, empregados por falantes nativos dos mais diversos idiomas) sofrem alteração com relação ao seu uso, passando a ser empregados em contextos diferentes daqueles em que são empregados pelos falantes nativos da língua original. Essas diferenças de uso acontecem em graus diferentes de palavra para palavra (ou expressão). As diferenças entre a utilização de uma palavra ou expressão em seu país de origem e no país estrangeiro resultam, mais uma vez, de um processo natural, consequência direta do empréstimo de termos estrangeiros por uma comunidade lingüística, naturalmente diferenciada social, econômica, política e culturalmente da comunidade lingüística da qual o termo emprestado é nativo. Não se trata, portanto, de “corrupção”, “descaracterização” ou “uso errôneo” do estrangeirismo, mas tão-somente da adaptação da palavra estrangeira às peculiaridades de sua nova “pátria” comunicativa.

O objetivo da pesquisa aqui relatada foi investigar a variação nos usos de palavras de origem inglesa incorporadas ao português do Brasil (anglicismos, portanto). Para tanto, utilizamo-nos da busca em corpora lingüísticos para levantar os contextos de utilização dessas palavras em português e em sua língua original. Nossa opção pela busca em corpora baseou-se em nossa intenção de que os exemplos levantados, embora limitados em número, representem, da maneira mais fiel possível, o uso dos termos pesquisados nas duas línguas (inglesa e portuguesa). Considerando que os corpora escolhidos (conforme será discutido abaixo) contêm apenas instâncias autênticas de utilização dos idiomas por falantes nativos, ao contrário da maioria das obras lexicográficas de porte editadas nos dois idiomas, a utilização destes corpora como fontes únicas de pesquisa pareceu-nos mais condizente com nossos objetivos.

Levantamento dos dados

Tomamos como ponto de partida para nossa análise uma lista de 97 estrangeirismos de origem inglesa comumente utilizados no Brasil, compilados ao longo da leitura do texto da obra *Estrangeirismos – Guerras em Torno da*

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

Língua, organizada por Carlos Alberto Faraco (2001). Nos oito artigos que constituem a obra, os autores discutem questões lingüísticas – e também políticas e sociológicas – relacionadas ao uso de estrangeirismos no Brasil, citando como exemplos termos de diversas origens, a maioria oriunda da língua inglesa. Após a exclusão das palavras de outras origens que não a inglesa, restaram-nos os 97 itens da lista citada acima.³

Em seguida, submetemos as palavras e expressões da lista original a um levantamento de frequência de utilização em português do Brasil. Para tanto, adotamos como banco de textos o conjunto das edições de 1994 a 1998 da Folha de São Paulo, jornal de grande circulação nacional, constituído de grande variedade de cadernos e seções contemplando enorme número de temas. Através de busca informatizada, em que utilizamos o CD-Rom contendo as edições mencionadas, levantamos o número de artigos do jornal em que aparece cada uma das palavras ou expressões compiladas acima.⁴ Para os fins da pesquisa aqui relatada, foram consideradas as 15 palavras ou expressões presentes no maior número de artigos do jornal, conforme listado abaixo:

CLASSIF.	ANGLICISMO	OCORRÊNCIAS
1	e-mail	9774
2	marketing	8946
3	ranking	6896
4	shopping	6859
5	software	3897
6	site	3265
7	van	2876
8	big	1974
9	light	1893
10	fashion	1747
11	hall	1692
12	shopping center	1638
13	slogan	1400
14	boom	1066
15	mouse	1049

Tabela 1 – Anglicismos mais freqüentes conforme o banco de textos da Folha de São Paulo (em ordem decrescente de freqüência)

Exemplos de utilização dessas palavras foram então extraídos do mesmo banco de textos da Folha de São Paulo, seguindo-se o seguinte critério: dentre os

³ Ver Anexo I.

⁴ Ver Anexo I.

artigos mais recentes em que apareceram os itens pesquisados, foi retirada de cada artigo pelo menos uma frase contendo o item em questão, perfazendo um total de 40 exemplos para cada item pesquisado.⁵ As frases assim obtidas foram listadas de acordo com a palavra de busca que continham, gerando nosso pequeno corpus de trabalho em português. Segue abaixo um exemplo de resultado de busca no banco de textos da Folha, em que as frases selecionadas aparecem já organizadas e formatas, com a palavra de busca destacada em negrito:

shopping center
1. Também não disfarça no merchandising de achocolatados e de dentifrícios -que aparecem com destaque até em vitrine de shopping center .
2. Jessica, 3, sem-teto, durante passeio no shopping center Iguatemi, ontem na Folha.
3. Era um passeio de Natal de um grupo de cinco sem-teto pelo shopping center Iguatemi.
4. Consumidores no shopping center Paulista; movimento no comércio em São Paulo aumenta cerca de 40% desde o final de semana
5. Homens entre 45 e 65 anos, boa disposição e que amem crianças, para função de bom velhinho em shopping center .

Figura 1 – Exemplos de resultados de busca no banco de textos da Folha de São Paulo para a expressão *shopping center*.

A partir da lista dos 15 anglicismos mais freqüentes, realizamos a busca de exemplos de utilização dessas palavras em sua língua original, empregando para esse fim a versão *on-line* do corpus COBUILD⁶, que, por ser uma versão de demonstração, permite acesso a apenas 40 instâncias de cada palavra pesquisada. No entanto, em vez de frases completas, o COBUILD forneceu-nos os resultados em forma de *concordances*⁷ (ver exemplo abaixo), o que, no entanto, não nos pareceu prejudicar a análise que se seguiu.

⁵ A opção pelo número de 40 exemplos deveu-se à intenção de manter uma certa coerência com a busca no corpus de inglês que se seguiu, em que dispúnhamos apenas da versão demo do corpus COBUILD – que fornece apenas 40 instâncias de cada palavra buscada.

⁶ Disponível on-line no endereço: <http://titania.cobuild.collins.co.uk> durante a elaboração deste trabalho. Atualmente disponível no endereço: www.collins.co.uk/Corpus/Corpus Search.aspx.

⁷ O termo *concordance*, freqüentemente traduzido para o português como “concordância”, refere-se a “uma lista contendo uma palavra específica (chamada de palavra de

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

Query Results		
NOTE: no more than 40 lines will be displayed here, since a threshold has been implemented. If there were more than 40 instances found, a random selection will have been applied.		
Hoski had walked to a neighborhood theme visible in Horicon. The town area. For instance, we build that Health Systems and the Stanford proposed expansion of the Stanford	shopping center shopping center shopping center Shopping Center Shopping Center	and, at a J. C. Penney is called Marshland. The -- we had one in Compton [p] If you can help us and prevent development of

Figura 2 – Exemplos de resultados de busca no corpus COBUILD para a expressão *shopping center* (*concordances*)

Comparação dos resultados de busca

De posse das instâncias de utilização dos 15 anglicismos em inglês e em português, procedemos à compilação dos contextos em que cada um dos termos foi empregado nas duas línguas. Seguiu-se, então, a comparação desses contextos, na busca de possíveis discrepâncias que poderiam apontar para uma diferenciação entre o uso do termo por falantes nativos de sua língua de origem e o novo uso do termo como empréstimo na língua portuguesa.

A Tabela 2, abaixo, apresenta os 15 anglicismos aqui estudados divididos em dois grupos de acordo com a presença ou ausência de discrepâncias de uso consideradas significativas para os objetivos de nossa análise. Os números entre parênteses representam a posição na lista dos 15 itens mais frequentes apresentada acima na Tabela 1.

busca ou nóculo) juntamente com parte do texto ao seu redor (co-texto)” (Tony Berber Sardinha, em manuscrito não publicado). O formato mais popular de *concordance* é a chamada KWIC Concordance (KeyWord in Context, ou “palavra-chave em contexto”) (Baker, 1995). O programa que realiza as *concordances* (ou *concordancer*) na versão on-line do corpus COBUILD utiliza este formato de apresentação dos resultados de busca, ou seja, apresenta uma lista das ocorrências da palavra-chave especificada, sendo esta colocada no meio de uma linha de contexto para cada ocorrência. Não existe a obrigatoriedade de fornecer a sentença completa em que a palavra de busca apareceu, conforme mostra o exemplo (Figura 2).

GRUPO 1 Anglicismos que apresentaram variação significativa de usos	GRUPO 2 Anglicismos que mantiveram praticamente inalterados seus usos originais
e-mail (1) ranking (3) shopping (4) site (6) big (8) light (9) fashion (10) hall (11) boom (14)	marketing (2) software (5) van (7) shopping center (12) slogan (13) mouse (15)

Tabela 2 – Anglicismos mais freqüentes classificados quanto à ocorrência de variação semântica significativa.

Como podemos observar, dentre os 15 termos pesquisados, seis não apresentaram variação de usos significativa. Sobre estes, não há muito o que comentar. Entretanto, nove dos termos apresentaram variações significativas de usos, merecedoras de uma análise mais detalhada, a que procederemos a seguir.

É importante ressaltar, no entanto, que os resultados apresentados referem-se exclusivamente aos dados recolhidos para este estudo, refletindo, assim, uma limitação do presente trabalho. Evidentemente uma pesquisa mais aprofundada, que não apenas utilize como fontes um maior número de corpora em ambas as línguas e obras lexicográficas nos dois idiomas, mas que também focalize um espectro maior de estrangeirismos, talvez até de outras origens que não só a inglesa, poderá oferecer resultados bastante diferentes dos apresentados aqui. No entanto, embora restrito, nosso material de pesquisa foi suficiente para o levantamento de alguns dados interessantes sobre o tema, conforme relataremos a seguir.

Anglicismos do Grupo 1

Como já foi dito acima, fazem parte deste grupo os nove termos estudados que apresentaram diferença significativa entre seus usos no país de origem e no português do Brasil. Analisaremos agora cada um deles separadamente, procurando esclarecer as diferenças encontradas.

- **e-mail**

As instâncias recolhidas do corpus COBUILD apresentam o termo utilizado como substantivo, significando “correio eletrônico”, “endereço de correio eletrônico”, ou ainda “mensagem enviada por correio eletrônico”:

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

***E-mail** is having a much bigger effect on...*

*He said: I'd prefer my **e-mail** to be out there on the network...*

*Faxes can also be converted to **e-mail** and picked up from the Internet.*

No entanto, observamos a utilização da palavra também como verbo, em inglês:

***E-mail** me at [e-mail]⁸ and let me know that you are...*

O termo, quando utilizado em português do Brasil, de acordo com os dados recolhidos, também ocorreu como substantivo, mas não como verbo. Por esta razão decidimos manter este item dentro do grupo 1 por seu emprego não ser exatamente idêntico ao da língua de origem, ou seja, manifesta-se no português do Brasil através de apenas um de seus usos originais. O mesmo ocorre com outros itens aqui estudados, como é o caso de *ranking, site, big* e *boom*.

*... os alunos que estudam à distância aprendem mais, porque eles conversam muito mais entre si sobre os conteúdos do curso por **e-mail**...*

*Cartas para esta coluna devem ser enviadas para o Folhainvest, alameda Barão de Limeira (...), ou pelo **e-mail**:investimentos@uol.com.br.*

*Continuo recebendo **e-mail** comercial que me diz como posso lucrar distribuindo **e-mail** comercial.*

- **ranking**

De forma muito semelhante, a palavra *ranking* também conserva seu uso praticamente inalterado no português do Brasil na função de substantivo, significando “escala de importância dos membros de um determinado grupo”⁹:

*... is the structured **ranking** of entire groups of people...*

*Luiz Antonio dos Santos, 34, o primeiro do **ranking** brasileiro de maratona, disse que vai correr com cautela...*

⁸ Talvez por razões de proteção de privacidade, o endereço eletrônico foi omitido deste exemplo obtido através do corpus COBUILD.

⁹ “To have, or give, a place in a group, according to importance” (Password – English dictionary for speakers of Portuguese, 1996)

Crop, 10, 2004

Novamente, porém, na língua inglesa ela pode exercer a função de verbo, constituindo a forma do particípio presente do verbo *rank*, que significa “definir a posição de alguém ou algo em uma escala de importância”:

... *the top national Liberal Arts school, **ranking** first in academic reputation...*

Esta distinção entre o uso original do termo e seu emprego como empréstimo torna-se, no caso de *ranking*, ainda mais significativa do que no caso de *e-mail*, em virtude da diferença na frequência com que esses dois termos são utilizados como verbo na língua inglesa, de acordo com os dados recolhidos. Enquanto o último foi empregado como verbo apenas uma vez em nossa amostra, o primeiro foi utilizado como verbo na maior parte das instâncias recolhidas.

- **shopping**

O termo *shopping* foi, dentro do conjunto dos itens desta pesquisa, a palavra que mais sofreu modificação na forma de emprego, ao ser adotada pelo português do Brasil. Todas as amostras recolhidas de seu uso em português apresentam este termo como um substantivo denotando “centro de compras”, evidente redução da expressão *shopping center* (ou *shopping mall*):

*Há poucas semanas, num belo **shopping** numa avenida de Havana, um jipão de brinquedo com motor elétrico era vendido a US\$ 400.*

*O **shopping** Tatuapé está registrando um movimento 30% maior do que o registrado no ano passado.*

No entanto, conforme visto anteriormente, *shopping center* é um dos seis estrangeirismos selecionados para este estudo que apresentaram pouca ou nenhuma variação no uso pós-empréstimo (ver Tabela 2). Pode-se concluir, portanto, que o uso da forma *shopping* no português do Brasil é diferente do uso da forma *shopping center* como empregada em língua inglesa corrente, já que, de acordo com as amostras recolhidas, *shopping* não é empregada como sinônimo de *shopping center* ou *shopping mall*. Quando utilizada como substantivo, *shopping* adota o sentido de “ação de fazer compras”, aparecendo inclusive em posição atributiva em um sem-número de expressões como *shopping bag*, *shopping list*, *shopping cart* e, é claro, *shopping center*. Empregada como verbo, corresponde ao particípio presente de *shop*, que significa literalmente “comprar, ir às compras”:

*For the best **shopping** this side of London, look...*

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

*Buyers are still **shopping** around to squeeze an extra...*

- **site**

De acordo com as instâncias recolhidas no corpus COBUILD, a palavra *site* é empregada, em inglês corrente, como um substantivo significando simplesmente “lugar”, “região”, “terreno”; ou, mais especificamente, “lugar onde fica localizado algo”. Nesta segunda acepção, os elementos que se podem situar neste lugar determinado podem ter naturezas variadas: edifícios, monumentos, marcos históricos ou arqueológicos, atrações turísticas ou ecológicas, entre outras:

*As we approached the **site** there was an anguished cry.*

*Knight Frank is selling a 35 acre **site** in the Scottish borders...*

*An important breeding **site** for seabirds...*

*... Cleveland, the **site** of the latest power line...*

*... their positions near Harbel, the **site** of a massive rubber plantation...*

Mais modernamente, incluiu-se entre os possíveis elementos a ocupar o “lugar” mencionado na segunda definição as páginas da rede mundial de informação, surgindo daí a expressão *Web site*, ou “lugar na WWW (World Wide Web) que abriga a página virtual de um determinado indivíduo, empresa ou entidade”. Essa expressão é muitas vezes empregada de forma reduzida, aparecendo apenas como *site*:

*Survival's **site** has news on hotspots around...*

Ao ser tomada em empréstimo pelo português do Brasil, *site* conservou apenas o uso desta última acepção, sendo encontrada em todos os exemplos recolhidos como sinônimo ou forma reduzida de *Web site*, significando, portanto, “lugar na WWW que abriga a página de um determinado indivíduo, empresa ou entidade”:

*Este é o endereço de um **site** sobre Archie Moore, ex-campeão mundial dos meio pesados...*

*O **site** Bibliomania, em www.bibliomania.com, tem mais de 40 novelas, como “Little Women”, e livros clássicos para transferir.*

- **big**

A variação no emprego da palavra *big* no português do Brasil aponta em direção a uma modificação na quantidade de acepções que o termo toma em sua

Crop, 10, 2004

língua de origem, conforme demonstrado pelos dados coletados. Assim, em língua inglesa moderna, *big* é utilizado como adjetivo em uma variada gama de contextos, sendo o mais comum deles o que se refere a tamanho, largura ou intensidade. *Big*, nesses casos, significa “grande”, “largo”, ou “intenso”:

*It was a **big** part of her life now...*

*The road isn't **big** enough to cope. Our cars get blocked...*

*For me this was a **big** relief.*

Outra possibilidade de utilização de *big* em inglês, embora menos freqüente nas amostras recolhidas, é como sinônimo de *good*, *nice* ou *great*, no sentido de uma avaliação positiva de uma determinada coisa, instituição, evento etc.

*No-one ever gets a **big** idea through a rational, conscious...*

Além disso, em alguns exemplos, *big* foi empregada como equivalente aos adjetivos *main* ou *major*, correspondendo a uma noção de “o maior/um dos maiores” (em tamanho), ou “o melhor/um dos melhores” ou “o mais importante/um dos mais importantes” (em valor ou qualidade)”, como no exemplo abaixo:

*One **big** producer, Tim Finney of Eastbrook...*

Em apenas uma das instâncias recolhidas, *big* foi utilizada como sinônimo para *knowledgeable*, conotando um “indivíduo que possui grande conhecimento em uma determinada área”:

*... was named Alejandro. He was **big** on latin culture.*

Em língua portuguesa, *big* apresentou diferenças em seus usos, sendo empregada no português do Brasil com menos sentidos que em sua língua original. De acordo com os dados coletados, essa limitação acabou por preterir apenas os sentidos menos freqüentemente atribuídos ao termo na língua de origem. Assim, os exemplos recolhidos em português apresentam *big* utilizada como sinônimo para “grande”, “largo”, “intenso”, “muito bom/ótimo” e outras expressões de valor semântico similar. É também empregada nesses sentidos na formação de nomes próprios, juntamente com outros termos estrangeiros ou mesmo lado a lado com palavras vernáculas, como nos exemplos abaixo:

*Carmo Sodré e Jovelino Mineiro abocanharam este fim-de-semana um **big** prêmio na Argentina.*

*Se não vender, se for um **big** fiasco, para mim é o mesmo de sempre.*

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

O **big** empreendimento inclui centro multimídia, 20 estúdios, 22 mil metros quadrados de palcos de som...

Novo supermercado **Big** Cândia, que será inaugurado em SP...

- **light**

O sentido mais divulgado do adjetivo *light* em língua portuguesa do Brasil é aquele que é geralmente confundido com o também estrangeiro *diet*¹⁰, ou seja, “de baixa caloria” ou “apropriado para dietas de emagrecimento”. No entanto, na amostra coletada para esta pesquisa, o termo é também – e mais freqüentemente – empregado com a conotação de “sutil”, “suave”, “discreto”, “pouco agressivo”, “aprofundado ou detalhado”, e até mesmo “politicamente correto”:

*Os planos da Pepsi para 99 incluem o lançamento da sua versão **light***

*Qualquer humor negro perto do discurso do Rei-Eleito FHC vira **light!***

*Maluf pretende ser **light**, sem comprar briga.*

*O PT adota um discurso “**light**”, jovem e de tom desenvolvimentista...*

Segundo ele, a crítica da peça é “**light** e próxima do público”.

Analisando, no entanto, as amostras de *light* utilizadas por falantes de sua língua de origem, essas diferem das amostras do português do Brasil. Por exemplo, não há, entre os exemplos coletados do inglês, nenhum em que *light* venha denotar “de baixa caloria”, quando relacionado a alimentos. Na função de adjetivo, *light* é associado a cores e utilizado com o sentido de “leve” em valorações de peso ou, menos freqüentemente, com alguns – mas não com todos – os sentidos também observados em português (“sutil”, “discreto”, “suave”):

***Light** pink, double blooms produced in...*

*... like a swarm of little hands, **light** but thick, touching even her face...*

¹⁰ O termo *diet* também fazia parte de nossa lista original de anglicismos freqüentes em língua portuguesa do Brasil. Não foi, porém, tratado com detalhe nesta pesquisa por não apresentar um número de ocorrências suficiente para incluí-lo na lista dos 15 estrangeirismos selecionados para este estudo (ver Anexo I).

Crop, 10, 2004

Ao contrário, porém, do que foi observado nas amostras em português, nos exemplos em inglês a esmagadora maioria das instâncias de utilização do termo *light* apresentaram-no na função de substantivo, significando “luz”:

*Hold it up to the **light** and check it carefully...*

*... with excellent natural **light** and interesting views.*

Não se observou, nas amostras em português, nenhum exemplo deste emprego tão freqüente na língua de origem desta palavra.

- **fashion**

Os exemplos extraídos do corpus COBUILD apresentam *fashion* como substantivo, com duas acepções principais. Na primeira delas, *fashion* surge como sinônimo de *way* ou *style*, significando “modo, maneira, estilo de agir”:

*... our clinical staff acted in a proper **fashion**.*

*... repeat lessons in parrot **fashion**, learning the Russian language...*

A segunda acepção é a que mais profundamente transportou-se para o português (como veremos abaixo), na qual *fashion* corresponde a “moda”, tanto em sentido genérico, referindo-se ao que está em voga no momento em termos de roupas, sapatos, acessórios, entre outros, como, em sentido estrito, à área de estudo correspondente ao “Estilismo”, em português:

*Formerly a **fashion** editor on one of America's...*

*... a body blow to our native **fashion** industry.*

*Of all the **fashion** graduates, most will join the...*

Como já foi mencionado acima, o emprego de *fashion* como sinônimo de “moda” é freqüente entre as amostras coletadas em português brasileiro, na maioria das vezes entrando na composição de uma grande variedade de nomes próprios, constituindo, dentre os exemplos recolhidos, a única acepção do termo em função de substantivo, podendo compor o significado de “feira de moda”:

*Mais uma baixa no Morumbi **Fashion**, depois da Viva Vida.*

*Tudo para o “GNT **Fashion**” comercial.*

No entanto, ao contrário do que se observou nos exemplos em inglês, em português o termo é mais freqüentemente empregado como adjetivo, geralmente significando “na moda” (correspondendo, portanto, a uma redução do

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

adjetivo inglês *fashionable*, que tem precisamente o mesmo sentido). Isso acontece mesmo quando o termo refere-se a elementos outros que roupas, sapatos e acessórios (lugares e pessoas, por exemplo), caso em que os equivalentes mais próximos em português seriam “famoso”, “célebre” ou “popular”. Variações dessa primeira acepção apontam para a utilização de *fashion* como adjetivo equivalente a “moderno, atualizado”, mesmo quando o contexto de utilização nada tem a ver com o mundo da moda:

... trazendo as cores, o fluo e o branco à negritude **fashion** da modernidade.

O debochado jornalzinho de Ronaldo Fraga foi a sensação da “cobertura” da *Semana da Moda*, com “reportagens” sobre o universo do estilista e sátiras ao mundinho **fashion**.

O **fashion** é a meta, numa atitude moderna, livre e desencanada.

Ele desembarcou – sozinho da silva – em St. Barthélemy, ilha mais **fashion** do Caribe.

Betty Lago anda mais **fashion** do que nunca.

O filme filipino “Casamento de Conveniência” (...) mostra uma visão do universo gay que passa bem longe de qualquer estética **fashion** ou moderninha.

- **hall**

A partir dos exemplos extraídos do corpus COBUILD, a palavra *hall* é empregada, em sua língua original, como um substantivo dotado de vários sentidos relacionados, aos quais os equivalentes em português podem, dependendo do caso, ser “salão”, “casa de espetáculos ou concertos”, “mansão ou casa senhorial”, “edifício público”, “corredor”, “pavilhão”, “saguão” ou “vestíbulo”:

... a great **hall** with minstrel's gallery for dining...

... in the **hall** was a playpen with a baby...

... the chance to see the great Central **Hall** – until recently home to the gallery...

...Glasgow Royal Concert **Hall**, Sauchiehall Street.

... she came running down the **hall** in her new outfit...

The main **hall** has been completed and looks really...

Além das ocorrências acima, em que aparece isoladamente como substantivo, *ball* faz parte da expressão *ball of fame*, encontrada várias vezes em nossa amostra e que, como se verá mais à frente, também é utilizada como empréstimo em português. Esta expressão, equivalente a algo como “galeria ou corredor da fama”, é utilizada em inglês para designar um “grupo de indivíduos mais bem sucedidos em uma determinada atividade”:

... *research of the Pro Football **Hall of Fame** in Canton, Ohio.*

... *known until 1975 as the Lacrosse **Hall of Fame** Federation...*

Nos exemplos levantados em português, *ball* é também empregada em alguns dos sentidos que lhe são atribuídos na língua original, mais especificamente aqueles que equivalem a “corredor”, “saguão” ou “vestíbulo”. Os demais significados não foram detectados nas amostras recolhidas, com exceção daqueles que se referem à expressão *ball of fame*. No entanto, nesse caso, a expressão geralmente aparece na forma híbrida “*ball* da fama”:

*Estacionei o carro, entrei no prédio, fui até o **ball** e vi que três pessoas desceram do elevador: um rapaz e duas senhoras.*

*No **ball** ao lado do estúdio, o presidente e Lewinsky se beijaram.*

*... eu tiro minha tábua de passar do armário do **ball**, monto-a a uns cinco passos do aparelho de TV (...) e passo tudo.*

*O desfile da Chanel no **ball** da Ópera da Bastilha, também perdeu em élan.*

*É o atual dono do recorde de nocautes (141), o que lhe valeu um lugar no **Hall** da Fama.*

Dentre os nove estrangeirismos que fazem parte deste primeiro grupo, *ball* foi o termo que apresentou o menor grau de diferença entre os usos quando utilizado como empréstimo no português do Brasil.

- **boom**

Nas duas acepções tomadas pelo termo *boom* nas amostras recolhidas em inglês, este anglicismo é empregado como substantivo. Em alguns exemplos, apresentou-se como palavra onomatopaica, denotando “um som alto e surdo, como o de um grande tambor ou arma de fogo”, a que um bom equivalente em português seria talvez a palavra “estrondo”. No entanto, na grande maioria das instâncias, o substantivo *boom* foi utilizado em seu sentido figurativo, significando um “súbito crescimento ou progresso”:

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

*... crawling from crib to crib. **Boom**. Then he ran into Firebug. Boom.*

*Canada had a baby **boom**, a high incidence of immigration...*

*... the company is enjoying a business **boom**.*

*We had an unsustainable **boom** in housing, said John Maples...*

Em relação ao original inglês, identificamos que apenas o segundo dos dois sentidos originais discutidos acima se mantém após a incorporação da palavra ao vocabulário do brasileiro. Em todas as instâncias coletadas em português, portanto, *boom* foi empregada com o significado de “súbito crescimento ou progresso”:

*Sem o embargo americano, poderia haver um “**boom**” do turismo, mas sempre sob controle estatal.*

*O **boom** de inaugurações de hipermercados em todo o país vem preocupando os comerciantes de pequeno e médio portes.*

*Popular nos EUA, o MBA tem experimentado um “**boom**” no Brasil nos últimos anos.*

Conclusão

Em termos percentuais, a variação de usos encontrada corresponde a 60% da amostra.¹¹ Não é possível determinar um padrão absolutamente regular que explique sistematicamente as variações ou conservações de usos dos anglicismos quando incorporados em nossa língua. Entretanto, os resultados parecem indicar uma tendência para alterações de uso após os empréstimos.

Reconhecemos as limitações de um estudo deste porte. Diferentes focos poderiam ser considerados em termos de critérios de análise, tais como variação de classes gramaticais ou variação de número (como é o caso de “e-mail”, que se pluraliza no português do Brasil enquanto isso não parece ocorrer no inglês). Esperamos, entretanto, que este modesto estudo possa contribuir para o desenvolvimento da investigação de estrangeirismos na nossa língua.

¹¹ Entendemos, porém, que o limitado número de instâncias recolhidas impossibilita um tratamento estatístico mais aprofundado. Ao explicitar este valor, intentamos apenas facilitar a visualização geral dos resultados alcançados.

Crop, 10, 2004

Referências Bibliográficas

- BAKER, Mona.(1995) *Corpora in translation studies: an overview and some suggestions for future research*. Target 7, John Benjamins, p. 223-243.
- BERBER-SARDINHA, Antonio P.Berber. *Handout – Wordsmith Tools*. Manuscrito não publicado.
- COLLINS PUBLISHERS. *COBUILD Corpus of the English language – on-line sampler*. Disponível no endereço: <http://titania.cobuild.collins.co.uk>
- FARACO, Carlos Alberto (Org.). (2001) *Estrangeirismos – guerras em torno da língua*. São Paulo: Parábola Editorial.
- FOLHA DE SÃO PAULO. (1999) *CD-ROM Folha edição 99 – CD 1 – Texto Integral*. São Paulo: Publifolha.
- GARCEZ, Pedro M. & ZILLES, Ana Maria S. (2001) *Estrangeirismos – desejos e ameaças*. In: FARACO, Carlos Alberto (Org.). *Estrangeirismos – guerras em torno da língua*. São Paulo: Parábola Editorial. p. 15-36.
- KERNERMAN SEMI-BILINGUAL DICTIONARIES. (1996) *Password – English dictionary for speakers of portuguese*. São Paulo: Martins Fontes.

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos.*

ANEXO I

ESTRANGEIRISMOS DE ORIGEM INGLESA CITADOS EM *FARACO*, 2001

(acompanhados de sua ocorrência dentro do Corpus da Folha de São Paulo de 1994 a 1998)

	ANGLICISMO	OCOR		ANGLICISMO	OCOR
1	home banking	106	50	pet shop	94
2	coffee break	6	51	factoring	351
3	shopping (4)	6859	52	leasing	648
4	shopping center (12)	1638	53	play-off	25
5	home page	803	54	delivery	156
6	pedigree	91	55	tie-break	313
7	software (5)	3897	56	top-spin	12
8	download	142	57	match point	23
9	internet banking	11	58	set game	0
10	card	328	59	dumping	354
11	happy hour	217	60	anti-doping	11
12	on sale	2	61	baby-doll	29
13	sale	75	62	big-bang	204
14	% off (*)	?	63	black-tie	217
15	out door	403	64	blazer	409
16	royalty	40	65	bodyboard	2
17	walkman	175	66	boom (14)	1066
18	cheeseburger	15	67	brain-trust	1
19	marketing (2)	8946	68	boomerang	10
20	budget	54	69	briefing	38
21	nonsense	146	70	bulldog	27
22	self-service	203	71	by-pass	5
23	van (7)	2876	72	site (6)	3265
24	feedback	55	73	topless	120
25	printar	3	74	ranking (3)	6896
26	drive	895	75	trekking	156

27	delete	29	76	kummel	3
28	insert	4	77	steeple-chase	1
29	reset	19	78	holding	994
30	office-boy	277	79	off-road	94
31	baby-sitter	29	80	station wagon	34
32	hall (11)	1692	81	approach	52
33	e-mail (1)	9774	82	lunch	35
34	mouse (15)	1049	83	ferryboat	20
35	ok	887	84	hot dog	85
36	halloween	140	85	deletar	29
37	smoking	213	86	bidar	0
38	know-how	496	87	yogurt	24
39	slogan (13)	1400	88	light (9)	1893
40	brunch	184	89	diet	319
41	fashion (10)	1747	90	clipping	40
42	catering	33	91	estartar/startar	0/2
43	preview	119	92	atachar	2
44	big (8)	1974	93	recall	125
45	weekend	371	94	franchise	100
46	serial killer	254	95	database	46
47	gadget	24	96	newsletter	50
48	songbook	127	97	personal banking	3
49	personal trainer	78			

OBS:

1. A ocorrência é dada em termos do número de artigos do jornal (entre 1994 a 1998) em que consta a palavra ou expressão pesquisada (a palavra pesquisada, no entanto, aparece freqüentemente mais de uma vez no mesmo artigo; esse dado, porém, não foi levado em consideração nesse momento da pesquisa).

SILVA, Diana Costa Fortier; COSTA, Renato Barros da. *Anglicismos no Português do Brasil: uma Perspectiva de Análise de Usos Baseada em Corpora Eletrônicos*.

2. Os itens em **negrito** representam os 15 estrangeirismos mais frequentes (maior número de ocorrências). O número entre parênteses refere-se à ordem relativa do item em relação aos demais termos da lista (compilados em ordem decrescente).
3. Não foi possível localizar nenhuma ocorrência da expressão *% off* na base de textos da Folha, nem com recurso ao acréscimo de um número antes do sinal de porcentagem (*25% off*). O fato deve-se provavelmente à incapacidade do software de busca de interpretar caracteres numéricos e outros não-alfabéticos. Uma alternativa para corrigir essa ausência de ocorrências poderia ter sido a busca pela preposição *off* individualmente; no entanto, para mantermos coerentes com a metodologia empregada para o levantamento dos demais itens da lista acima, resolvemos não optar por esse procedimento.

Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus

*Benivaldo José de Araújo Júnior**

Resumo: *Um exame dos livros didáticos e manuais de espanhol como língua estrangeira nos mostra a pouca relevância dada às estruturas passivas nessas publicações, quando não sua total ausência. O mais freqüente é que não se fale de estruturas passivas, e sim de voz passiva, adotando-se a definição, classificação e usos presentes nas gramáticas prescritivas. Na contramão dessa abordagem simplista, propomos uma análise por meio da lingüística do corpus, partindo da hipótese de que os dados – amostras de uso real da língua, em grande quantidade – podem revelar outros usos e particularidades das estruturas passivas que nos auxiliem a compreender melhor o fenômeno.*

Palavras-chave: *estruturas passivas; espanhol; livros didáticos.*

Abstract: *It one takes a good look at books used for teaching Spanish as a foreign language, one will notice the little or no importance given to passive structures. What is usually seen is a presentation of “passive voice” not of “passive structures” by adopting the definition, classification, and usage as they appear in prescriptive grammars. As an alternative to such a simplistic approach, we propose a corpus analysis of the matter. Our hypothesis is that the data – a large amount of samples from Spanish native speaker productions – are able to reveal other uses and particularities of passive structures which might give us a better understanding of the subject.*

Keywords: *passive structures; Spanish; textbooks.*

* Mestrando do Programa de Língua Espanhola e Literaturas Espanholas e Hispano-Americanas da Universidade de São Paulo.

1. O que dizem os livros didáticos e manuais de E/LE¹

Por questões de espaço e concisão, nosso propósito aqui não é realizar uma análise exaustiva de como o tema das passivas é tratado nos livros didáticos de E/LE. Portanto, selecionamos apenas alguns manuais² dentre os vários disponíveis no mercado brasileiro, e o critério que adotamos foi a boa aceitação desses materiais nas nossas escolas. Disponibilizamos a seguir uma síntese³ do que encontramos na análise:

1. Em geral não se fala de estruturas passivas, e sim de voz passiva, definindo-a como a *forma verbal* que indica que a pessoa/sujeito *recebe* a ação verbal, sendo portanto seu *paciente*. Dessa forma, a voz passiva surge em oposição à voz ativa, na qual a pessoa/sujeito *pratica* a ação verbal, convertendo-se em seu *agente*.
2. A voz passiva é classificada em dois tipos: passiva sintática (*frase verbal pasiva* ou *pasiva con ser*, formada com o verbo auxiliar *ser* + particípio) e passiva pronominal (*pasiva refleja* ou *pasiva con se*, com o pronome oblíquo *se* + verbo na 3ª pessoa).
3. A voz passiva se constrói com verbos transitivos diretos.
4. Usa-se a voz passiva nos seguintes casos: a) quando se quer destacar o paciente; b) quando o agente é desconhecido ou não se quer mencioná-lo; c) quando o agente é conhecido e não é necessário explicitá-lo. A passiva com *se* é sempre impessoal, ou seja, não explicita o agente.
5. A passiva com *se* é a mais utilizada na oralidade, enquanto a passiva sintática é típica da linguagem escrita, presente sobretudo nos jornais, folhetos, relatos históricos, guias turísticos etc.

Nesse tipo de abordagem, fica patente que a ênfase que se dá ao tema recai sobre o aspecto formal, ou seja, é esse o critério por excelência (que acaba se convertendo em fórmula) quando se pretende ensinar a reconhecer uma es-

¹ Espanhol como Língua Estrangeira.

² Planet@, de Cerrolaza et alii: *Libro del alumno*, vol. 1 (1998), 2 (1999), 3 (2000) e 4 (2000); *Libro de referencia gramatical*, vol. 1 (1998), 2 (1999), 3 (2000) e 4 (2000); *Uso de la gramática española*, de Castro, vol. 1 (1996), 2 (1997) e 3 (1997); *Claves del Español*, de Domínguez e Bazo (1994) e *Gramática básica del español*, de Sánchez e Sarmiento (2001)

³ Ressaltamos que o tratamento do tema varia muito entre um livro e outro. Por exemplo, Sánchez e Sarmiento (2001) dedicam dois comentários brevíssimos sobre as passivas, enquanto Domínguez e Bazo (1994) abordam a questão de forma mais distendida e com mais exemplos.

trutura passiva. O critério semântico (fundamentado nos papéis temáticos *agente* e *paciente*), de suma importância na questão, passa, nitidamente, a um segundo plano, assim como o caráter passivo intrínseco às formas verbais. Não se problematiza o fato de existirem outros possíveis papéis temáticos além dos de *agente* e *paciente* (na frase *la policía es temida por los manifestantes*, o constituinte *los manifestantes* não é um **agente**, como seria esperado na “fórmula” da passiva sintática, e sim um **afetado** pelo processo), nem a ocorrência de estruturas ativas com sentido passivo (*al ladrón lo detuvieron*, que tem clara correspondência com a passiva *el ladrón fue detenido*) e muito menos a impossibilidade de se construir a passiva com verbos transitivos diretos⁴ em determinadas situações (caso de: *los árboles tienen hojas*, **hojas son tenidas por los árboles*). Aprender as *estructuras pasivas*, portanto, transforma-se em aprender a *voz pasiva* tal como figura nos livros didáticos e em boa parte das gramáticas (sobretudo aquelas cujo enfoque é prescritivo): basta dominar as poucas regras apresentadas e praticá-las por meio de exercícios repetitivos, que consistem em converter frases ativas em passivas e vice-versa. Destaca-se o processo, a transformação, os artifícios que se usam para passar de uma forma a outra, como se fossem equivalentes em sentido e o falante optasse indiferentemente por uma delas. Esse modo de proceder corresponde a um imaginário de língua, de gramática, de ensino, que se reproduz na contramão das ditas e apregoadas inovações – o que por vezes as converte em pura maquiagem nos materiais didáticos e na fala do docente. Para concluir nosso raciocínio, ressaltamos que a tarefa mecânica de transformar ativas em passivas (e o contrário) acaba por adquirir o aspecto de um jogo ou brincadeira, ao fim da qual muitos estudantes se perguntam que finalidade poderia ter, dado o caráter arbitrário e vazio de sentido da tarefa. É claro, tal aspecto lúdico e automatizante não seria problemático caso se passasse desse momento a uma reflexão sobre as formas da língua, como funcionam em diferentes contextos e que sentidos e valores comportam nesses contextos. No caso específico das estruturas passivas, pensar, por exemplo, nos efeitos obtidos ao se enfatizar um sujeito paciente ou ao se omitir um agente.

⁴ É preciso esclarecer que, em espanhol, com relação à transitividade, os verbos são classificados apenas em **transitivos** e **intransitivos**, conforme a admissão ou não de um objeto direto. Nas palavras de Alarcos Llorach (1994: 280-81): “La posibilidad o imposibilidad de que el verbo admita objeto directo ha sido el criterio de clasificación de los verbos en *transitivos* e *intransitivos*. Cuando la actividad denotada por la raíz verbal requiere la especificación aportada por el sustantivo que funciona como objeto directo, se considera el verbo transitivo; en caso contrario, el verbo es intransitivo.”

ARAÚJO JÚNIOR, Benivaldo José de. *Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

2. Nossa proposta de estudo

Pensando no que foi dito na seção anterior sobre a abordagem das passivas nos manuais, propomos um caminho alternativo de tratamento da questão, por meio da lingüística de corpus: em vez de restringir as passivas a duas formas (“fórmulas”) – a sintática (com *ser*) e a pronominal (com *se*) –, identificar outras possíveis estruturas de conteúdo passivo nos dados obtidos em um corpus de referência da língua espanhola. A idéia é partir das amostras de uso da língua em textos autênticos e nelas buscar padrões lexicais e sintáticos relacionados ao nosso objeto de análise (estruturas passivas). O corpus de referência servirá para testar, validar e exemplificar hipóteses, além de servir como fonte de evidência quantitativa. Igualmente, a observação de padrões recorrentes de uso nos dados do corpus pode levar a que se formulem hipóteses sobre o objeto em estudo, demandando, assim, uma análise qualitativa. Um estudo dessa natureza poderia servir, posteriormente, para a elaboração de unidades e materiais sobre esse tema feitos com uma perspectiva mais ampla.

2.1 O corpus de referência

Utilizaremos o CREA – *Corpus de Referencia del Español Actual*, projetado e mantido pela *Real Academia Española* (RAE), que conta com aproximadamente 130 milhões de palavras⁵. Na composição do corpus entraram textos completos – tanto escritos (90%) quanto orais (10%) –, coletados a partir de 1975. Também no intuito de garantir um quadro representativo da língua espanhola no mundo, o CREA reserva 50% dos textos para a variante peninsular e dedica os outros 50% para as demais variantes (especialmente as americanas).

No intuito de oferecer a maior flexibilidade possível na obtenção de dados, o CREA está estruturado em diferentes módulos, o que torna possível que as consultas se refiram à totalidade dos textos ou unicamente àqueles que satisfaçam determinados critérios selecionados pelo usuário, por exemplo: CRONOLÓGICOS (trabalhar com apenas uma faixa temporal, por exemplo, de 1980 a 1990); GEOGRÁFICOS (trabalhar com uma só variante); MEIO (optar por textos publicados em jornais); TEMÁTICOS⁶ (escolher entre textos de ficção, científicos, artísticos

⁵ Este corpus está disponível para consulta gratuita no site <http://corpus.rae.es/creanet.html>

⁶ Os textos do corpus estão distribuídos em 7 hipercampos, que ainda contêm subdivisões. Por exemplo, o hipercampo FICCIÓN (22,5% do corpus) está subdividido em *novela, relatos e teatro*.

etc.). Apesar dessas vantagens, o sistema apresenta limitações referentes à quantidade de documentos recuperáveis – que está limitada a 2000 –, assim como à quantidade de exemplos, restrita aos primeiros 1000 de cada consulta⁷.

2.2 O caso a estudar

Pretendemos analisar neste trabalho a incidência de estruturas passivas com o verbo *pensar*. A razão de escolha desse verbo deve-se ao fato de o mesmo admitir tanto construções transitivas quanto intransitivas. Ademais, o verbo em questão parece ser bastante produtivo em construções intransitivas (*El hombre es un animal que piensa; El director nunca piensa sobre los problemas; Me ha dicho que piensa en mí todos los días*) e em outras que usualmente não admitem passivização (*Pienso que deberías volver a tu país; Pienso marcharme ahora mismo*), o que nos levaria a supor, em princípio, uma baixíssima incidência de passivas. É isso que tencionamos verificar a partir dos dados do CREA, além de observar se as passivas encontradas se limitam à classificação apontada nos manuais. Por questões de brevidade, faremos as seguintes delimitações:

1. Será utilizada apenas a variante peninsular; porém, não haverá restrições cronológicas, de meio ou de tema;
2. Restringiremos a pesquisa e análise às passivas com participípio. Fica excluída, portanto, a passiva pronominal (*pasiva refleja*).

Para familiarizar o leitor quanto aos usos mais frequentes de *pensar* em espanhol, transcrevemos acepções e exemplos oferecidos para o verbo no dicionário *María Moliner: Diccionario de uso del Español* (1998: 629-30). Apenas as acepções foram traduzidas para o português:

1. Formar e relacionar idéias; considerar; examinar; reflexionar; refletir: (intr.) *El oficio del filósofo es pensar*; (intr.) *Estaba pensando en ti en este momento. – ¿Y qué pensabas?*; (intr.) *Tengo que pensar sobre el asunto antes de dar una contestación.* (tr.): *Piénsalo cuanto antes*; (tr.) *Tengo que pensármelo mucho antes de aceptar.*
2. Decidir algo como consequência de ter pensado sobre um assunto: *He pensado que no me conviene ese negocio.*
3. Considerar; julgar; supor: *Yo pienso que no es ahora momento oportuno para eso.*

⁷ No caso de uma consulta para a qual o número de documentos/exemplos exceda o limite, o sistema oferece a possibilidade de filtragem (por documentos ou por casos), de maneira que o usuário tenha uma amostra representativa.

ARAÚJO JÚNIOR, Benivaldo José de. *Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

4. Ter intenção de fazer algo: *Pensamos marcharnos a primeros de mês.*
5. Inventar; conceber; imaginar un plano, procedimento ou meio para algo: (tr.) *Ya he pensado la manera de convencerle. Tengo pensado un plan magnífico para este verano.*
6. Aspirar a algo: (intr.) *Él piensa en una cátedra.*

3. Explorando o corpus

Primeiramente, fizemos consultas no CREA para as quatro formas de particípio do verbo *pensar*: *pensado*, *pensada*, *pensados*, *pensadas*. Os resultados podem ser visualizados na figura 1:

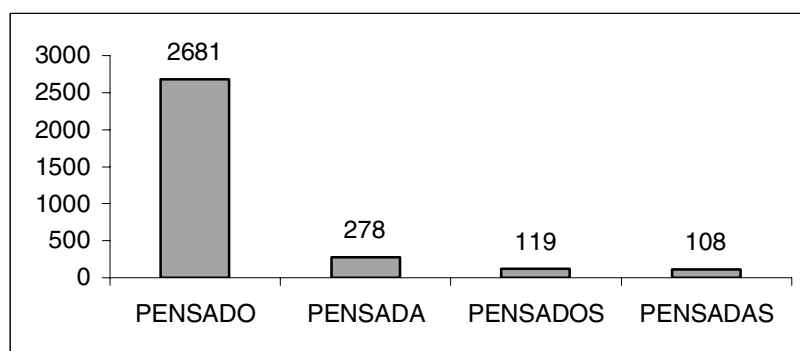


Figura 1: Total de ocorrências para as formas do particípio de *pensar*

O número elevado de ocorrências para *pensado* foi nosso primeiro problema, uma vez que o sistema não nos oferecia a possibilidade de separar do conjunto as incidências que não nos interessavam (caso dos tempos compostos). Para contornar essa dificuldade, decidimos analisar as ocorrências para as demais formas de particípio e, a partir dos padrões mais frequentes de estruturas passivas encontrados, fazer consultas específicas com o particípio *pensado*. Por exemplo, se ocorrem no corpus as construções *fue pensada*, *fueron pensados* e *fueron pensadas*, é muito provável que também ocorra *fue pensado*; dessa maneira, com essa expressão fazemos nova consulta ao corpus e assim por diante. Como complementação desse procedimento, realizamos outras consultas a partir de nossa intuição acerca da língua espanhola.

As análises dos resultados foram feitas a partir de concordâncias fornecidas pelo sistema para as palavras ou expressões pesquisadas (figura 2).

10									
11									
12									
13									
14									
15									
16									
17									
18									
19									

Figura 2: Fragmento das concordâncias fornecidas pelo CREA para a forma *pensada*

Após analisar as concordâncias para todas as consultas e agrupar os dados, encontramos os seguintes padrões para os participios de *pensar*:

- (1) verbo *ser* + participio (*pensado/a/os/as*), incluindo a conjugação de *ser* com tempos simples (*es pensado; fue pensada*, etc.) e com tempos compostos (*ha sido pensado*);
- (2) verbo + infinitivo de *ser* + participio (*pensado/a/os/as*), como em *puede ser pensado*;
- (3) verbo *estar* + participio (*pensado/a/os/as*), incluindo a conjugação de *estar* com tempos simples (*están pensados*) e com tempos compostos (apenas 1 ocorrência: *han estado pensados*);
- (4) verbo + infinitivo de *estar* + participio (*pensado/a/os/as*), como em *pueden estar pensados*;
- (5) verbo (diferente de *ser* e *estar*) + participio (*pensado/a/os/as*), como em *vienen pensadas*;
- (6) participio (*pensado/a/os/as*) com sentido passivo (*una exposición pensada para ser itinerante*);
- (7) verbo *tener* + participio (*pensado/a/os/as*), como em *tengo pensadas*;
- (8) advérbio *mal* + participio (*pensado/a/os/as*), equivalente a um adjetivo (*¡Qué mal pensada!*);

ARAÚJO JÚNIOR, Benivaldo José de. *Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

(9) advérbio *menos* + particípio (*pensado/a/os/as*), equivalente a um adjetivo (*el día menos pensado*).

As ocorrências de cada padrão podem ser visualizadas na figura 3:

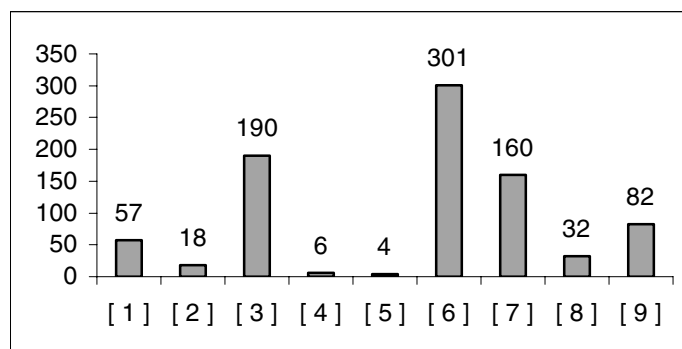


Figura 3: Padrões observados e suas ocorrências no corpus

Os padrões (7), (8) e (9) não foram considerados por nós como sendo estruturas passivas. No caso de (7), a estrutura *tener* + particípio expressa uma idéia próxima aos tempos compostos (*haber* + particípio), porém com um matiz de concretude que enfatiza o resultado da ação⁸, conforme se pode ver através dos exemplos:

(a) *Como muchos otros clubes, el Atlético de Madrid no ha podido concretar los partidos de la pretemporada porque está pendiente de la decisión de la Liga Profesional sobre el comienzo del campeonato. De cualquier forma, Javier Clemente ya **tiene pensadas** cosas al respecto y, sobre todo, ha decidido que quiere jugar pocos partidos.*

[CREA/España: ABC, 24/05/1989 (PRENSA), Tema 05: Deportes]

⁸ Nesse sentido, podemos considerá-la como uma **perífrase de resultado**, como o fazem alguns autores como Matte Bon (1995:101). Tal perífrase não é exclusiva do verbo *pensar*, ocorrendo com a maioria dos verbos, especialmente os transitivos. Apesar de esse padrão estar fora do nosso escopo, vale ressaltar que o mesmo ocorre 160 vezes no corpus de estudo, em torno de 5% das incidências; se considerássemos o uso de *tener* com outros particípios, esse percentual seria ainda maior. Por isso mesmo, parece-nos curioso que essa estrutura esteja ausente dos manuais e livros didáticos, uma vez que parece ser bastante utilizada na língua espanhola, como revela o corpus.

- (b) *Antes de esa intervención médica, Gallagher estaba planeando su retorno. El músico **tenía pensado** ofrecer una gira por el Reino Unido e Irlanda antes de iniciar su visita al resto de Europa.*

[CREA/España: *La Vanguardia*, 16/06/1995 (PRENSA), Tema 04: Música]

As estruturas (8) e (9) possivelmente continham rasgos passivos na origem, porém atualmente são expressões adjetivas (ou atributivas) fossilizadas, que integram o repertório fraseológico da língua espanhola. No caso de (8), a expressão *malpensado/a* pode ser traduzida como “desconfiado/a”, “malicioso/a”, “que pensa mal de algo/alguém”, e ocorre junto a nomes ou como atributo:

- (c) – *Os habrá contado que soy un cobarde o quizá algo peor, un afrancesado. – No ha dicho nada de eso, no seas **mal pensado**. Al contrario, comprende tus dudas.* [CREA/España: 1986; Gabriel y Galán, José Antonio (FICCIÓN), Tema 07: Novela]

- (d) – *Quien sabe, a lo mejor es para tapar todo lo que salga mal, o para ocultarse de la mirada severa del señor ministro que tiene las riendas de la economía. O puede ser, y será un simple error, que hay que ver qué **mal pensada** es la gente.* [CREA/España: ABC, 24/12/1983 (PRENSA), Tema 03: Política]

Em se tratando de (9), *menos pensado* aparece freqüentemente com palavras relacionadas a tempo (preferencialmente com *día*) para indicar que o fato ao qual nos referimos ocorrerá de maneira inesperada⁹:

- (e) *Ya no podemos negar la existencia del SIDA, ya que no es posible seguir ocultando la cabeza como dicen que hace el avestruz, preferimos creernos que por “arte de magia”, el día **menos pensado**, la pesadilla habrá concluido.*

[CREA/España: 1987, Lorenzo, Ricardo; Anabitarte, Héctor; Tema 06: Salud]

⁹ O padrão (9) ocorre 82 vezes no corpus de estudo (cerca de 2,6% das incidências), e vale aqui o mesmo comentário feito anteriormente para *tener* + participio: apesar da freqüência considerável, esse padrão não aparece nos manuais didáticos consultados. É curioso que, tampouco, esteja registrado no dicionário RAE (2001) ou em outros consultados, como o *VOX* (1997) e o *Clave* (2000); a única exceção foi o *María Moliner* (2001).

ARAÚJO JÚNIOR, Benivaldo José de. *Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

- (f) *Soy esposa y madre día y noche, qué remedio, pero en el momento **menos pensado**, yendo por la calle, por ejemplo, la mirada de un desconocido puede hacerme soñar... ¿Me comprende? No, supongo que no -sonríe nuevamente con aire de tomarle el pelo-. Usted no me conoce.*
[CREA/España: 2000; Marsé, Juan: *Rabos de lagartija* (FICCIÓN), Tema 07: *Novela*]

Com a exclusão dos padrões (7), (8) e (9), ficamos com 576 ocorrências de passivas no corpus. Uma estimativa para as formas ativas do verbo *pensar*, também através de pesquisas no CREA¹⁰, fornece aproximadamente 28.346 casos. Portanto, uma conta simples nos revela percentuais de 98% para as ativas e apenas 2% para as passivas com participio, o que vem ratificar o caráter da passiva como uma estrutura marcada. Vale recordar que neste estudo não foi investigada a passiva pronominal (com o pronome *se*, ou *pasiva refleja*) com *pensar*. É certo que, caso fossem computadas as pronominais, aumentaria o percentual de passivas no corpus; ainda assim, tal acréscimo pouco afetaria os resultados, já que a preferência dos hispanofalantes parece ser mesmo pelas construções ativas¹¹. Fato observado nas construções passivas do corpus foi a presença do **desencadeador**¹² (**agente**, se pensamos na denominação clássica para a função) em posição secundária (3,8%) ou sua ausência (96,2%), o que parece corroborar a hipótese de Duarte (1990), segundo a qual a função básica das passivas seria a **detematização** do sujeito/agente (desencadeador), por meio da sua

¹⁰ Para essa pesquisa, utilizamos formas lematizadas do verbo (*piens**, *pensab**, *pensar** e *piensas**) e em alguns casos tivemos que entrar com as próprias formas do verbo (*pensamos*, *pensáis*, *pensó*, etc.). Isso se deveu ao fato de que o corpus de referência não dispõe de uma versão marcada morfologicamente, que nos permitisse selecionar apenas as formas verbais (excluindo automaticamente as não verbais, como *pensamiento/s*, *pensador/a/es/as*, *pensativo/a/os/as*, *pensión*, *pensionista/s*, etc.).

¹¹ Pode-se comprová-lo pelo estudo realizado por Barrenechea e Manacorda de Rosetti (1979: 65) para o espanhol falado em Buenos Aires, para o qual as ativas ocorrem em 97,81% dos casos. Como se pode ver, tais resultados estão muito próximos aos que obtivemos com a análise dos dados do CREA, para a variante peninsular.

¹² Os conceitos de **desencadeador** e **afetado** provêm do trabalho de Caçado (Um estatuto teórico para os papéis temáticos. In: Müller, A.L.; Negrão, E.; Foltran, M.J. *Semântica Formal*, São Paulo, Contexto, 2003, p. 95-124), cuja proposta é a definição dos papéis temáticos não como noções (agente, paciente, instrumento, força etc.) e sim como uma composição de propriedades (desencadeador, afetado, estado e controle) atribuídas a um dado argumento a partir dos acarretamentos estabelecidos por toda a proposição em que esse argumento se encontra.

omissão [ver exemplo (g) abaixo] ou recodificação (passando o agente/desencadeador para uma posição secundária na estrutura, regido de preposição; ver exemplo (h)).

Façamos agora algumas considerações específicas sobre os padrões que classificamos como passivos.

A estrutura (1), ou passiva sintática (*ser* + particípio), ocorre 57 vezes no corpus e parece estar vinculada a contextos de alto grau de formalidade, relacionados principalmente às ciências humanas (filosofia, história, etc.); o verbo *pensar* é usado aqui para designar entidades concebidas pela razão, numa acepção que está distante da linguagem corrente:

(g) *La ciudadela, una especie de sarcófago pentagonal, queda encima de una peña encrespada. Pero parece que no **fue pensada** como castillo, sino como santuario.*

[CREA/España: *La Vanguardia*, 23/06/1994 (PRENSA), Tema 02: Historia]

(h) *Mientras las máquinas son productos de nuestra inteligencia, los seres humanos **somos «pensados»** y contruidos por nuestros genes egoístas.*

[CREA/España: 1995; Acero, Juan José, *Teorías del contenido mental*; Tema 02: Filosofía]

No exemplo (g) podemos atribuir à passiva sintática a função de **detematização** do **agente/desencadeador**, com conseqüente omissão desse elemento. Conseqüentemente, a posição de destaque e relevo no discurso (tema) passa ao elemento afetado/paciente *la ciudadela* (não presente na frase, porém recuperável no contexto); em resumo, é o processo de **tematização**¹³ do **afetado/paciente**. Em (h) a passiva sintática também detematiza o agente/desencadeador, conferindo-lhe uma posição secundária; porém, embora em segundo plano do ponto de vista temático, esse agente detematizado converte-se num elemento de realce (foco) e introduz informação nova na estrutura (os genes egoístas que, segundo o autor, são os que “pensam” e constroem os seres humanos): é o processo de **focalização** do **agente/desencadeador**.

¹³ O uso da *impersonal activa*, se aplicado a esse exemplo, teria a mesma função de tematização: (...) *Pero parece que no **la pensaron** como castillo, sino como santuario.* Mesmo sendo uma estrutura ativa, é evidente seu caráter passivo, o que também ratifica, de outra perspectiva e com outras formas, o parentesco entre as passivas e as chamadas *impersonales* do espanhol (nossas indeterminadas).

ARAÚJO JÚNIOR, Benivaldo José de. *Estructuras Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

A estrutura (2), verbo + *ser*+ participípio, ocorre associada a um efeito de **modalização** próprio do discurso científico (formal, portanto), no qual é comum que se evitem as afirmações categóricas sobre o tema/objeto tratado. Não é por acaso que o verbo *poder* aparece como o mais freqüente nas incidências (do total de 18, 50% ocorrem com esse verbo), seguido de outros modais como *parecer* e *deber*:

- (i) *La noción misma de estímulo se convierte así en la principal dificultad: el spot ya no **puede ser pensado** tan sólo como el estímulo que genera una conducta, sino como el objeto mismo de una nueva conducta difícilmente homologable: la de consumir spots.*
[CREA/España: 1995; González Requena, Jesús; Ortiz de Zárate, Amaya, El spot publicitario. Las metamorfosis del deseo; Tema 04: Publicidad]

Verbos não modais também aparecem com (2), relacionados a outros efeitos de sentido. É o caso de *volver*, que introduz um matiz reiterativo, conforme o exemplo a seguir:

- (j) *La Paleontología posterior al histórico coloquio de Princeton de 1947 no podía orillar a partir de entonces ideas y conceptos biológicos cuyo conocimiento se hacía inexcusable para la correcta interpretación de los conjuntos fósiles. De nuevo, la Paleontología **volvía a ser pensada** realmente como ciencia biológica.*
[CREA/España: 1987; Truyols, Jaime, Desarrollo histórico de la Paleontología contemporánea en España; Tema 01: Zoología y Paleontología]

No que diz respeito à estrutura (3) (*estar*+ participípio), não existe um consenso quanto a classificá-la como passiva. Não aparece como tal nos manuais didáticos, e o mais comum é que a definam como uma perífrase de resultado [Matte Bon (1992:101)]. Consideramos que (3) é uma passiva, pondo-nos de acordo com as afirmações de Gili Gaya (1978:125): “La acción verbal que expresa la pasiva con *ser* se produce en el tiempo que expresa el verbo auxiliar: *el suceso es, era, fue, será comentado*. Con *estar*, la acción se da como terminada y cumplida antes del tiempo que indica el auxiliar: decimos que un problema *está resuelto* (presente), cuando *ha sido resuelto* (antepresente); decimos que *estaba resuelto* (imperfecto), cuando *había sido resuelto* (pluscuamperfecto); que *estará resuelto* cuando *habrá* o *haya sido resuelto*; es decir, que *estar resuelto* es el resultado de *haber sido resuelto*”.

Vejam os exemplos do corpus:

- (k) *La «nacionalidad infantil» **está pensada** para los extranjeros de tercera generación. Podrán obtenerla los niños nacidos en Alemania y uno de cuyos padres también haya nacido aquí, además de otros requisitos.*
[CREA/España: *La Vanguardia*, 15/11/1994 (PRENSA), Tema 03: Política]

Além de expressar o resultado de uma ação acabada, a estrutura tem um matiz de atualização, muito provavelmente associado ao sentido local de *estar*. Ou seja, quando dizemos que *la nacionalidad infantil **está pensada**...*, fica clara a vinculação do tema com o momento presente; tal matiz não seria obtido caso disséssemos que *la nacionalidad infantil **fue pensada**...*: nesse caso, enfatizaríamos apenas a ação, e não a atualidade do tema. Portanto, não é à toa que a passiva com *estar* seja recorrente no gênero publicitário (*un coche que está pensado para toda la familia* etc.). Diferente do observado para *ser + pensado/a/os/as* – identificada ao discurso mais formal (sobretudo o científico) ^{3/4}, a passiva com *estar + pensado/a/os/as* se revela mais permeável a outros gêneros, transitando do informal ao formal; também incide no corpus três vezes mais que sua “concorrente”.

Observando os exemplos do corpus para o padrão, vimos que a ocorrência mais freqüente é *estar + pensado/a/os/as + para* (308 casos), seguido de um nome (normalmente na função de **beneficiário**, como no exemplo (k) acima) ou de uma forma verbal infinitiva ou conjugada (antecedida por *que*), que resulta numa expressão de finalidade:

- (l) *Los equipamientos de la localidad **están pensados** para satisfacer las necesidades de los vecinos y, en lo que a educación se refiere, en principio, la situación de partida es bastante equilibrada, pudiendo garantizar la escolarización de todo el municipio.*
[CREA/España: *El mundo – Su vivienda* (Suplemento), n.º. 236, 18/01/2002 (PRENSA), Tema 05: Vivienda]

O padrão (4) – verbo + infinitivo de *estar* + particípio – é bastante semelhante ao (2) nos usos, portanto relacionado aos efeitos de modalização do discurso. Também em (4) a preferência nas ocorrências é para o verbo *poder*:

- (m) *En esta ocasión, al igual que en tantas otras, no lo tuvieron demasiado fácil, a pesar de que la respuesta que se les exigía ante el clima de crispación reinante **podía estar pensada** y reflexionada. Y la respuesta que ofrecieron*

ARAÚJO JÚNIOR, Benivaldo José de. *Estructuras Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

no convenció a muchos, porque todavía colea el asunto y no se ha retirado totalmente la marejada.

[CREA/España: El Norte de Castilla, 15/06/2002 (PRENSA), Tema 03: Iglesia]

O padrão (5) – verbo (diferente de *ser* e *estar*) + *pensado/a/os/as* – foi o que teve menor número de ocorrências (apenas 4), e talvez mereça nosso comentário exatamente por isso: usar uma estrutura de baixa frequência não raro se relaciona a uma busca de expressividade ou efeitos de sentido que as soluções mais correntes na língua não oferecem. Vejamos o exemplo a seguir:

(n) (...) *asuntos que de ninguna manera pudiesen parecer inventados sobre la marcha para llenar huecos, sino que tenían la solidez, la consistencia, la armazón de las cuestiones que, de tiempo atrás, vienen pensadas y ordenadas, y siguen al individuo como la piel al cuerpo.*

[CREA/España: 1984; Ayerra, Ramón, *La lucha inútil* (FICCIÓN), Tema 07: Novela]

A passiva com *venir* é classificada por Lorenzo(1980) como *inconclusa* (não concluída), e destaca um processo que se iniciou no passado e segue em andamento no presente. O verbo *venir* também enriquece a estrutura com um matiz iterativo, de recursividade: as **questões** referidas no exemplo sempre voltam a **ser pensadas**, em busca do refinamento e da solidez.

Finalmente, o padrão (6) – mais freqüente dentre todos os tratados neste estudo – consiste no participípio (*pensado/a/os/as*) com sentido passivo, sem nenhum auxiliar. A estrutura em questão é classificada por Moino (1989: 40) como passiva lexical. Pode ocorrer dentro do sintagma nominal¹⁴:

(o) *Esta vez la llamada fue al final de la cumbre pensada por la OTAN para celebrar su 50 aniversario, pero que resultó opacada por la crisis de Kosovo.*

[CREA/España: El Diario Vasco, 27/04/99 (PRENSA), Tema 03: Política]

E também encabeça construções apositivas, conforme o exemplo:

(p) *La editorial Trotta ha comenzado a publicar la obra del filósofo J.L.L. Aranguren. La obra, pensada por el momento en seis volúmenes, contiene*

¹⁴ Observar que o exemplo na seqüência (o) vem com agente explícito: *la OTAN*.

Crop, 10, 2004

en su primero lo que bajo el título Filosofía y religión pertenece a la época temprana del profesor Aranguren.

[CREA/España: *El Mundo*, 24/09/1994 (PRENSA), Tema 02: Literatura]

O padrão (6) incidiu 301 vezes no corpus, superando as passivas com *ser* (57) e *estar* (190). A razão dessa preferência parece estar relacionada à busca de maior concisão no discurso escrito. De fato, a passiva lexical torna possível um “enxugamento”, uma condensação do discurso, sem prejuízos para o entendimento do leitor. Basta ver como (o) resultaria mais extensa, desnecessariamente, caso introduzíssemos a passiva sintática: *Esta vez la llamada fue al final de la cumbre **que fue pensada** por la OTAN para celebrar su 50 aniversario, pero que resultó opacada por la crisis de Kosovo.* A explicação faz sentido se pensamos que a maioria dos exemplos do corpus provém de artigos jornalísticos, cuja limitação de espaço impõe que se diga o máximo com um mínimo de palavras.

Pensando neste estudo como um todo, o levantamento das passivas com participio para o verbo *pensar* a partir do CREA, mais que confirmar o que se diz nos manuais e gramáticas acerca do assunto, revela padrões freqüentes na língua espanhola (caso de (3) e (6)) que não aparecem nessas publicações. As amostras do corpus também sinalizam a relevância das funções temáticas, das funções informativas (pragmáticas) e dos gêneros do discurso para a abordagem das passivas, o que nem sempre é levado em conta nos materiais didáticos de E/LE.

E aqui termina nossa análise. Reconhecemos que o estudo teria ficado mais completo caso tivéssemos incorporado as passivas pronominais com *pensar*; optamos por não fazê-lo, já que teria sido necessário incorporar à análise as estruturas impessoais (que muitas vezes se confundem com as passivas pronominais), aumentando em muito o escopo deste trabalho. Entretanto, a partir dos resultados que obtivemos, nada impede retomar este estudo em outra oportunidade, não só incluindo as pronominais e impessoais, como também o trabalho com enunciados orais.

4. Conclusão

Em nosso estudo nos propusemos a analisar as estruturas passivas com participio para o verbo *pensar*, a partir de amostras do discurso escrito coletadas de um corpus de referência da língua espanhola. Feitas as descrições e os comentários acerca dos principais padrões observados, fica clara a necessidade de ampliar o repertório presente nos manuais didáticos com respeito ao tema: por um lado, explorar com mais propriedade os usos das passivas, não a partir de noções

ARAÚJO JÚNIOR, Benivaldo José de. *Estruturas Passivas no Espanhol Escrito: um Estudo de Caso Utilizando a Lingüística de Corpus*.

que se pretendem gerais (os papéis temáticos fixos de agente e paciente, a transitividade dos verbos, etc.), e sim por meio dos gêneros do discurso em que ocorrem e das funções informativas que estão em jogo (por exemplo, o foco); por outro lado, introduzir outras estruturas passivas, como aquelas com o verbo *estar* e as passivas lexicais. Dessa forma, o manejo do corpus de referência possibilitou-nos tanto a confirmar as hipóteses já lançadas sobre o assunto, quanto revelar de novos padrões de uso das passivas. Finalmente, os resultados que obtivemos nesta análise também nos permitiriam vislumbrar na lingüística de corpus uma ferramenta poderosa no e para o ensino da gramática – o que nos servirá de tema e inspiração para futuros trabalhos nessa área.

Referências Bibliográficas

- ALARCOS LLORACH, E. *Estudios de gramática funcional del español*. Madrid: Gredos, 1970. p.90-94, 124-132.
- _____. *Gramática de la Lengua Española*. Madrid: Espasa Calpe, 1994.
- ALVAR EZQUERRA, M. (Ed.). *Vox: Diccionario Avanzado de la Lengua Española*. 15.ed. Barcelona: Bibliograf, 1997.
- BARRENECHEA, A. M.; MANACORDA DE ROSETTI, M. V. La voz pasiva en el español hablado en Buenos Aires. In: *Estudios lingüísticos y dialectológicos*. Buenos Aires, Hachette, 1979. p. 61-72
- CASTRO, F. *Uso de la gramática española: elemental*. Madrid: Edelsa, 1996.
- _____. *Uso de la gramática española: elemental*. Madrid: Edelsa, 1996.
- _____. *Uso de la gramática española: intermedio*. Madrid: Edelsa, 1997.
- _____. *Uso de la gramática española: avanzado*. Madrid: Edelsa, 1997.
- CERROLAZA, M.; CERROLAZA, O.; LLOVET, B. *Planet@ 1: Libro del alumno*. 2.ed. Madrid: Edelsa, 1998.
- _____. *Planet@ 1: Libro de referencia gramatical*. 2.ed. Madrid: Edelsa, 1998.
- _____. *Planet@ 2: Libro del alumno*. Madrid: Edelsa, 1999.
- _____. *Planet@ 2: Libro de referencia gramatical*. Madrid: Edelsa, 1999.
- _____. *Planet@ 3: Libro del alumno*. Madrid: Edelsa, 2000.
- _____. *Planet@ 3: Libro de referencia gramatical*. Madrid: Edelsa, 2000.
- _____. *Planet@ 4: Libro del alumno*. Madrid: Edelsa, 2000.
- _____. *Planet@ 4: Libro de referencia gramatical*. Madrid: Edelsa, 2000.
- DOMÍNGUEZ, P.; BAZO, P. *Claves del Español: Gramática Práctica*. Madrid: Santillana, 1994.
- DUARTE, Y. As passivas do português e do inglês: uma análise funcional. In: *Revista D.E.L.T.A*, v. 6, n. 2. São Paulo, 1990. p. 139-167.
- GILI GAYA, S. *Curso superior de sintaxis española*. 12.ed. Barcelona: Bibliograf, 1978. p. 121-129.

Crop, 10, 2004

- LORENZO, E. Sobre el talante y el semblante de la lengua española. In: *El español y otras lenguas*. Madrid: SGEL, 1980. p. 17-22.
- MALDONADO GONZÁLEZ, C. (Ed.). *Clave: Diccionario de uso del español actual*. 4.ed. Madrid: SM, 2000.
- MATTE BON, F. *Gramática comunicativa del español*. Madrid: Edelsa, 1992. v.1, p. 101.
- MOINO, R. E. L. Passivas nos discursos oral e escrito: No princípio era o verbo. E o verbo se fez adjetivo? Ou... O que estamos fazendo no oral! In: TARALLO, F. (Org.). *Fotografias Sociolingüísticas*. Campinas: Pontes, 1989, p. 35-50.
- MOLINER, M. *Diccionario de uso del español*. 2.ed. Madrid: Gredos, 1998.
- REAL ACADEMIA ESPAÑOLA. *CREA – Corpus de Referencia del Español Actual*. Disponível em < <http://corpus.rae.es/creanet.html> >. Acesso no período de 30/05/04 a 15/07/04.
- REAL ACADEMIA ESPAÑOLA. *Diccionario de la lengua española*. 22.ed. Madrid: Espasa Calpe, 2001.



A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil

*Catherine Anne Lonngren Sampaio**

Resumo: *Este artigo descreve a investigação de alternância de código (codeswitching) em um corpus eletrônico de linguagem bilíngüe infantil (português e inglês) construído seguindo o sistema de transcrição do projeto CHILDES (Child Language Data Exchange System), um sistema de troca de dados de linguagem infantil originalmente desenvolvido na Universidade Carnegie Mellon nos Estados Unidos. A pesquisa enquadra-se na área da Lingüística de Corpus, que pode ser caracterizada pela investigação de fenômenos lingüísticos baseada em exemplos reais de linguagem, armazenados em corpora eletrônicos, compilados de forma sistemática e segundo determinados critérios. O corpus analisado no presente trabalho é composto de transcrições da linguagem falada de duas crianças bilíngües que estão aprendendo português e inglês simultaneamente. Os dados foram coletados através de gravações durante um período de aproximadamente dois anos e transcritos seguindo as convenções estabelecidas pelo CHILDES. Fora as convenções mínimas, foram inseridos códigos específicos nas transcrições que permitissem a investigação do uso que as duas crianças fazem da alternância de código, tanto de uma perspectiva estrutural quanto de uma perspectiva pragmática. No caso do estudo específico relatado aqui, foram analisadas nove conversas telefônicas gravadas durante o período de dois meses que as crianças passaram na Inglaterra. As conversas têm como principais interlocutores as duas crianças e o pai brasileiro.*

* Mestre em Lingüística Aplicada da Universidade Estadual do Ceará.

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil.*

Através das ferramentas do CHILDES investigou-se tanto a alternância de código para inglês quanto para o português.

Palavras-chave: *Linguística de Corpus; CHILDES; linguagem bilíngüe; alternância de código.*

Abstract: *This article describes the investigation of codeswitching in a computerised corpus of child bilingual language (English/Portuguese) built following the transcription system of the CHILDES (Child Language Data Exchange System) project, a data exchange system of child language originally developed at Carnegie Mellon University in the United States. The research belongs to the area of Corpus Linguistics, which is characterized by the investigation of linguistic phenomena based on examples of real-life language stored in computerised corpora compiled in a systematic manner and following certain criteria. The corpus analysed in the present article is composed of transcriptions of the spoken language of two bilingual children who are learning Portuguese and English simultaneously. The data were collected through recordings carried out over a period of approximately two years and transcribed according to the conventions established by CHILDES. Apart from the minimum conventions, specific codes were inserted in the corpus which permitted the investigation of the use which both children made of codeswitching from both structural and pragmatic perspectives. In the case of the specific study reported here, nine telephone conversations, recorded while both children were in England for a period of two months, were analysed. With the children and their father, a Brazilian, as the main interlocutors, codeswitching to both English and Portuguese was investigated through the tools of analysis which integrate CHILDES.*

Keywords: *Corpus Linguistics; CHILDES, bilingual language; codeswitching.*

1. Introdução

A investigação relatada no presente trabalho faz parte de uma pesquisa de mestrado realizada pela Universidade Estadual do Ceará na qual foi descrito

¹ Seguindo Sardinha (1999a; 1999b), os termos 'corpus' e 'corpora' não serão italizados no presente trabalho.

o processo da construção e exploração de um corpus¹ eletrônico de linguagem bilíngüe infantil seguindo os padrões do CHILDES (Child Language Data Exchange System)². Este projeto multilíngüe foi originalmente concebido em 1981, por Brian MacWhinney, Dan Slobin, Willem Levelt e Susan Ervin-Tripp, nos Estados Unidos. O objetivo original do projeto foi estabelecer um sistema de transcrição padronizado e sistematizado que permitisse a prática de troca de dados, armazenados em corpora eletrônica, entre pesquisadores do mundo todo, possibilitando progresso no estudo científico da aquisição da linguagem. O sistema CHILDES, disponível de forma gratuita pela Internet, constitui-se dos seguintes três componentes:

1. O banco de dados do CHILDES, que conta com corpora de diversas fontes, contribuições de mais de 100 pesquisadores que trabalham na área da linguagem infantil. Há dados de várias línguas³, e, sendo um banco de dados aberto, qualquer pesquisador pode contribuir com seus dados, contanto que estejam transcritos em formato CHAT (veja abaixo).
2. CHAT (Codes for the Human Analysis of Transcripts) que consiste de um sistema de transcrição cujo objetivo é uniformizar a codificação de corpora para permitir a análise através do uso de ferramentas eletrônicas.
3. CLAN (Computerized Language Analysis) que é um programa especificamente desenvolvido para analisar dados transcritos no formato CHAT.

O CHAT é um sistema de transcrição muito claro e completo, fácil de aplicar nos dados e fácil de ler. Além da transcrição propriamente dita, que focaliza “a produção de um registro escrito que nos permite compreender, embora de maneira vaga, o fluxo da interação original”⁴ (MacWhinney, 2003:13), o sistema CHAT inclui uma série de convenções para codificação, definida por MacWhinney como “o processo de reconhecer, analisar e tomar nota de fenômenos no discurso falado transcrito”⁵ (p. 13). Desde que o corpus seja anotado

² A dissertação completa está disponível no seguinte endereço: www.lonngren.com.br/cathy.

³ A natureza de cada corpus é descrita no volume II do manual mencionado na nota acima.

⁴ Tradução minha do original: “...the production of a written record that can lead us to understand, albeit only vaguely, the flow of the original interaction”.

⁵ Tradução minha do original: “...the process of recognizing, analyzing, and taking note of phenomena in transcribed speech”.

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

segundo o padrão sugerido no manual do CHILDES, o programa de análise, CLAN, pode ser utilizado na sua exploração. As possibilidades de análises aumentam com a inserção de códigos, a escolha dos quais vai variar de acordo com o objetivo do pesquisador. Em relação ao corpus da pesquisa, foram inseridos códigos criados pela própria pesquisadora que permite a análise das línguas separadamente, possibilitando, assim, a investigação de alternância de código. No presente artigo, pretende-se demonstrar a metodologia utilizada na investigação desse fenômeno e discutir os resultados obtidos das análises eletrônicas realizadas através do programa que integra ao CHILDES. Antes, entretanto, será descrita a seguir a natureza do corpus.

2. O LOBILL Corpus

Os dados usados na pesquisa fazem parte de um corpus aberto, denominado LOBILL⁶ Corpus, composto da linguagem falada de duas crianças bilíngües nas suas interações com interlocutores mono e bilíngües, em diversas situações de família. Os informantes principais são MEG⁷, uma menina de 8;2 anos⁸, e JAM, seu irmão de 5;10 anos⁹ (tendo respectivamente 5;10 e 3;5 anos no início da coleta de dados). Ambos nasceram em Fortaleza e desde 1;6 anos freqüentam uma escola brasileira. A mãe, MOT, que é a própria pesquisadora, é inglesa, casada com um brasileiro que fala inglês fluentemente. Com mais de 10 anos no Brasil, MOT fala português fluentemente. Embora desde o nascimento, a mãe interaja com os dois filhos, MEG e JAM, exclusivamente em inglês, o pai, PAI, se dirige a eles em português a maior parte do tempo. A interação entre os irmãos é predominantemente em português. Fora a interação com a mãe, o contato diário com o inglês se restringe a programas de televisão ('Cartoon Network' e 'Discovery Channel') e livros de histórias. Visitas ocasionais de parentes ingleses proporcionam uma outra fonte importante de contato com inglês e ambas as crianças têm passado temporadas curtas na Inglaterra; a última

⁶ LOBILL significa Lonngren BILingual Language, Lonngren sendo o sobrenome dos sujeitos.

⁷ Nesse artigo, os informantes são indentificados pelos códigos de identidade que foram utilizados nas transcrições: seguindo as normas do CHAT, o código é composto de uma combinação de três letras maiúsculas, que pode ser baseada no nome próprio do falante, por exemplo, MEG (Meggie), ou no seu papel, por exemplo, MOT (Mother).

⁸ Em dezembro 2003.

⁹ Em dezembro 2003.

sendo durante os meses de junho, julho e agosto de 2003. Os quatro membros da família são considerados 'bilíngües balanceados' (*balanced bilinguals*) (Baker, 2000:5). Embora o conceito 'bilíngüe balanceado' não seja bem definido (p. 5-7), aqui se refere a uma pessoa que tem 'competência aproximadamente igual em duas línguas' (p. 167)¹⁰.

Atualmente, contamos com, aproximadamente 18 horas de gravações. A coleta dos dados teve início em agosto de 2001 e continuou ao longo da pesquisa. Os dados foram coletados principalmente por meio de gravações em áudio, com um microcassete para permitir a gravação das conversas sem as crianças perceberem a sua presença. Entretanto, as gravações também incluem 'entrevistas' com os sujeitos, que, sendo pré-planejadas, ocorreram com o conhecimento pleno das crianças. Anotações de campo complementam as gravações.

No início da coleta de dados, não foi seguido nenhum sistema de transcrição devido ao fato de existirem tantos sistemas diferentes que não pareciam se adequar ao objetivo da pesquisa. As transcrições eram muito básicas e não faziam uso de convenções específicas. Ao conhecer o CHILDES, ficou evidente que o CHAT oferecia o que estava sendo procurado: uma maneira sistemática de transcrever os dados do corpus, que, ao mesmo tempo, abrisse a possibilidade de realizar análises alternativas. Fora as convenções de transcrição, foram inseridos códigos para permitir a análise das línguas separadamente: em interações em que a língua principal foi identificada como inglês (indicado pelo 'en' no nome do arquivo, por exemplo 27enJ&MNOV02.cha), somente o material em português foi marcado (<*material em português*>[@pt]); em interações em que a língua principal foi identificada como português (indicado pelo 'pt' no nome do arquivo, por exemplo 52ptJ&MJUN03.cha), somente material em inglês foi marcado (<*material em inglês*>[@en]). É fora do âmbito desse artigo descrever em detalhes todos os códigos que foram inseridos nas transcrições durante o processo da construção do LOBILL Corpus seguindo CHAT (veja dissertação para mais detalhes). Entretanto, a inserção dos códigos mencionados acima foi fundamental para a realização das análises com CLAN na investigação do uso de alternância de código pelas duas crianças, assunto da próxima seção.

3. A investigação de alternância de código no LOBILL Corpus

Atualmente o termo alternância de código (*codeswitching* em inglês) é utilizado para descrever "qualquer alternância durante uma única conversa,

¹⁰ Tradução minha do original: "...approximately equal competence in two languages".

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*.

podendo ser no nível da palavra ou sentença ou no nível de blocos de discurso”¹¹ (Baker, 2001:101). Não é um fenômeno puramente lingüístico, mas envolve muitas variáveis cognitivas, psicológicas, sociais e políticas, que determinam a sua natureza em um contexto específico. No estudo sobre alternância de código, muitos lingüistas investigam as possíveis restrições lingüísticas que acreditam governar a alternância de código (veja Poplack, 1980; MacSwan, 1999; Myers-Scotton, 2002). Embora não caiba aqui discutir os modelos de alternância de código propostos por esses teóricos, considerou-se útil para a discussão dos resultados o uso dos termos ‘Língua Matriz’ (*Matrix Language*) e ‘Língua Encaixada’ (*Embedded Language*) (Myers-Scotton, 2002). Baseado num construto gramatical, a Língua Matriz é aquela que mais contribui com as estruturas morfossintáticas para sentenças bilíngües (freqüentemente considerada a língua ‘dominante’), enquanto que a Língua Encaixada é caracterizada pela sua contribuição com morfemas de conteúdo (como substantivos, verbos, adjetivos e preposições) (p. 62). Essa assimetria entre as línguas participantes na alternância de código e as possíveis restrições estruturais que resultam podem ser investigadas no LOBILL Corpus através do CLAN, pelo uso dos comandos ‘**FREQ**’ (de ‘**FREQ**uency’ – ‘freqüência’) e ‘**KWAL**’ (‘**Key**Word **And** **Line**’ – palavra e linha chave), conforme descrito a seguir.

Na sua forma mais básica, FREQ produz uma lista de freqüência das palavras de um ou mais arquivos. Ao especificar o falante e a língua, os resultados produzidos podem ser analisados para averiguar o uso de cada língua em relação à categoria gramatical das palavras. A repetição dessas análises em outros arquivos possibilita a investigação de tendências no uso das duas línguas para certas categorias gramaticais e, portanto, a investigação da oposição entre Língua Matriz e Língua Encaixada. Uma análise detalhada das listas produzidas por FREQ pode revelar várias categorias gramaticais, como verbos, artigos, preposições, conjunções e marcadores discursivos.

O comando KWAL pode ser utilizado para investigar o uso dessas palavras em contexto, e outras questões relacionadas à estrutura morfossintática de frases bilíngües, como os possíveis pontos de alternância. Através do uso dos dois comandos (FREQ e KWAL), portanto, a relação entre o uso das duas línguas e a estrutura gramatical de frases bilíngües pode ser investigada.

Além do estudo das restrições lingüísticas de alternância de código, outras pesquisas têm como foco a investigação de aspectos pragmáticos, isto é, como os falantes utilizam a alternância de código de acordo com: ...a situação (domínios sociais, graus de formalidade do discurso etc.), os participantes da interação (idade,

¹¹ Tradução minha do original: “...any switch within the course of a single conversation, whether at word or sentence level or at the level of blocks of speech”.

sexo, *status*, origem étnica, grau de relação interpessoal etc.), o tópico (afetivo, escolar etc.) e a função da interação (um pedido, uma ordem, um argumento etc.) (Brito de Mello, 1999:122)

Para investigar no LOBILL Corpus a alternância de código em relação a essas quatro variáveis (situação, participantes, tópico e função), como veremos, novamente os comandos *FREQ* e *KWAL* oferecem uma maneira eficiente de fazê-lo. Entretanto, análises específicas somente se tornaram possíveis devido à inserção dos códigos e cabeçalhos que marcam a língua da interação, o interlocutor e seu interlocutor, e a situação.

Com respeito à variável 'tópico', ambos os comandos *FREQ* e *KWAL* podem revelar evidências sobre a relação do uso do português ou inglês e o tópico da interação. Através de outras análises similares no corpus, talvez seja possível determinar os 'domínios' de cada língua e investigar a estabilidade desses domínios nos dados longitudinais.

A investigação das várias funções da alternância de código é outra área que pode ser explorada no LOBILL Corpus através dos comandos do *CLAN*. Por exemplo, foi observado nas interações que a alternância de código freqüentemente ocorre quando um dos falantes utiliza discurso direto. No LOBILL Corpus trechos de discurso direto foram transcritos da seguinte forma:

***JAM:** é assim ó +”/.

***JAM:** +” there's too much butter on those trays +”.

Utiliza-se +”/ para indicar que o material que segue na linha seguinte é discurso direto e na linha seguinte insere-se +” para indicar o início do discurso direto e +” para marcar seu término. Dessa forma, ao pedir uma análise de *KWAL* para o falante *JAM* em vários arquivos, seria possível averiguar outras ocorrências desse tipo de uso de alternância de código (indicadas pelos símbolos mostrados acima). Uma outra função fácil de analisar no corpus é o uso de alternância de código para fazer referências metalingüísticas do tipo ***JAM: tu sabe que é um buffer[“]?** Na construção da linha do comando do *KWAL*, a inserção de +s”[“] selecionaria todas as ocorrências desse tipo nos arquivos especificados do corpus. Embora, nas duas análises citadas acima, os símbolos específicos inseridos nas linhas principais facilitem a especificação dos dados para a análise, não é necessário (ou viável) codificar todas as possíveis funções de alternância de código encontradas dentro do corpus. O importante é poder selecionar nos arquivos desejados os enunciados em que ocorre alternância de código, isto é, em que se encontra a Língua Matriz e a Língua Encaixada no mesmo enunciado. Essa seleção é possível devido à inserção dos símbolos **<material em portugues>[@pt]** ou **<material em ingles>[@en]** em todas as transcri-

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

ções. Após a seleção através do uso de KWAL, os resultados apresentados podem ser analisados com o objetivo de identificar e classificar as várias funções de alternância de código (por falante).

Fica evidente pela discussão acima que, no LOBILL Corpus, tanto os aspectos pragmáticos quanto os aspectos estruturais da alternância de código são passíveis de investigação. Com o objetivo de ilustrar uma análise específica de forma mais detalhada, na próxima seção, serão mostrados e discutidos os resultados da investigação de alternância de código em 9 arquivos que contêm interações da mesma natureza.

4. A análise de alternância de código em nove conversas telefônicas

4.1. Os arquivos analisados

Os arquivos escolhidos para análise compartilham certas características: são gravações de conversas telefônicas que os dois sujeitos tiveram com parentes brasileiros, durante o período de dois meses que passaram na Inglaterra de férias. Durante todas as ligações, a mãe dos sujeitos esteve presente, sentada ao lado, e também conversou ao telefone. As conversas da mãe com os parentes no Brasil não foram gravadas, mas a interação entre ela e os dois sujeitos durante as conversas faz parte das gravações. A tabela a seguir traz informações sobre data e duração (em minutos) das gravações, os participantes das conversas e o nome dos arquivos:

Tabela 1: Dados dos arquivos do LOBILL Corpus utilizados nas análises

No.	Data	Duração	Participantes	Idade dos sujeitos	Nome do arquivo
1	4.07.03	4.29	JAM, VIN, MOT	JAM: 5;4.3	59ptJUL03.cha
2	4.07.03	7.40	JAM, MEG, PAI, MOT	JAM: 5;4.3 MEG: 7;8.27	60ptJ&MJUL03.cha
3	12.07.03	16.15	JAM, MEG, PAI, MOT	JAM: 5;4.11 MEG: 7;9.5	62ptJ&MJUL03.cha
4	17.07.03	11.0	JAM, MEG, PAI, MOT	JAM: 5;4.16 MEG: 7;9.10	64ptJ&MJUL03.cha
5	24.07.03	12.44	JAM, MEG, PAI, MOT	JAM: 5;4.23 MEG: 7;9.17	65ptJ&MJUL03.cha
6	1.08.03	13.53	JAM, MEG, PAI, MOT	JAM: 5;5.0 MOT: 7;9.24	69ptJ&MAUG03.cha
7	8.08.03	19.03	MEG, JAM, PAI, MOT	JAM: 5;5.7 MEG: 7;10.1	71ptJ&MAUG03.cha
8	9.08.03	2.20	JAM, MEG, VOV, SAR, MOT	JAM: 5;5.8 MEG: 7;10.2	72ptJ&MAUG03.cha
9	15.08.03	19.42	MEG, JAM, PAI, MOT	JAM: 5;5.14 MEG: 7;10.8	74ptJ&MAUG03.cha

Das 9 conversas telefônicas, totalizando 1 hora e 47 minutos, apenas duas não tinham como interlocutor o pai (PAI): conversa número 1 ocorreu entre JAM e um amigo dele, VIN (4.29 minutos) e a conversa número 8 teve como interlocutores o avô brasileiro (VOV) e a prima brasileira (SAR) (2.20 minutos). Os nomes dos arquivos indicam que o português (pt) é a língua principal das conversas, e apenas o material em inglês foi marcado (<material em inglês>[@en]). Os turnos dos interlocutores localizados no Brasil não foram transcritos, no entanto, para indicar o turno destes falantes, o símbolo **www** foi inserido nas linhas principais. No cabeçalho **@Warning** de cada arquivo, consta que a inserção dos turnos pode ser incorreta, mas para fins desta análise específica, isso não influencia os resultados.

4.2. Primeira análise dos arquivos

4.2.1. Construção da linha de comandos

Com o objetivo de investigar e comparar o uso que os dois sujeitos fizeram das duas línguas com a variável 'participante', foram realizadas as seguintes análises: através do KVAL e FREQ foram pedidas listas de frequência das palavras dos dois sujeitos por interlocutor e por língua. Os exemplos das linhas de comando reproduzidas abaixo mostram como foram construídos os comandos:

Exemplo 1:

```
kwal @ +t%add +t*JAM +s"PAI" +d +u | freq +s"<@en>" -s"*pn"+o
```

Dividida em duas partes, essa linha de comando é composta de vários itens que especificam a seleção dos dados a serem analisados e a produção dos resultados. Na primeira parte, KVAL seleciona os enunciados que JAM (+t*JAM) dirigiu ao seu pai (+t%add e +s"PAI") e faz uma análise de todos os nove arquivos selecionados (+u), isto é, não faz a análise por arquivo. Os resultados são despojados de informações extras (+d) e mandados automaticamente para uma segunda análise, realizada por FREQ (| freq). Nessa segunda parte, FREQ realiza uma análise apenas no material em inglês (+s"<@en>") encontrado nos dados resultantes da primeira análise e exclui os nomes próprios marcados (-s"*pn"). As palavras que compõem a lista final aparecem em ordem de frequência (+o).

Exemplo 2:

```
kwal @ +t%add +t*JAM +s"PAI" +d +u | freq -s"<@en>" -s"*pn" +o
```

A única diferença entre essa linha de comando e àquela do primeiro exemplo é o uso de - (sinal de menos) no lugar de + (sinal de mais) imediatamente

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

antes do item `s"<@en>`" na segunda parte da linha. Isso significa que, desta vez, o material em inglês será excluído e a lista será composta somente das palavras em português.

Exemplo 3:

`kwal @ +t%add +t*JAM +s"MOT" +d +u | freq +s"<@en>" -s"*pn" +o`

Essa linha diferencia-se da do exemplo 1, porque desta vez são selecionados os enunciados direcionados à mãe (`+s"MOT"`) em vez de ao pai (`+s"PAI"`).

Exemplo 4.

`kwal @ +t%add +t*MEG +s"PAI" +d +u | freq +s"<@en>" -s" *pn" +o`

Neste exemplo, a análise é realizada com os enunciados da MEG (`+t*MEG`) dirigidos ao pai, contendo material em inglês.

Como ficou claro nos exemplos acima, apenas três itens da linha de comando são trocados: o falante (`+t*xxx`), o interlocutor (`+s"xxx"`) e a língua (`+/-s"<@en>"`). Mantendo a mesma estrutura básica da linha do comando e utilizando a cada análise uma combinação diferente dos três itens mencionados acima, foi realizado um total de 24 análises, cujos resultados produzidos serão discutidos na próxima seção.

4.2.2. Resultados gerais

Para ter uma visão geral do papel que o português e o inglês representam para cada falante em relação aos seus interlocutores, foi feito um levantamento do número total de *tokens* (palavras) de cada lista produzida. Os resultados estão sintetizados nas Tabelas 2, 3 e 4 abaixo.

Na Tabela 2, que mostra os resultados do sujeito JAM, é possível constatar que, em termos gerais, cada língua tem um papel definido, estritamente ligado à condição monolíngüe ou bilíngüe do interlocutor.

Tabela 2: Número total e porcentagem de palavras por língua e por interlocutor do falante JAM, em nove arquivos do LOBILL Corpus.

Língua	Interlocutores									
	PAI		VOV		VIN		MOT		MEG	
Português	1531	91%	18	100%	258	100%	20	4%	23	92%
Inglês	155	9%	0	0%	0	0%	431	96%	2	8%
Total	1686		18		258		451		25	

Embora a variação entre os números totais de palavras dirigidas a cada interlocutor não permita uma comparação totalmente confiável, os resultados mostram que, ao interagir com o avô brasileiro (VOV) e o amigo VIN (nascido na Suécia e bilíngüe em sueco e português), JAM só fez uso do português. Enquanto que nenhuma palavra em inglês foi encontrada nas interações de JAM com esses interlocutores, nas conversas com os demais (pai, mãe e irmã) percebe-se o uso do inglês. Com o pai (PAI), o uso do português predomina, mas quase 10% do total de 1686 palavras foram em inglês, indicando a presença do uso de alternância de código. Com a mãe (MOT), o uso das línguas inverteu-se: de um total de 451 palavras, 96% são palavras em inglês e 4%, em português, mostrando que ao interagir com a mãe o inglês faz o papel da Língua Matriz. Com a irmã, apesar do número reduzido de palavras, os resultados indicam que a língua dominante é o português, mas que existe também a presença de alternância de código. Tudo indica que JAM considera o avô brasileiro e o amigo (bilíngüe em sueco e português) falantes com os quais a estratégia de alternância de código não pode ser utilizada, por eles não dominarem o inglês. O fato de a alternância de código ser empregada na interação com os falantes bilíngües (pai, mãe e irmã) demonstra que JAM é consciente (de forma inconsciente!) da sua escolha em usar uma ou duas línguas com um determinado falante. A natureza estrutural e pragmática do uso da alternância de código por JAM será o assunto da seção 4.3.2.

Ao examinar a Tabela 3, que contém os dados de MEG, vemos que os resultados assemelham-se aos da Tabela 2 (de JAM).

Tabela 3: Número total e porcentagem de palavras por língua e por interlocutor do falante MEG em nove arquivos do LOBILL Corpus.

Língua	Interlocutores									
	PAI		VOV		SAR		MOT		JAM	
Português	3053	96%	23	92%	12	100%	4	3%	4	80%
Inglês	127	4%	2	8%	0	0%	117	97%	1	20%
Total	3180		25		12		121		5	

Com o pai e o irmão, o português predomina (96% e 80%, respectivamente) e com a mãe, o inglês predomina (97%), mas em todos os casos há presença de material em inglês, indicando o uso de alternância de código. Em relação à prima SAR, não constam palavras em inglês, mas o número de palavras é muito reduzido para se chegar a uma conclusão: outras análises de interações entre MEG e SAR precisariam ser realizadas. Com o avô brasileiro, 2 do total de 25 palavras são palavras em inglês, o que indica o uso, embora mínimo, de alternância de código. Na seção 4.2.3., examinaremos essas palavras em mais detalhes.

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

Considerou-se interessante incluir os resultados das análises realizadas em relação ao falante MOT, que estava sempre presente nas conversas telefônicas e interagindo com os dois sujeitos. Como podemos ver na Tabela 4, a mãe utilizou inglês na maior parte do tempo, mas usou o português, especialmente ao interagir com JAM (13%).

Tabela 4: Número total e porcentagem de palavras por língua e por interlocutor do falante MOT em nove arquivos do LOBILL Corpus

Língua	Interlocutores			
	JAM		MEG	
Português	31	13%	2	2%
Inglês	200	87%	86	98%
Total	231		88	

Nos dois casos, há evidência de alternância de código, cuja natureza será investigada e discutida na seção 4.2.3.

Os números totais de palavras na Tabela 4 indicam que a mãe interagiu mais com JAM que com MEG. Ao examinar nas Tabelas 2 e 3 os totais de palavras que os dois sujeitos direcionaram à mãe, também encontramos indicação de que JAM interagiu mais com a mãe (451 palavras) se comparado à irmã (121 palavras). Ao calcular a porcentagem que estes totais representam do total geral de palavras de cada falante (2438 no caso do JAM e 3343 no caso da MEG), constatou-se que o número de palavras dirigidas à mãe por JAM representou 18,4% do total, muito mais do que a porcentagem de 3,6% de MEG. Isto é, durante as conversas telefônicas com o pai (e os outros interlocutores), em comparação a MEG, JAM interagiu muito mais com a mãe. Esta diferença entre os dois sujeitos foi investigada e os dados revelaram uma relação entre os sujeitos e o uso de alternância de código, o que será discutido nas seções a seguir.

4.2.3. Resultados mais específicos

Na discussão dos resultados mostrados acima, ficou evidente que, para os dois sujeitos, o português é a língua utilizada para interagir com falantes considerados 'monolíngües' (em português) e que a alternância de código apenas ocorre na interação com falantes bilíngües (mãe, pai e irmã). Nesta seção, o objetivo é examinar as listas de palavras produzidas pelo FREQ, buscando o que elas podem revelar sobre a natureza da alternância de código, fenômeno detectado pela análise das palavras em termos gerais.

As tabelas apresentadas nessa seção contêm as listas de palavras fornecidas pelas análises realizadas com KWAL e FREQ (veja 4.2.1). Por razões práticas, apenas as primeiras 20 palavras das listas foram selecionadas para análise e discussão. Os interlocutores aos quais não foi dirigida nenhuma palavra em inglês foram excluídos das tabelas.

Tabela 5: Listas de palavras em inglês em ordem de frequência produzidas por JAM e MEG ao interagir com os interlocutores (primeiras 20 palavras).

JAM			MEG			
PAI	MOT	MEG	PAI	VOV	MOT	JAM
10 train	19 i	1 bell	6 e+ mail	1 hey	9 he	1 yes
8 tram	17 to	1 hole	4 library	1 yeah	4 and	
6 **	15 how		4 no		4 mummy	
6 and	15 we		4 triceratops		4 the	
4 bath	14 in		3 and		4 what	
4 but	13 Portuguese		3 dictionary		3 after	
4 live	13 say		3 he		3 her	
4 train+ track	13 want		3 rock		3 him	
3 because	12 do		3 the		3 in	
3 digger	11 "		3 tram		3 says	
3 I	9 is		2 bell		3 to	
3 I'm	9 mummy		2 french		2 does	
3 lightening	9 the		2 oink		2 else	
3	8 a		2 p@l		2 get	
swimming+ po	8 you		2 river		2 just	
ol	6 bath		2 station		2 mother	
2 's	6 don't		2 then		2 said	
2 all	6 it		2 they		2 she	
2 boyfriend	6 which		2		2 something	
2 buffer	5 have		wildlife+ park		2 what's	
2 don't			1 "			
2 England						

* Veja seguinte página para explicações sobre o código " (aspas).

Essas palavras serão examinadas mais adiante nos seus contextos, através do uso do comando KWAL. Entretanto, as listas mostradas nas tabelas acima já indicam vários fatos em relação ao uso de alternância de código pelos dois sujeitos.

Na seção 3, foram apresentados os conceitos de Língua Matriz (a língua que contribui com mais estrutura morfossintática para as sentenças bilíngües) e Língua Encaixada (caracterizada pela contribuição de morfemas de conteúdo como substantivos, verbos, adjetivos e preposições). Ao examinar as listas de palavras dos sujeitos que foram dirigidas ao pai, encontramos uma dominância de substantivos e muito poucos morfemas gramaticais: das primeiras vinte palavras mais freqüentes de cada sujeito, 11 são substantivos. Fica evidente que o inglês exerceu o papel de Língua Encaixada nas interações com o pai. Ao exami-

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

nar as listas de palavras dirigidas à mãe, encontramos uma situação oposta: uma dominância de morfemas gramaticais e apenas 3 substantivos (**Portuguese**, **mummy** e **bath**) no caso de JAM e dois (**mummy** e **mother**) no caso de MEG. Isso confirma que o inglês exerceu o papel de Língua Matriz nas interações entre os sujeitos e sua mãe. A interação entre os irmãos foi pouca, mas as palavras em inglês ficaram reduzidas a substantivos (**bell** e **hole** no caso de JAM) e um marcador discursivo (**yes** no caso da MEG). As duas palavras em inglês dirigidas por MEG ao avô também ficaram na categoria de marcadores discursivos (**hey** e **yeah**), que é freqüentemente citada como um ponto de alternância de código. Conjunções formam outra categoria de palavras que, freqüentemente, marcam o ponto de alternância entre as línguas participantes. Entre as palavras dirigidas ao pai, foram encontrados **and** nas listas dos dois sujeitos e **but** e **because** na lista de JAM. A análise dessas conjunções nos seus contextos maiores (com KWAL) confirmará, ou não, a presença de alternância de código intrassentencial nessas fronteiras.

Outro aspecto que foi revelado pela lista de palavras de JAM em relação à mãe, diz respeito a uma das razões principais da necessidade de sua interação com a mãe durante as conversas com pai. Entre as primeiras dez palavras mais freqüentes, encontram-se **I**, **to**, **how**, **we**, **in**, **Portuguese**, **say**, do e o símbolo “ (aspas) que marca uma referência metalingüística. Ao reorganizar essas palavras, podemos formar a pergunta **How do I/we say “ in Portuguese?**, que, pela freqüência das palavras, deve ter sido utilizada mais de dez vezes por JAM. Isso pode ser confirmado pela análise das palavras em contexto através do KWAL.

A Tabela 6, abaixo, mostra as primeiras vinte palavras mais freqüentes em português que JAM e MEG produziram ao interagir com os pais. Ao comparar as listas dirigidas ao pai, é interessante notar que a maioria é de morfemas gramaticais, sendo eles os mesmos. Isso indica que os dois sujeitos fizeram uso muito semelhante do português ao conversar com o pai ao telefone. É claro que seria necessário comparar as listas inteiras para chegar a alguma conclusão, mas essa amostra indica que, além de atuar como a Língua Matriz, a natureza morfossintática do português empregado pelos dois sujeitos é muito similar.

Tabela 6: Listas de palavras em português em ordem de freqüência produzidas por JAM e MEG ao interagir com os pais (primeiras 20 palavras).

JAM		MEG	
PAI	MOT	PAI	MOT
110 é	2 de	155 e	2 um
84 um	2 o	123 a	1 cachorro
70 não	2 quando	115 que	1 depois
68 que	2 terminou	104 o	
63 eu	1 assistir	81 um	
53 a	1 caiu	81 é	
45 e	1 estava	75 para	
32 o	1 falar	73 eu	
32 tem	1 ligou	57 de	
31 gente	1 noite	57 está	
28 só	1 nome	50 não	
27 mas	1 qual	43 tem	
26 de	1 quê	37 aí	
23 para	1 ratinhos	36 ele	
20 estava	1 scooby+doo	36 tu	
17 no	1 trilho	34 com	
16 também		34 na	
15 aí		33 gente	
15 pai		31 ela	
14 ele		30 estava	

As listas de palavras dirigidas à mãe são muito reduzidas, indicando o uso muito limitado de português ao interagir com esse interlocutor. A natureza desse uso só pode ser examinada pelo estudo das palavras nos seus contextos (veja seção seguinte).

Embora as listas mostradas nas duas tabelas acima contenham palavras fora dos seus contextos, pela discussão nessa seção ficou evidente que uma análise da freqüência das palavras pode ser muito útil para uma investigação preliminar dos papéis que cada língua exerce na interação entre duas crianças bilíngües e seus interlocutores. Pelo estudo da natureza das palavras e suas freqüências foi possível identificar, nas interações, a presença de uma Língua Matriz e uma Língua Encaixada, variando de acordo com o interlocutor. A presença de material de ambas as línguas nas listas de JAM e MEG indica o uso de alternância de código com certos interlocutores, a saber, o pai, a mãe e a irmã. Em relação ao pai, a freqüência de palavras de conteúdo em inglês indica que a alternância de código empregada pelas duas crianças é do tipo intrassentencial, com a inserção de substantivos em inglês na estrutura morfossintática do português (a Língua Matriz). Em relação à mãe, o número reduzido de palavras para estudo

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

dificulta a identificação do tipo de alternância de código utilizado pelos dois sujeitos, o que torna necessário o estudo das palavras em contexto, assunto da próxima seção.

4.3. Segunda análise dos arquivos

4.3.1. Construção das linhas de comandos

Com o objetivo de analisar a natureza da alternância de código utilizada pelos dois sujeitos, foram realizadas seis análises através de KWAL. Como foi identificada a presença significativa de alternância de código nas interações de JAM ou MEG com seus pais, foram construídas as linhas de comandos mostradas a seguir¹²:

1. kwal @ +t%add +t*JAM +s"PAI" +d +u | kwal +s"<@en>" +d
2. kwal @ +t%add +t*JAM +s"MOT" +d +u | kwal -s"<@en>" +d
3. kwal @ +t%add +t*MEG +s"PAI" +d +u | kwal +s"<@en>" +d
4. kwal @ +t%add +t*MEG +s"MOT" +d +u | kwal -s"<@en>" +d
5. kwal @ +t%add +t*MOT +s"JAM" +d +u | kwal -s"<@en>" +d
6. kwal @ +t%add +t*MOT +s"MEG" +d +u | kwal -s"<@en>" +d

A primeira parte de cada linha de comando é basicamente igual àquela explicada em 4.2.1. Apenas trocam-se os códigos dos falantes (**+t*JAM**, **+t*MEG** e **+t*MOT**) e os interlocutores (**+s"MOT"**, **+s"PAI"**, **+s"JAM"**, **+s"MEG"**). Na segunda parte da linha, **FREQ** é substituído por **KWAL** e **+o** por **+d**. Isso significa que no lugar de listas de palavras, os resultados são apresentados em forma de concordâncias, e os enunciados são despojados de informações extras (**+d**). Foi eliminado o item **-s"*pn"** da linha do comando por uma razão importante: ao fazer testes de análises descobriu-se que, ao pedir KWAL para eliminar nomes próprios, qualquer enunciado que continha um nome próprio marcado com **@pn**, era excluído. Isso significa que além de excluir os nomes próprios, KWAL exclui outro material importante para ser analisado. Portanto, esse item não foi inserido na linha de comando.

Nas análises 1 e 3, que tratam de interações entre os dois sujeitos e o pai, foram selecionados os enunciados que contêm material em inglês (**+s"<@en>"**), isto é, que mostram alternância de código para inglês (a Língua Encaixada nesse

¹² As linhas de comando foram numeradas para facilitar a referência textual, os números não fazendo parte das linhas em si.

caso). Nas análises 2, 4, 5, e 6, que tratam de enunciados entre as crianças e a mãe, apenas enunciados que contêm material em português (a Língua Encaixada nesse caso) foram selecionados (-s"<@en>") uma vez que a Língua Matriz é o inglês. Por questões de espaço, foi necessário fazer uma seleção de alguns dos resultados produzidos pelas análises, mas os exemplos apresentados na próxima seção foram considerados representativos do todo.

4.3.2. Resultados das seis análises

Os resultados obtidos das análises através de KWAL serão discutidos de acordo com cada sujeito: primeiro, as três análises relacionadas a JAM (1, 2 e 5) e, depois, em relação a MEG (3, 4 e 6). Para facilitar a discussão dos resultados, as concordâncias foram numeradas (esse número não faz parte da concordância em si).

4.3.2.1. JAM e a alternância de código com a mãe

Foi visto em 4.2.2. que, ao interagir com a mãe, JAM utiliza inglês na maior parte do tempo (431 palavras de um total de 451), mas que existe a presença de material em português, embora pouco (20 palavras). Ao pedir KWAL para localizar esse material, os resultados apresentados foram os seguintes:

- (1) *JAM: <Vincent@pn said which[*]>[@en] Sara@pn caiu .
- (2) *JAM: <he's going to>[@en] assistir Scooby+Doo .
- (3) *JAM: <<but if we>[/] <if we>[/] if Meggie@pn>[@en] terminou, <then you call me when Meggie@pn>[@en] terminou de falar, <okay>[@en]?
- (4) *JAM: qual nome <which[*] we went>[@en]?
- (5) *JAM: <quando o>[/] <daddy he's talking[*] if[/] if Meggie@pn want[*] to talk for>[@en] +/.
- (6) *JAM: <(be)cause yesterday>[@en] quando estava de noite <I didn't have a bath>[@en] .
- (7) *JAM: o quê?
- (8) *JAM: <was saturday which we[/] Vincent@pn, he>[@en] ligou <to here>[@en]?
- (9) *JAM: <how do you say>[@en] ratinhos?
- (10) *JAM: trilho[“].

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

A primeira observação a ser feita é que o tipo de alternância de código utilizado por JAM ao interagir com a mãe é intrassentencial: dentro do mesmo enunciado (que aqui equivale à unidade ‘sentença’), encontramos material das duas línguas. Em quatro dos exemplos, a alternância de código trata da simples inserção de verbos em português (**caiu**, **assistir**, **terminou** e **ligou**) na estrutura do inglês, uma contribuição típica da Língua Encaixada em sentenças bilíngües (Myers-Scotton, 2002:62). Em dois casos, a alternância para o português envolve a conjunção **quando** (**quando** o e **quando estava de noite**); novamente um ponto de alternância freqüentemente encontrado em discurso bilíngüe. O uso pragmático de alternância de código se revela nos últimos dois enunciados: no primeiro, JAM pede a tradução da palavra **ratinhos** e, no segundo, o símbolo [“] indica que ao dizer trilha o sujeito está fazendo uma referência metalingüística à palavra. Embora cada exemplo de alternância de código mostrado nos resultados mereça uma discussão detalhada à luz dos modelos desenvolvidos por teóricos como Myers-Scotton (2002), aqui fica limitado a uma observação geral em relação à alternância de código utilizada por JAM ao interagir com a mãe: pelas concordâncias produzidas por KWAL, é possível constatar que o inglês exerce o papel da Língua Matriz e que a natureza das alternâncias para português é típica da contribuição da Língua Encaixada.

Ao examinar as concordâncias que mostram o material em português produzido pela mãe ao interagir com JAM (veja análise 5), percebemos logo que a alternância de código empregada é de um tipo bem específico. Como pode ser visto nos resultados apresentados abaixo, quase todos os casos tratam de referências metalingüísticas, indicadas pelo símbolo [“].

- (11) *MOT: **bonde**[“] .
- (12) *MOT: **bonde**[“] .
- (13) *MOT: **bonde**[“] .
- (14) *MOT: **no** +...
- (15) *MOT: **sshh** .
- (16) *MOT: <**buraco grande**>[“] .
- (17) *MOT: <**bucket**[“]>[“en], **balde**, **balde** .
- (18) *MOT: **madeira**[“] .
- (19) *MOT: **proteger**[“] .
- (20) *MOT: <**assistindo um filme**>[“] .
- (21) *MOT: <**I think it's**>[“en] **trem+a+vapor**[“] .
- (22) *MOT: <**eu te amo**>[“] .
- (23) *MOT: **rato**[“] .
- (24) *MOT: **ahh**.

- (25) *MOT: sair["].
- (26) *MOT: frio["].
- (27) *MOT: trem+a+vapor["] vapor["].
- (28) *MOT: trilho["].
- (29) *MOT: trilho["].
- (30) *MOT: banheira["].
- (31) *MOT: banheira["].
- (32) *MOT: <tomar um banho>["].

O uso desse símbolo nas transcrições para marcar referências metalingüísticas foi fundamental. Se não houvesse esse tipo de marcação, ao ser apresentado com essa lista de concordâncias, o pesquisador poderia chegar a conclusões errôneas sobre o uso do português pelo falante. Fica evidente que as palavras e frases em português produzidas pela mãe são respostas a vários pedidos de JAM do tipo **How do I say XXX in Portuguese?**, que foi constatado em 4.2.3. Como será visto na próxima seção, esse uso pragmático do português não foi encontrado nos resultados da análise 6 (enunciados da mãe dirigidos à filha), o que indica um contraste na necessidade de cada criança de fazer esse tipo de alternância de código.

4.3.2.2. JAM e a alternância de código com o pai

Ao interagir com o pai ao telefone, constatou-se que 91% de um total de 1686 palavras foram de material em português. Após examinar a natureza das palavras encontradas nas listas produzidas por FREQ, conclui-se que, enquanto com a mãe inglês era a Língua Matriz, com o pai português exercia esse papel. E inglês foi claramente identificado como a Língua Encaixada. Ao pedir a seleção de todos os enunciados contendo material em inglês dirigidos ao pai (análise 1), uma lista de 87 concordâncias foi apresentada. Embora não seja possível reproduzir e analisar cada concordância, foram selecionadas 22 que mostram a natureza variada das inserções de inglês encontrada nos dados. Os exemplos foram agrupados para melhor exemplificar as categorias do material inserido e o primeiro grupo representa os casos mais comuns: a inserção de palavras de conteúdo, especialmente substantivos:

- (33) *JAM: mas só eu sou muito <lucky>[@en] que eu peguei um[/] um <train+track>[@en] e a gente comp(rou) +/.
- (34) *JAM: e um <cutting+grass+machine>[@en] também .
- (35) *JAM: eu tenho um <digger>[@en], um <digger>[@en] também e um <train>[@en] também .

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

- (36) *JAM: aí a gente ficou # # <five>[@en] dias lá .
(37) *JAM: não tinha <lightening>[@en] não, só trovão .
(38) *JAM: e eu[/] eu <found>[@en] um buraco grande, muito grande.
(39) *JAM: era <shiny, yeah>[@en] .
(40) *JAM: não, <a gente>[///] eles foi[*] pra[: para a] <Grandad@pn
's house>[@en] .
(41) *JAM: e tem um <swimming+pool>[@en] lá que a gente +/.
(42) *JAM: Mas o <North@pn>[@en] é o <England>[@en] também.

É importante lembrar que, ao conversar com o pai pelo telefone, JAM encontra-se na Inglaterra de férias. Exposto a novas experiências e novos contextos, muitas vezes as palavras em inglês inseridas no português coincidem com objetos ou conceitos que não faziam parte da sua realidade no contexto brasileiro. É o caso de palavras como **train+track**, **digger** e **cutting+grass+machine**. E, embora haja evidência de que JAM pedia à mãe a tradução para português de várias palavras em inglês, enquanto falava com o pai, fica evidente pelos exemplos que o sujeito não fazia questão de eliminar totalmente da interação o uso de inglês, especialmente em relação às palavras de conteúdo.

Em alguns casos, JAM inseriu mais do que palavras isoladas do inglês, como mostram os dois exemplos seguintes:

- (43) *JAM: <Uncle+William@pn>[@en] (es)tá <sleeping in another bed>[@en] .
(44) *JAM: não, <ele não <live>[@en]>[///] <she don't live with he(r)[/] her boyfriend>[@en] .

No primeiro exemplo, JAM está explicando ao pai que o tio veio passar alguns dias e que não vai dormir na sala (como o pai imaginava). Tendo já conversado sobre esse assunto com a mãe e a avó em inglês¹³, evidentemente ficou mais fácil mudar para inglês. No segundo exemplo, JAM está explicando sobre o namorado da tia, que não mora com ela (e a avó inglesa). Embora iniciando em português, após inserir **live**, JAM faz uma reformulação que envolve uma mudança total para inglês. Novamente parece que, nesse caso, JAM teve mais facilidade de se expressar em inglês e considerou (consciente ou inconscientemente) que seu pai não ia ter dificuldades em entendê-lo.

¹³ As anotações de campo contêm essa informação.

Em relação à inserção de conjunções em inglês, outro grupo que se destacou nos enunciados bilíngües, as concordâncias revelaram um fato interessante. Pela transcrição dos exemplos mostrados abaixo, é possível notar, que logo após a inserção de uma conjunção em inglês, JAM procura reformular seu discurso (indicado pelo símbolo [///]) e insere o equivalente em português:

- (45) *JAM: <but>[@en][///] mas, só que <o deles>[//] eles tem um debaixo e um de cima .
- (46) *JAM: não, <bec(ause)>[@en][///] porque <não tem>[/] não tem bateria[*] não.
- (47) *JAM: <and>[@en][///] e cavou muito que o[///] peguei o[///] a gente trouxe o +...

Isso também ocorre com frequência quando JAM confirma algo dito pelo pai, como mostra o seguinte exemplo, no qual o sujeito diz *yes* e imediatamente depois é:

- (48) *JAM: <yes>[@en][///] é, mas quando eu fui pro[: para o][/] pro[: para o] outro[///] pra[: para a] areia <I can very>[@en] muito muito.

Embora seja evidente que nesses exemplos a palavra em inglês é ativada primeiro (está na ‘ponta da língua’), JAM a substitui pelo equivalente em português, demonstrando uma ‘preocupação’ em falar a língua do pai. Até certo ponto essa atitude parece contradizer as ocasiões em que JAM demonstrou uma ‘falta’ de preocupação em falar apenas português (veja exemplos 33 a 42 acima). Entretanto, os casos das conjunções citadas acima parecem se enquadrar na categoria de ‘lapsos’, em que os equivalentes são fácil e rapidamente acessados, enquanto a inserção de, por exemplo, um substantivo em inglês, pode ocorrer por falta de um equivalente adequado em português ou por a palavra ser, momentaneamente, de mais difícil acesso que a equivalente em inglês. De fato, por JAM estar submerso num contexto inglês, as palavras inglesas estão mais ativadas que as portuguesas e um atraso no acesso a estas, portanto, pode ser natural.

As concordâncias mostradas a seguir revelam um uso bem específico de alternância de código que envolve referências metalingüísticas:

- (49) *JAM: tu sabe que é um <tram[“]>[@en]?
- (50) *JAM: é um banheiro mas não é um chuveiro, é um <bath, bath>[@en].

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

(51) *JAM: tu sabe que é um <buffer[“]>[@en]?

Ao fazer a alternância de código para inglês de algumas palavras, JAM demonstra certa consciência de que o pai talvez não conhecesse essas palavras e procura verificar e clarificar seus significados. Ao conferir a lista das palavras em português produzidas pela mãe (veja exemplos 11 a 32), fica evidente que, nesses casos, JAM acabou pedindo a tradução à mãe. Nos primeiros dois exemplos, a mãe forneceu **bonde** e **banheira**, mas não sabia a tradução de **buffer** (‘pára-choque’), o que JAM tentou explicar para o pai, recorrendo ao inglês novamente:

(52) *JAM: é um negócio que[/] que <keep on rails>[@en].

Precisando pedir ajuda à mãe com a tradução da palavra **rail** (**trilho** aparece duas vezes, marcado como referência metalingüística na lista de concordâncias da mãe), JAM consegue fazer com que o pai entenda o que está querendo dizer.

Pelos exemplos apresentados, representativos da lista de concordâncias produzida por KWAL, em quase todos os casos, a alternância de código empregada por JAM foi do tipo intrassentencial, com a inserção de morfemas de conteúdo. Entretanto, houve uma ocorrência de uma alternância de código intersentencial mostrada na seguinte seqüência:

(53) *JAM: eu disse para mamãe no com(putador)[///] <library>[@en] (es)tá fazendo assim ó disse +”/.

JAM: +” <mummy mummy would you write on the computer <which[] saying>[/] write which[*] I’m missing Daddy and I want to see him and I’m[/] I’m[*] love him too>[@en] +”.

Aqui JAM está explicando ao pai o que falou para a mãe quando ela estava na biblioteca escrevendo em e-mail para o pai. Os símbolos +” e +”. deixam claro que se trata de discurso direto e, nesse caso, JAM escolheu manter a língua original do pedido (inglês).

Embora fossem apresentadas apenas 22 das 87 concordâncias, os exemplos escolhidos revelaram bastante sobre a natureza estrutural e pragmática da alternância de código empregada pelo falante JAM ao interagir com a mãe e o pai. Na próxima seção, examinaremos as concordâncias produzidas pelas análises 3, 4 e 6 que tratam da interação de MEG com seus pais.

4.3.2.3. MEG e a alternância de código com a mãe

Pelas análises de *FREQ*, de um total de 121 palavras direcionadas à mãe, apenas 4 foram em português: **um**, que aparece duas vezes, **cachorro** e **depois**. Ao pedir as concordâncias desse material em português, os resultados foram as seguintes:

- (54) *MEG: <and as well <when he>[//] he said when he comes in he says>[@en] cachorro[""] <and he says she starts to mi+mi+mi[""]>[@en] .
- (55) *MEG: depois um[/] um <pound>[@en]?

Na primeira concordância, MEG está contando à mãe o que o pai disse em relação ao preá dela: quando ele chega do trabalho e abre a porta, ele chama **cachorro** (uma brincadeira de nome já estabelecida) e o preá responde com **mi+mi+mi**. Além de ser utilizado para marcar uma referência lingüística, o símbolo [""] também marca discurso direto quando se trata de uma palavra isolada. Portanto, a alternância de código empregada aqui com a inserção de **cachorro** tem a função específica de citar discurso direto. A segunda concordância contém as outras três palavras em português dirigidas à mãe. Por coincidência, exatamente a mesma frase aparece na lista de concordâncias direcionadas ao pai. Através de outro comando do CLAN, COMBO, que busca grupos de palavras, (nesse caso, **combo @ t*MEG +s"depois^um"**), foi possível localizar com facilidade o arquivo específico dessa concordância e depois examiná-la em seu contexto maior (isto é ler a parte pertinente da transcrição). Ao fazer isso, constatou-se que, na realidade, esse enunciado foi direcionado tanto ao pai quanto à mãe: MEG está contando ao pai sobre a compra de um conjunto de dicionários cujo preço tinha sido reduzido de trinta libras (pounds) para sete libras e **depois um pound**. Ao dizer esse último valor para o pai, MEG olha para a mãe procurando confirmação do valor e esta responde **yes**. Portanto, podemos considerar que a mãe foi um interlocutor indireto e que a decisão de MEG de utilizar português como Língua Matriz foi baseada no seu interlocutor principal (o pai).

Em termos de números totais de palavras, MEG interagiu muito menos com sua mãe (121) quando comparado com JAM (451), o que dificulta uma comparação confiável da extensão do uso de alternância de código pelas duas crianças. Ao examinar os resultados de *KWAL* em relação ao português utilizado pela mãe ao interagir com a filha, constatou-se uma única concordância, mostrada a seguir:

(56) *MOT: bem rapidinha Meggie@pn .

Não houve ocorrências de referências metalingüísticas em português, como foi visto no caso da interação entre a mãe e JAM. Isso indica que MEG não pediu a tradução para o português de palavras inglesas. Talvez ela considerasse desnecessário ter recurso à tradução (porque o pai compreendia as palavras em inglês utilizadas) ou talvez ela não precisasse utilizar essa estratégia (por não ter dificuldade em se expressar em português). Na próxima seção, a análise das concordâncias contendo material em inglês dirigidas ao pai, revelará mais sobre as diferenças no uso de alternância de código pelos dois sujeitos.

4.3.2.4. MEG e a alternância de código com o pai

Pelos resultados de FREQ foi visto que, ao interagir com o pai, MEG utilizou 127 palavras em inglês, de um total geral de 3053 palavras (representando 4%). Ao examinar as primeiras vinte palavras da lista produzida por FREQ, constatou-se a presença de muitos substantivos e poucas palavras gramaticais. Ao pedir a lista de concordâncias do material em inglês através de KWAL, os resultados mostram que, na grande maioria dos casos, a alternância de código trata-se da inserção de palavras inglesas isoladas na estrutura do português. De um total de 71 concordâncias, 59 contêm uma ou, às vezes, duas palavras isoladas e dessas 59 concordâncias, 56 contêm substantivos. Alguns exemplos típicos podem ser vistos a seguir:

- (57) *MEG: e também eu[/] eu fui com umas botas que eu comprei no <car+boot+sale>[@en] e fui lá no[/] no <river>[@en] num cantinho que tem água e fiquei andando .
- (58) *MEG: hoje <de man(hã)>[/] de[/] de tarde quando a mamãe foi pro[: para o] <library>[@en] para mandar um <e+mail>[@en] para tu, sabe o que que ela trouxe?
- (59) *MEG: também eu ganhei o sinal que tem o <bell>[@en] que tem no <tram>[@en] .
- (60) *MEG: um, comprou por noventa e nove <p@l[/] pence>[@en] .
- (61) *MEG: ah, e hoje na praia a gente comprou um <rock>[@en] .
- (62) *MEG: quando eu estava lá no <station, train+station>[@en] +...
- (63) *MEG: aí sabe que[/] que o tiranosauo+rex estava comendo o <triceratops>[@en] .
- (64) *MEG: tinha tudo preto, e estavam[/] quando James@pn corria, eles corriam embaixo desta[/] deste bush>[@en] .

- (65) *MEG: ei pai, quando o James@pn estava falando de cair da cama quando eu estava deitando para dormir eu e a mamãe na cama dela, o[/] o Jame(s)@pn caiu da cama e caiu de[/] de costas na[//] no <bricks>[@en] .
- (66) *MEG: era erm aquele que passa no filme <que tem aquela>[///] que parece <triceratops>[@en] só que não é afiado a[/] a[///] /] <the horn>[@en] .

A ausência de símbolos que marcam hesitações (#) e marcadores de discurso como **erm**, indica que, pelo menos nestes casos, MEG empregou esse tipo de alternância de código com a maior naturalidade. Ao examinar a natureza das palavras, encontramos algumas que são estritamente ligadas ao contexto experiencial em que MEG se encontrava: **car+boot+sale**, **pence**, **rock**, **river**, **library**, **bell**, **tram**, **train+station**, **bush**. É importante salientar que tal afirmação é possível devido ao fato de a pesquisadora encontrar-se na posição de um observador participante, compartilhando as experiências dos sujeitos. Para outro pesquisador poder fazer essa mesma afirmação seria necessário incluir esse tipo de informação extralingüística numa linha dependente.

Em relação à natureza estrutural da alternância de código utilizada por MEG, os dados nos revelam uma coisa interessante: com exceção do último exemplo, o substantivo em inglês sempre vem acompanhado pelo artigo em português e fica evidente que a preferência do falante é utilizar o artigo masculino: **no car+boot+sale**, **no library**, **o bell**, **um rock**. Nos exemplos (64) e (65), essa preferência fica ainda mais clara: apesar de inicialmente utilizar o artigo feminino, MEG imediatamente faz uma reformulação e o substitui pelo artigo masculino (**desta[//] deste bush** e **na[//] no bricks**). Parece que a maneira que MEG encontrou de lidar com a falta de gênero dos substantivos ingleses é de generalizar, inserindo o artigo masculino mesmo no caso de palavras cujas traduções em português requereriam o artigo feminino, como por exemplo, **library** – ‘biblioteca’ e **station** – ‘estação’. Ao re-examinar as concordâncias referentes a JAM, foi encontrado o mesmo fenômeno. Em relação à exceção (o último exemplo da lista acima), MEG utiliza o artigo feminino, mas, após repetir uma vez, insere o substantivo inglês junto com o artigo definido em inglês: **...afiado a[/] a[///] <the horn>[@en]**. Talvez a inserção do **the** fosse uma estratégia para evitar um encontro de sons que soasse meio desajeitado (**a** com **horn**) ou talvez MEG quisesse evitar uma possível ambigüidade de sentido: sem a inserção do **the**, **a**, seguido por **horn**, podia ser entendido tanto como ‘um’ (o artigo indefinido inglês) quanto ‘a’ (artigo definido em português). Pelo enunciado, fica claro que o sentido desejado é este último e, portanto, talvez para

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

evitar a construção de um enunciado potencialmente incoerente, MEG decide utilizar **the**.

Embora a grande maioria das palavras inglesas inseridas de forma isolada fosse de substantivos, as concordâncias abaixo mostram outros tipos de palavras isoladas inseridas (no máximo duas ocorrências de cada):

- (67) *MEG: <then>[@en] sete .
- (68) *MEG: é, dois dias, <no>[@en] duas noites .
- (69) *MEG: <yeah>[@en], uma grade .
- (70) *MEG: <fine>[@en] .
- (71) *MEG: tinha dois <little>[@en] ratinhos ali .
- (72) *MEG: ela nas escadas <already>[@en] .

Dessas poucas ocorrências, os últimos dois exemplos indicam um uso mais marcado de alternância de código: a inserção de **little** e **already** parece ter a função de enfatizar o que o falante está contando. Entretanto, como já foi dito, as ocorrências de concordâncias contendo esses tipos de palavras foram tão poucas que quaisquer conclusões seriam especulações.

As concordâncias que evidenciaram a inserção de mais que palavras isoladas podem ser divididas em dois grupos. Em um grupo, encontram-se 4 concordâncias que mostram a inserção de frases curtas em inglês no português, reproduzidas a seguir:

- (73) *MEG: mas o James@pn, ele é maluco, se pulou no mar, <to the sea>[@en].
- (74) *MEG: aí a gente vai vir essa semana que vem (por)que o <library>[@en] está fechado <friday, saturday and sunday>[@en].
- (75) *MEG: <then it was>[@en] +...
- (76) *MEG: aí a <Janet@pn>[@en] conseguiu ver <but>[@en][///] mas o <Grandad@pn didn't know>[@en] .

Embora o pesquisador possa apenas especular sobre as razões por que MEG inseriu essas frases curtas, em termos estruturais o que pode ser observado em três das concordâncias é que há coerência entre as estruturas da Língua Matriz e da Língua Encaixada. Novamente é importante salientar que não há evidência suficiente para fazer afirmações concretas em relação a essa coerência entre as línguas. Entretanto, a falta de ocorrências desse tipo de inserção (frases curtas) quando comparada ao número de ocorrências da inserção de palavras isoladas, reafirma

que, na maior parte do tempo, a alternância de código empregada por MEG foi de um tipo bem específico e restrito (de palavras isoladas).

Outro grupo pequeno de 6 concordâncias evidenciou o uso pragmático da alternância de código para citar o discurso direto. As primeiras quatro citações da lista mostrada a seguir tratam de falas de personagens de um seriado inglês *Fawlty Towers*, que as crianças assistiam em vídeo (repetidas vezes) na Inglaterra e que já tinha passado no canal EuroChannel no Brasil (o pai conhecia o seriado). Nos primeiros dois casos, são mostradas em parênteses as falas anteriores à citação:

- (77) ***MEG:** e tu já viu no <Fawly+Towers>[@en] o Manuel@pn?
***MEG:** +” <two dead pigeons in the water+tank>[@en] +”.
- (78) ***MEG:** aí e[/] e na linguagem do Manuel@pn, ele pensa que é porcos
e
[/] <e ele disse>[///] <and he said>[@en] +”/.)
***MEG:** +” <how did they get up there>[@en] +”?
- (79) ***MEG:** +” <they flew>[@en] +”, disse o Basil@pn .
- (80) ***MEG:** +” <oink oink>[@en] +”, e tudo assim .
- (81) ***MEG:** aí o[/] <Jake@pn and Max@pn>[@en] também vieram mas
tu sabe como <Jake@pn>[@en] é +”/.)
***MEG:** +” <why doesn’t he eat his tail>[@en] +”.
- (82) ***MEG:** +” <what on earth is he thinking>[@en] +”.

As últimas duas concordâncias são citações das falas do primo inglês, JAK, e da mãe. Embora esse tipo de alternância de código seja utilizado mais por MEG, quando comparado ao irmão (veja exemplo (53)), sua presença nas interações com pai é significativa: indica a consciência plena da condição bilíngüe do pai. Entretanto, as concordâncias revelam que, apesar de saber que a alternância de código para inglês não cria muito problema para o pai, ambas as crianças mostram o uso de certas estratégias para evitar o seu uso excessivo. Enquanto JAM recorre bastante à ajuda da mãe para traduzir palavras em inglês para o português (veja exemplos (11) a (32)), MEG usa a estratégia de reformulação. A seguir são reproduzidas as cinco concordâncias que revelam essa estratégia:

- (83) ***MEG:** e estava colecionado <err <ca(rds)>[@en]>[///] err tiquete
de trem .
- (84) ***MEG:** tinha o[/] o <whale>[@en][///] a baleia .
- (85) ***MEG:** <do outro <si(de)>[@en]>[///] do outro lado .

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilíngüe Infantil*.

- (86) *MEG: estava lá no <North@pn>[@en] e <mum(my)>[@en][//]
mamãe estava no <conference>[@en] ela, eu[/] eu saí com
a[/] a <Janet@pn>[@en] e o <Grandad@pn>.
- (87) *MEG: era um banco # # que[/] que estava na[/] em frente do
<river>[@en] e o ratinho <jump(ed)>[@en][//] err pu-
lou na água, aí[/] aí eu vi aí uma coisa tinha pulada na água .

As palavras para as quais MEG forneceu a tradução em português (**cards, whale, side, mummy, jump(ed)**) não são do tipo específico que justifica a necessidade de uma tradução. Portanto, essa estratégia de reformulação indica que MEG, talvez por empatia com o pai, estava disposta a limitar o uso de inglês quando possível, isto é, quando o equivalente em português fosse facilmente recuperado. No caso de palavras mais específicas ou mais recorrentes no ambiente de convívio imediato (exemplos (57) a (66)), essa estratégia foi raramente utilizada: o desconhecimento do equivalente em português, ou o fato de ser de difícil acesso no momento da fala, levou à inserção liberada de substantivos em inglês.

5. Conclusão

A análise das listas de palavras e das concordâncias permitiu identificar a natureza da alternância de código empregada por MEG e JAM ao interagir com o pai pelo telefone. Os dados indicam que, no caso de MEG, seu uso é bastante específico, limitado na maioria dos casos à inserção de substantivos ingleses isolados e a algumas ocorrências de citações de discurso direto. Pela análise dos dados relacionados à criança JAM, constatou-se que seu uso da alternância de código é muito parecido. Entretanto, ao examinar as listas de palavras e as concordâncias que resultaram das análises da interação entre mãe e filho, ficou evidente uma diferença notável entre os dois sujeitos. Embora utilizando alternância de código com o pai, JAM também recorria muito à ajuda da mãe em busca de traduções para o português de palavras inglesas. Essa diferença pode ser indicativa de uma relação entre o uso de alternância de código e o desenvolvimento lingüístico das duas crianças. Pelas evidências encontradas nos dados, MEG (7;10 na época da gravação das conversas) mostra uma certa consistência e segurança no uso de alternância de código, utilizando quando e quanto ela julga necessário na interação com seu pai, não achando necessário recorrer à ajuda da mãe. Em relação a JAM (5;6), por outro lado, os dados mostram uma certa incoerência no uso de alternância de código: ora insere palavras em inglês ora pede a tradução para a mãe. Isso

Crop, 10, 2004

talvez indique uma imaturidade no que diz respeito à consciência de sua condição bilíngüe e da de seu pai. É evidente que JAM sabe que pode empregar alternância de código ao interagir com o pai, mas parece inseguro em relação a 'quando' e 'quanto'. Talvez, por causa do seu estágio de desenvolvimento lingüístico, falte para JAM o conhecimento necessário para estabelecer os parâmetros que regem o uso de alternância de código com este determinado falante.

6. Considerações finais

Os resultados das análises apresentados e discutidos neste artigo deixa evidente o valor de poder realizar tais análises eletronicamente. Em questão de segundos, o pesquisador pode dispor de dados confiáveis, os quais podem ser examinados tanto de forma quantitativa quanto qualitativa. O exame das listas de palavras e das listas de concordâncias dos dados de JAM e MEG permitiu detectar várias facetas do fenômeno de alternância de código e seu uso pelas duas crianças. Além dos aspectos estruturais e pragmáticos revelados pelos dados, foram levantadas algumas hipóteses que merecem ser investigadas nos demais dados do LOBILL corpus. Através do CLAN, esse tipo de investigação torna-se fácil e rápida, e também permite a comparação entre dados de corpora diferentes. É evidente que essa metodologia eletrônica pode contribuir muito para a investigação de vários fenômenos lingüísticos e não somente a da alternância do código. As opções que o CHILDES pode oferecer em termos de construção e exploração de um corpus oral ou escrito são amplas. Cabe ao pesquisador delimitar o tipo de corpus que precisa construir para a realização dos seus objetivos e transcrevê-lo com suas especificidades.

Referências Bibliográficas

- BAKER, Colin. *The care and education of young bilinguals: an introduction for professionals*. Clevedon: Multilingual Matters, 2000.
- _____. *Foundations of bilingual education and bilingualism*. 3 ed. Clevedon: Multilingual Matters, 2001.
- BERBER-SARDINHA, Tony. *Usando WordSmith Tools na investigação da linguagem*. [online]. 1999a. [cited 05.08.2002]. <http://ael.pucsp.br/DirectPapers/40.pdf>
- BRITO DE MELLO, Heloísa Augusta. *O falar bilíngüe*. Goiânia: Ed. da EFG, 1999.
- MACSWAN, Jeff. *A minimalist approach to intrasentential Code Switching: Spanish-Nahuatl Bilingualism in Central Mexico*. [online]. 1997. [cited 23.03.2003]. (dissertação de doutorado) <http://cpnet.mit.edu/MITECS/Entry/gumperz.html>.
- MACWHINNEY, Brian. *Manual de CHAT*. [online]. 2003. [cited 23.03.03]. <http://childes.psy.cmu.edu/manuals/CHAT.pdf>

SAMPAIO, Catherine Anne Lonngren. *A Investigação de Alternância de Código em um Corpus Eletrônico de Linguagem Bilingüe Infantil*.

_____. *Manual de CLAN*. [online]. 2003. [cited 23.03.03]. <http://childes.psy.cmu.edu/manuals/CLAN.pdf>

_____. *Out of the Baby Book and Into the Computer: Child Language Research Comes of Age* [online]. 2003. [cited 23.03.03]. <http://childes.psy.cmu.edu>

MYERS-SCOTTON, Carol. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press, 2002.

POPLACK, Shana. 'Sometimes I'll Start a Sentence in Spanish y termino español': Toward a Typology of Code-Switching. *Linguistics* 18, 1980: 581-618.

_____. *Noções de compilação de corpus*. [online]. 1999b. [cited 05.08.2002].

Lingüística de Corpus e Análise Literária: uma Aplicação a *Dubliners* de Joyce

Lourdes Bernardes Gonçalves*

Resumo: *A Lingüística de Corpus sugere uma gama de possibilidades de pesquisa com a língua nas mais diversas áreas. Ferramentas computacionais estão sendo constantemente aperfeiçoadas para que se possa manipular facilmente uma grande quantidade de textos e obter informações estatísticas de vários tipos, que enriquecem a pesquisa. Análises literárias, que lidam com a exploração de textos, podem se beneficiar grandemente com a utilização de uma metodologia que forneça informações que passariam despercebidas com outros procedimentos de investigação. Para ilustrar como a Lingüística de Corpus pode contribuir para a pesquisa literária utilizaremos como Corpus de Estudo a coletânea de contos **Dubliners**, de James Joyce.*

Palavras-chave: *Lingüística de Corpus; análise literária; Dubliners.*

Abstract: *Corpus Linguistics enables a large variety of research possibilities in language in many different areas. Softwares are constantly being improved in order to easily handle large quantities of text and provide all kinds of statistical data, which will add relevant information to the research. Literary analyses, dealing with text investigation, can indeed benefit from the use of a methodology which provides information that otherwise would have gone unnoticed. We intend to illustrate how Corpus Linguistics can contribute to literary*

* Doutoranda do Programa de Estudos Lingüísticos e Literários em Inglês da Universidade de São Paulo, com bolsa PQI-CAPES da Universidade Federal do Ceará.

GONÇALVES, Lourdes Bernardes. *Linguística de Corpus e Análise Literária: uma Aplicação a Dubliners de Joyce*.

research by using the collection of short stories Dubliners, by James Joyce as our Study Corpus.

Keywords: *Corpus Linguistics; literary analysis; Dubliners.*

1. Introdução

A utilidade e eficiência de um corpus eletrônico para o estudo de textos já está estabelecida em várias áreas de pesquisa com a língua, nas suas mais vastas manifestações. Pesquisas nas áreas de terminologia e lexicologia têm usufruído grandemente dos recursos que essa metodologia fornece. Com a possibilidade de armazenar e manipular eletronicamente os textos, uma quantidade consideravelmente maior de textos pode ser compilada e arquivada de modo a servir como material de consulta. Softwares especiais, que oferecem ferramentas de análise de corpus, vêm sendo desenvolvidos para facilitar a manipulação de grandes corpora, o que é essencial para que o pesquisador possa obter facilmente as informações de que precisa. Palavras podem ser selecionadas dentre milhares sem que se precise ler todos os textos com que se está trabalhando, ocorrências de palavras são fornecidas imediatamente, em termos de percentagens e em linhas de concordância. O acesso a exemplos de uso real e atual da língua em números estatisticamente relevantes é inestimável para o pesquisador na elaboração de dicionários, por exemplo.

É uma metodologia poderosa por conseguir trazer à luz fatos que haviam até agora permanecido escondidos ou passaram despercebidos. Lynne Bowker (2001: 346) ressalta a conveniência de se usar o estudo de corpora juntamente com outras metodologias mais tradicionais: “A abordagem de pesquisa baseada em corpus pode ser considerada como complementar a alguns elementos de abordagens mais tradicionais; por exemplo, a informação fornecida por especialistas pode ser explorada de modo mais abrangente num corpus, a intuição pode ser verificada usando um corpus, ou termos encontrados em dicionários ou textos paralelos podem ser usados como pontos de acesso para um corpus.”^{1,2}

¹ “The corpus-based approach can be seen as complementary to some elements of more traditional approaches; for example, information provided by subject field experts can be explored more fully in a corpus, intuition can be verified using a corpus, or terms found in dictionaries or parallel texts can be used as access into a corpus.” (BOWKER, 2001: 346)

² Salvo especificado ao contrário, as traduções aqui presentes, inclusive esta, são da autora deste trabalho.

Em pesquisas literárias, em que o texto é o elemento central das análises, corpora já vêm sendo utilizados desde o século XVII e até concordâncias foram pesquisadas e listadas manualmente, como atesta Kennedy (1998: 13-15), porém com muito sacrifício, consumo de tempo e inevitáveis introduções de erros. Assim, a Lingüística de Corpus se evidencia imediatamente como uma metodologia extremamente facilitadora do trabalho de pesquisadores e críticos literários, pois vai além de apenas listar palavras, fornecendo entre outras possibilidades palavras-chave dos textos estudados e linhas de concordâncias de diferentes tamanhos, dependendo do objetivo da pesquisa.

Neste trabalho pretendemos ilustrar como a Lingüística de Corpus pode fornecer elementos para a análise literária. Mostraremos como essa metodologia pode apontar características do texto literário que passariam provavelmente despercebidas com outros procedimentos de investigação. Para tanto, montaremos um Corpus de Referência que será utilizado juntamente com o Corpus de Estudo. Este último será a obra *Dubliners*, uma coletânea de quinze contos de James Joyce publicada em 1914. O Corpus de Referência será compilado de textos do mesmo gênero (contos), de autores contemporâneos a Joyce. Uma contribuição significativa da Lingüística de Corpus é permitir a observação de elementos curiosamente ausentes no Corpus de Estudo, por meio da identificação de palavras encontradas apenas no Corpus de Referência. Podemos também ter acesso a uma reunião de aspectos diluídos ao longo da obra, que dificilmente seriam percebidos com metodologias mais tradicionais.

2. Considerações gerais

James Joyce (1882–1941) é um escritor cuja individualidade não permite classificá-lo em um grupo de escritores. Extremamente criativo e inovador, preocupou-se profundamente com os problemas da Irlanda, não apenas no aspecto político, mas social e mesmo pessoal de seu povo. A angústia de reconhecer a opressão que em seu país se instalara em todos os níveis, macroscopicamente pela dominação inglesa e microscopicamente no coração e atitudes de cada irlandês, fez com que o autor escolhesse como tema geral dos contos a paralisia e estagnação, que se confrontam com a idéia do escapar para a liberdade, que se realizam no contraste paralisia versus fuga. Joyce procurou mostrar que não só a Inglaterra, mas a própria Irlanda exerce sobre seus filhos um efeito paralisante, quer pela pressão da religião, como em “The Sisters”, quer como resultado do consumo exagerado de álcool, como em “Counterparts”, ou ainda devido a pressões de rígidos códigos sociais, como em “The Boarding House”.

GONÇALVES, Lourdes Bernardes. *Linguística de Corpus e Análise Literária: uma Aplicação a Dubliners de Joyce*.

Quando escreveu *Dubliners*, Joyce não usou a fragmentação do discurso ou explorou a técnica de fluxo da consciência. Talvez por ter sido escrito anteriormente a *Ulysses* (1922) ou *Finnegans Wake* (1939), mas provavelmente por querer atingir o homem comum de Dublin, exibir um espelho diante dele e fazê-lo reconhecer-se nas múltiplas descrições. Patrick A. McCarthy (1998) reproduz um trecho de uma carta de Joyce a Grant Richards, que publicou *Dubliners*, mostrando justamente sua intenção e escolha de estilo ao escrever a obra: “Minha intenção foi escrever um capítulo da história moral do meu país e escolhi Dublin para a cena porque a cidade me parecia o centro da paralisia. [...] Escrevi a maior parte num estilo de mesquinhez escrupulosa e com a convicção de que um homem muito corajoso é aquele que se atreve a alterar, ou ainda mais deformar, a apresentação do que quer que ele tenha visto ou ouvido.”³

Com essas características em mente, começamos a estabelecer nossas hipóteses de como tais aspectos se manifestariam no texto para ter uma idéia de o que procurar nos corpora a serem utilizados. O primeiro passo é definir um Corpus de Referência, contra o qual vamos comparar o nosso Corpus de Estudo, isto é, o conjunto dos quinze contos de Joyce.

A linguagem literária é particular, no que concerne a sua riqueza lexical e amplo uso de figuras de estilo, e difere portanto do tipo de linguagem apresentada nos grandes corpora de referência, como o BNC (British National Corpus), por exemplo. Além disso, a forma literária de *Dubliners* é a do conto. Como é uma modalidade com características estruturais próprias, o estilo e peculiaridades de Joyce só podem ser realmente avaliados se se comparar o corpus com outras coletâneas de contos, em que ficarão anulados os aspectos comuns a todos os contos. Assim, para possibilitar o confronto de textos semelhantes, foi necessário compilar um Corpus de Referência com características especiais, como veremos a seguir.

3. Compilação dos Corpora de Estudo e de Referência

O Corpus de Estudo consiste dos quinze contos de Joyce que formam a obra *Dubliners*, em forma eletrônica, obtidos por meio de busca na rede. Cada conto foi nomeado individualmente, de modo a poder ser identificado quando necessário.

³ “My intention was to write a chapter of the moral history of my country and I chose Dublin for the scene because that city seemed to me the centre of paralysis. [...] I have written it for the most part in a style of scrupulous meanness and with the conviction that he is a very bold man who dares to alter in the presentment, still more deform, whatever he has seen or heard.” (BOSNELLI & MOSHER Jr., 1998).

Os contos no início de século XX guardavam ainda uma estrutura bastante rígida, herança da preocupação com a diferenciação do que se chamava na Inglaterra “tale” e “short-story”. Este último gênero era caracterizado nos moldes tradicionais da definição, geralmente apresentando um só enredo, tendo dissertações e mesmo descrições reduzidas a um mínimo e exibindo grande compressão de linguagem. Depois da poesia, o conto era considerado o gênero textual mais compacto. Todos os detalhes irrelevantes à trama eram suprimidos, os personagens se revelavam pelas ações e falas. Assim é razoável que o Corpus de Referência para ser comparado aos contos de Joyce fosse um corpus de contos, preferencialmente da mesma época. Naturalmente, os autores escolhidos não eram a única escolha possível, mas alguns critérios foram considerados para que se criasse um Corpus de Referência verdadeiramente relevante. Procuramos nomes que influenciariam os rumos do conto como forma literária. Virginia Woolf (1882–1941) é uma figura reconhecidamente diferenciada na literatura, que se impõe de imediato por ter sido a continuadora da técnica do fluxo da consciência, com o seu uso dos monólogos interiores. Katherine Mansfield (1888–1923), além de voluntariamente expatriada e falante nativa não-britânica como Joyce (ela era neozelandesa), é reconhecidamente uma grande contista. Seguindo a tradição de Tchekhov, elaborou contos perfeitamente dentro da estrutura formal e, ainda assim, trouxe inovações como, por exemplo, iniciar o conto *in media res*, colocar em foco a trivialidade da comédia da vida e, algumas vezes, deixar o final em aberto, suscitando a direta participação do leitor no enredo. D. H. Lawrence (1885–1930) aparece como um contista relevante e polêmico, embora não se possa dizer que considerava o efeito de compressão tão essencial como as escritoras mencionadas. Desafiou as convenções de sua época e privou, assim como Mansfield e Woolf, dos círculos de escritores mais seletos de seu tempo. O Corpus de Referência foi, então, composto da seguinte maneira:

- 23 contos de Virginia Woolf (total de 53.659 palavras)
- 12 contos de D. H. Lawrence (total de 79.087 palavras)
- 17 contos de Katherine Mansfield (total de 79.845 palavras)

O total de palavras do Corpus de Referência assim constituído é de 212.591 palavras. Nota-se que os contos de Virginia Woolf, embora em maior número, constituem o menor sub-corpus. Porém, conforme atesta Margaret Drabble, editora do ***The Oxford Companion to English Literature***, estes foram todos os contos que Woolf escreveu. Comparando com o nosso Corpus de Estudo, de 68.001 palavras, constatamos que o Corpus de Referência é 3,13 vezes maior, o que é um tamanho aceitável, como afirma Berber Sardinha: “Os tamanhos críticos de corpora de referência são 2, 3 e 5 vezes o tamanho do corpus de estudo.

GONÇALVES, Lourdes Bernardes. *Linguística de Corpus e Análise Literária: uma Aplicação a Dubliners de Joyce*.

Corpora de referência com estas dimensões retornam significativamente mais palavras-chave do que corpora de tamanhos menores.” (2004: 102). Nomeamos cada um dos quinze contos com um código de, no máximo, seis letras, para possibilitar o reconhecimento de cada conto no decorrer da utilização do WordSmith Tools, conjunto de ferramentas utilizado para a manipulação dos corpora.

Em relação às características do Corpus de Referência, é um corpus estático, relacionado a uma época determinada e representativo de três escritores, em linguagem escrita, monolíngüe, contempla somente o gênero contos e é composto de textos completos. A nomeação dos contos no Corpus de Referência seguiu o modelo do Corpus de Estudo. O Sub-Corpus Lawrence foi composto de contos tirados do texto eletrônico *The Prussian Officer and Other Stories* (<http://www.gutenberg.net.au/ebooks03>). O sub-corpus Mansfield foi constituído de 15 contos tirados do texto eletrônico *The Garden Party and Other Stories* (<http://www.gutenberg.net/etext/1429>) e, para completar o número de palavras desejável, acrescentamos mais dois contos da autora, “Bliss” e “Prelude”, também obtidos da WEB (<http://www.digital.library.upenn.edu/women/mansfield.html>). O Sub-Corpus Woolf consta dos 23 contos da autora tirados do texto eletrônico *A Haunted House and Other Stories* (<http://www.gutenberg.net.au/ebooks02>).

Uma vez compilado o Corpus de Referência, várias pesquisas podem ser feitas a partir de sua comparação com o Corpus de Estudo, como a que veremos a seguir.

4. Usando o Corpus de Referência numa pesquisa direcionada por corpus

Tognini-Bonelli (2001: 65-99) distingue dois tipos básicos de pesquisa com corpus, aquela direcionada pelo corpus (*corpus-driven*), que pressupõe do pesquisador uma postura aberta, pronta para tirar vantagem das informações que a Linguística de Corpus fornecer, e a baseada em corpus (*corpus-based*), que é calcada em hipóteses pré-formuladas pelo pesquisador, que as palavras recorre aos corpora para sua confirmação ou refutação. Na análise literária a que procederemos aqui será utilizada a primeira abordagem, embora reconheçamos o valor e importância da segunda em certas situações.

Uma vez instalados o Corpus de Estudo e o Corpus de Referência, usamos as ferramentas do software WordSmith, para obter as Listas de Palavras (Wordlists) dos dois corpora, fornecidas por ordem de frequência em que as palavras ocorrem nos textos. A partir da comparação dessas duas listas, o software, por meio da ferramenta Keywords (Palavra-Chave), identifica as palavras mais comuns no Corpus de Estudo que nele se sobressaem por ocorrerem em frequência signifi-

cativamente maior que no Corpus de Referência. São as Palavras-Chave que vão revelar o que há de peculiar no Corpus de Estudo. Os resultados apresentaram um grande número de nomes próprios, o que era de se esperar, mas as outras palavras são bastante reveladoras acerca do conteúdo do texto. Há nesta lista uma variável chamada *keyness*, traduzida por “chavicidade”, que dá a medida em que uma palavra é original e característica do Corpus de Estudo em relação ao Corpus de Referência. A palavra de maior chavicidade (ch) foi *Mr* (ch = 922,3). Como referência, ressaltamos que se costuma aceitar como chavicidade significativa valores maiores que 4,0. Partindo daí, pesquisamos palavras como *he* (ch = 117,6), *his* (ch = 193,8) e *man* (ch = 39,4), que mostraram a predominância de personagens masculinos ou, pelo menos, a importância dada por Joyce a esses personagens, apesar de contos como “Eveline”, “A pensão” (“The Boarding House”), “Argila” (“Clay”), “Um caso trágico” (“A Painful Case”) e “Mãe” (“A Mother”), em que os personagens centrais são femininos.

Há uma medida de chavicidade negativa, que indica palavras que aparecem marcadamente no Corpus de Referência e não apresentam uma frequência relevante no Corpus de Estudo. Estas aparecem no final da lista de Palavras-Chave. No Corpus de Referência, a palavra de maior chavicidade negativa foi *she* (ch = -376,5), seguida de outras como *her* (ch = -111,8) e *herself* (ch = -45,3), o que confirma a predominância de alusões de Joyce a seus personagens masculinos. Notamos, porém, que nem *Mrs* nem *Miss* aparecem como palavras de chavicidade negativa, o que seria de se esperar dadas as ocorrências de *she*, *her* e *herself*. Uma consulta à Lista de Palavras do Corpus de Estudo mostra a frequência de *Mr* (576 ocorrências), *Mrs* (117 oc.) e *Miss* (99 oc.). Isso significa que, embora os personagens femininos do Corpus de Estudo sejam menos mencionados, mesmo proporcionalmente, que no Corpus de Referência, praticamente todas as vezes que são mencionados vêm acompanhados de *Mrs* ou *Miss*, levantando a chavicidade destes dois últimos termos em relação ao Corpus de Referência. A conclusão é que a relação de Joyce com seus personagens sugere formalidade. O autor dá-lhes amiúde o título de *Mr*, *Mrs* e *Miss*, o que os outros três autores do Corpus de Referência não fazem. O tratamento formal de Joyce se confirma ainda com a alta chavicidade da palavra *gentlemen* (ch = 61,7), enquanto *man* apresenta uma chavicidade mais baixa (ch = 39,4) assim como com a chavicidade de *ladies* (ch = 27,6), apesar da destacada ausência de pronomes femininos no texto, comprovada pelas chavidades já mencionadas, e ainda pela chavicidade negativa da palavra *woman* (ch = -12,0). Uma explicação para esse fenômeno é que o distanciamento entre narrador e personagem é importante para Joyce, que pretende fazer um relato não-emotivo, racional, do perfil do dublinense. Qualquer envolvimento emocional turvaria a descrição e diminuiria a qualidade objetiva de sua apresentação.

A nosso ver, a chavidade negativa é uma das maiores contribuições da Linguística de Corpus em relação à análise literária, porque é muito mais difícil se perceber num texto a ausência de algum elemento que qualquer presença. Um exame das palavras de chavidade negativa nos deu resultados interessantes. Podemos incluir várias dessas palavras em grupos semânticos, dos quais alguns passam agora a ser nomeados, com suas chavidades:

- Cores: *white* (-31,6); *blue* (-28,0); *black* (-21,7); *pink* (-21,3); *red* (-19,0); *yellow* (-13,0); *green* (-11,1).
- Natureza: *garden* (-36,7); *trees* (-31,5); *tree* (-29,2); *water* (-28,9); *sea* (-24,2); *flowers* (-17,3); *sky* (-17,6); *grass* (-13,5).
- Emoção: *dear* (-35,9); *cried* (24,2); *feeling* (-16,1); *loved* (-15,6).

O último grupo consiste de poucas palavras, mas nem por isso é menos significativo. Especialmente quando se pensa que Woolf é extremamente parcimoniosa na expressão de emoções e Mansfield sublima a emoção sob a forma de ironia e do implícito. Se considerarmos a Lista de Palavras-Chave somente em relação ao Corpus de Estudo e sub-corpus Lawrence, vamos encontrar, entre outras, palavras com a seguinte chavidade negativa: *pain* (-29,4); *felt* (-28,6); *cried* (-26,9); *fear* (-25,8); *feeling* (-23,4); *heart* (-18,8); *loved* (-17,8); *afraid* (-17,3); *shame* (-17,3); *anger* (-13,8); *pale* (-12,7).

Vemos, assim, que Joyce realmente utiliza uma linguagem contida emocionalmente. Esta conclusão é reafirmada, quando, retornando para a Lista de Palavras-Chave construída a partir do Corpus de Estudo versus Literário, observamos as seguintes palavras, agora de chavidade positiva: *committee* (27,5); *spoke* (24,9)⁴; *accent* (23,6); *opinion* (18,2); *conversation* (17,7); *bearers* (17,0); *nationalist* (17,0); *address* (16,0)⁵; *alluding* (14,2); *reported* (14,2); *express* (13,8); *phrase* (11,5); *arguing* (11,3); *monologue* (11,3); *narrative* (11,3).

Notamos que essas palavras se referem a operações cognitivas e algumas a questões políticas, como discursos, argumentações e discussões. O ato de falar é muito importante em Joyce, provavelmente por ressaltar o contraste entre palavra e ação, que ele vê como central na problemática da Irlanda, além do aspecto da expressão racional já mencionado. Para reiterar essa afirmação, encontramos *said* com chavidade de 94,8, assim como outros verbos dicendi como *told* (26,6),

⁴ A palavra *spoke* em todos os casos foi utilizada para introduzir o modo como alguém falava e não como verbo dicendi.

⁵ A palavra *address* em 6 das 10 ocorrências se refere a discurso oratório e não endereço postal.

Crop, 10, 2004

retort (14,2), *asked* (14,0), *express* (13,8) e *stated* (11,3). Vários períodos compostos por orações adjetivas também sugerem explicações e definições, o que se verifica através da forte presença no discurso de Joyce de *which* (53,5) e *who* (27,5).

Surge neste momento, no leitor crítico de Joyce, uma inquietação em relação ao seu processo de criação dos personagens, pois esse leitor sente que emana de muitos dos personagens em **Dubliners** uma sensibilidade e melancolia que não podem ser expressas por uma linguagem puramente objetiva e racional. De alguma maneira a emoção percola por esse discurso e se evidencia de maneira sutil, bem à Joyce. Só resta esperar que ao longo da pesquisa surja uma maneira de se reconhecer esse traço do autor.

Uma análise das cores mostra que as mais ocorrentes no Corpus de Referência são aquelas mais freqüentemente associadas com o simbolismo. A dedução é de que a simbologia de Joyce não inclui cores; ele só as usa como parte das descrições. Esse aspecto reforça o que já se sabia sobre a sua escolha de imagens e figuras. Ele é absolutamente original em suas metáforas, que são geralmente complexas, com múltiplos sentidos embutidos, e pertencentes a campos semânticos os mais variados. É, portanto, muito difícil a identificação da imagística de Joyce pela Lingüística de Corpus. Porém, em relação às cores, apesar de não se lhes perceber o uso direto, pode-se identificar uma clara preferência pela noite, quando se observa a chavicidade das palavras *tonight* (28,0), *night* (26,7) e *evening* (19,2).

A ausência de menções à natureza aponta para o interesse de Joyce no cidadão, mais que na paisagem, e no urbano, mais que no rural. Palavras-Chave que se referem à cidade de Dublin e à Irlanda, por extensão, são bastante comuns em seu texto: *Dublin* (107,7) *street* (86,8); *Irish* (56,1); *city* (53,5); *tram* (35,5); *quay* (34,0); *cabman* (17,0); *streets* (13,9); *Ireland* (13,8); *canal* (11,3); *quays* (11,3).

Naturalmente, as palavras *Dublin*, *Ireland* e *Irish* têm chavicidade alta por serem nomes próprios e gentílico, mas a freqüência no Corpus de Estudo torna-as significativas: *Dublin*, 38 ocorrências, *Ireland/Irish*, 34 ocorrências.

A preocupação do autor com os negócios e a vida profissional de Dublin reflete também a sua inclinação para o urbano, além de indicar um possível desejo de mostrar o dublinense como trabalhador, apesar de trabalhar com eficiência questionável, como mostram os contos “A Little Cloud” e “Counterparts”, entre outros. O mundo dos negócios é representado pelas Palavras-Chave: *money* (65,3); *office* (38,0); *clerk* (31,3); *copy* (30,2); *contract* (30,2); *business* (27,0); *desk* (25,1); *cashier* (19,8); *chief* (18,2).

Como exemplo de confirmação das expectativas, podemos mostrar a questão do abuso do álcool como uma preocupação de Joyce, que já se esperava no texto, e que se confirmou com as Palavras-Chave: *bottles* (34,8); *stout* (29,6)⁶; *whisky* (28,7); *bar* (22,9); *drank* (21,5); *punch* (19,8); *bottle* (15,3); *corkscrew* (14,4); *sober* (14,2); *porter* (14,1)⁷; *glasses* (13,2); *barman* (11,3).

O mesmo poderia ser feito em relação a outros campos semânticos, como o da religião, entretanto as dimensões deste trabalho não permitem a exploração de todo o rico material de pesquisa fornecido pela Linguística de Corpus.

Na pesquisa direcionada por corpus, o investigador depara muitas vezes com fatos que não faziam parte de suas hipóteses e até mesmo haviam escapado à sua intuição. Foi o que ocorreu quando deparamos com um campo semântico que se destacou de modo impressionante no exame das Palavras-Chave: o de referências musicais, que passamos a especificar com suas palavras e chavidades: *tenor* (45,4); *concert* (43,7); *artistes* (34,0); *concerts* (24,9); *song* (22,8); *baritone* (22,7); *clapping* (22,7); *opera* (17,0); *music* (15,7); *piano* (15,7); *sing* (14,4); *musical* (14,4); *artiste* (14,2);⁸ *accompanist* (14,2); *waltz* (13,8); *melody* (11,8); *applause* (11,8)⁹; *bass* (11,5); *audience* (11,4); *singers* (11,3).

Há mais palavras ainda usadas por Joyce nesse campo semântico, que podem ser obtidas através de uma busca na Lista de Palavras (*Wordlist*) do Corpus de Estudo: *PLAY**¹⁰ – 27 ocorrências musicais em 56; *SING** – 37 ocorrências; *choir* – 3 ocorrências; *chorus* – 3 ocorrências; *contralto* – 3 ocorrências; *soprano* – 3 ocorrências; *DANCE** – 12 ocorrências; *air* – 9 ocorrências como *música* entre 40.

O leitor de Joyce já sabia que o escritor tinha grande sensibilidade e conhecimento musical. Vários são os registros de reconhecimento desta característica. Já em 1965, Anthony Burgess lembra o fato de que o próprio autor era um tenor (note-se que *tenor* é a palavra de maior chavidade no campo semântico musical). Referindo-se ao pai do escritor, e depois a James Joyce, Burgess relata: “John Joyce não era simplesmente um tenor, mas um bom tenor, e seu filho mais velho foi quase um grande tenor. A importância da canção em seus

⁶ A palavra *stout* é utilizada no Corpus de Estudo 21 vezes, 12 como tipo de bebida e 9 descrevendo um tipo físico.

⁷ A palavra *porter* aparece 11 vezes, 3 das quais como tipo de bebida.

⁸ As palavras *artistes* e *artiste* sempre apareceram ligadas a exibições musicais.

⁹ A palavra *applause*, referente a apresentações musicais, aparece em 4 das 6 ocorrências no Corpus de Estudo.

¹⁰ A palavra em caixa alta seguida de um asterístico indica a forma lematizada, correspondendo a todas as formas com que a palavra se apresenta, no texto.

Crop, 10, 2004

livros não pode ser exagerada – **Ulysses** canta em todo o seu percurso ou, quando não canta, declama ou entoa. Tornou-se uma peça de teatro – **Bloomsday**; também poderia ser transformado em uma ópera.”¹¹

Paulo Vizioli (1991: 27) também ressaltou esse aspecto joyceano: “A influência da música na obra de Joyce é mais que óbvia. Em **Ulysses**, por exemplo, é responsável não só pelas constantes alusões a árias e cenas operísticas, mas também por alguns símbolos empregados (como a ‘Dança das Horas’ e pelo esquema de alguns capítulos (como o das ‘Sereias’, que reproduzem uma ‘fuga per canonem’).”

Pouco foi comentado, no entanto, sobre a música em **Dubliners**. As únicas referências que puderam ser encontradas dizem respeito às quatro canções mencionadas na coletânea, ou claramente relacionadas a ela: “*I’ll sing thee songs of Araby*”, “*Silent, O Moyle*”, “*I dreamt that I dwelt in marble halls*” e “*Yes! Let me like a soldier fall*”. As canções foram discutidas pelo Prof. Zack Bowen e consideradas esclarecedoras de motivações nas histórias, assim como elas mesmas possíveis símbolos metafóricos.¹² Os comentários são interessantes, mas há bem mais a ser observado em **Dubliners**.

A Lingüística de Corpus torna a desempenhar aqui um papel fundamental por possibilitar uma observação minuciosa do contexto das palavras pertencentes ao campo semântico musical. Há no WordSmith uma ferramenta chamada *Concord* que pode mostrar a palavra selecionada em uma linha do texto e pode informar de qual texto (no caso, de qual conto) a palavra foi tirada. Além disso, a linha de concordância pode ser expandida, se for necessária uma contextualização maior.

Examinando então as concordâncias de Palavras-Chave relativas à música, pudemos perceber-lhes a importância fundamental na tessitura de **Dubliners**. Sua participação não é “periférica”, como sugere o Prof. Bowen, mas parte integrante da estrutura da obra. A música define personalidades. Esta observação resolve o problema mencionado anteriormente de como Joyce transmite a sensibilidade de seus personagens, e por extensão dá vazão à sua sensibilidade, mantendo uma linguagem seca e objetiva. Mas a função da música vai além de delinear

¹¹ “John Joyce was not just a tenor but a fine tenor, and his eldest son was almost a great tenor. The importance of song in his books cannot be exaggerated. **Ulysses** sings all the way or, when it does not sing, it declaims or intones. It has been turned into a stage play – **Bloomsday**; it could also be turned into an opera.” (BURGESS, 1965: 28)

¹² Essas discussões são interessantes e podem ser encontradas no site <www.james-joyce-music.com>

personagens; ela define ambientes, estados de espírito e dá muitas vezes o tom da narrativa. A título de exemplo, passaremos a apresentar como a influência da música se faz sentir na definição de alguns personagens e, em seguida, como ela dá o tom da narrativa.

Em “Eveline”, o marinheiro que tenta tirar a jovem da paralisia em que ela se encontra, oferecendo-lhe casamento e uma nova vida em Buenos Aires, “levou-a para assistir *The bohemian girl* e ela ficou radiante por sentar-se ao lado dele num setor do teatro onde não costumava ficar. Ele adorava música e tinha uma voz razoável” (Joyce, 1993: 47)¹³. Certamente se percebe a idéia de gostar de música associada ao fato de o marinheiro ser uma pessoa boa e generosa, já que a fez sentar num bom lugar no teatro.

Assim como ele, Jimmy e Villona, em “Depois da corrida”, têm uma descrição simpática ao leitor, que é relacionada à afinidade deles com a música. Assim, Jimmy, o irlandês de um grupo de jovens de diferentes nacionalidades, “tinha dinheiro e popularidade; e dedicava todo o seu tempo a atividades musicais e automobilísticas”(Joyce, 1993: 52)¹⁴. E Villona, o húngaro, “era um sujeito divertido – além de ser um pianista brilhante – mas, infelizmente, muito pobre” (Joyce, 1993:52)¹⁵. Note-se que o fato de ser bom pianista está tão ligado à prosódia positiva que o autor introduz uma adversativa para mencionar o fato de o húngaro ser pobre.

Pode ocorrer que a maneira de uma pessoa cantar represente, numa metonímia, o modo de ela ser e agir. Em “Mãe”, temos: “A pobre mulher cantou *Killarney* com uma voz fraca e ofegante, com todos aqueles floreios ultrapassados de entonação e pronúncia que a seu ver atribuíam elegância a sua interpretação” (Joyce, 1993: 150)¹⁶.

No mesmo conto, a sensibilidade do segundo tenor, Mr. Bell, revela-se também através da música: “O primeiro tenor, o barítono e Miss Healy estavam lado a lado, aguardando tranqüilamente num canto; Mr. Bell, entretanto,

¹³ “... took her to see *The Bohemian Girl* and she felt elated as she sat in an unaccustomed part of the theatre with him. He was awfully fond of music and sang a little.” (citação tirada do site – referência na bibliografia) Todas as traduções dos contos de Joyce aqui apresentadas são da autoria de José Roberto O’Shea.

¹⁴ “... had money and he was popular, and he divided his time curiously between musical and motoring circles.”

¹⁵ “... was entertaining also – a brilliant pianist – but, unfortunately, very poor.”

¹⁶ “The poor lady sang *Killarney* in a bodiless gasping voice, with all the old-fashioned mannerism of intonation and pronunciation which she believed lent elegance to her singing.”

estava com os nervos bastante abalados por recear que o público achasse que o atraso era culpa dele” (Joyce, 1993: 149)¹⁷. A ironia desta descrição é mais claramente percebida se se observar a descrição de Mr. Bell. Mais uma vez, há uma forte relação entre música e caráter. Trata-se de um músico medíocre, de caráter invejoso, mesquinho: “Mr. Bell, o segundo tenor, era um sujeito alourado e de baixa estatura que uma vez por ano participava dos concursos em Feis Ceoil. Na quarta tentativa, recebera uma medalha de bronze. Estava extremamente nervoso e sentia-se extremamente ameaçado pelos outros tenores e disfarçava o nervosismo através de uma amabilidade exagerada” (Joyce, 1993:146)¹⁸. Observe-se que o texto de Joyce é ainda mais forte que a tradução, por usar duas vezes variações da palavra *ciúme*, *jealous* e *jealousy*: “*He was extremely nervous and extremely jealous of other tenors and covered his nervous jealousy with an ebullient friendliness.*” Uma sensibilidade autêntica revelada pela apreciação sincera da música encontra-se em “Os mortos”, com tia Kate: “– Foi um tenor inglês, bonito, com uma voz doce, límpida, aveludada – disse tia Kate, vibrando de emoção” (Joyce, 1993: 200)¹⁹.

Numa passagem de “Mãe”, encontramos de novo o uso de uma conjunção adversativa, agora para justapor a música a uma alusão desabonadora de comportamento. Essa conjunção não foi preservada na tradução, mas o sentido adversativo fica claro:

“Cantou com muito sentimento e grande volume de voz e foi calorosamente recebido pelas galerias. Infelizmente, não causou maior impressão porque sem querer limpou o nariz uma ou duas vezes com a mão enluvada” (Joyce, 1993: 146)²⁰.

Encontramos, ainda, um caso de confronto entre o gosto refinado, detalhista, com a não-aptidão para a música, que aparece como uma surpresa,

¹⁷ “The first tenor and the baritone and Miss Healy stood together, waiting tranquilly, but Mr. Bell’s nerves were greatly agitated because he was afraid the audience would think that he had come late.”

¹⁸ “Mr. Bell, the second tenor, was a fair-haired little man who competed every year for prizes at the Feis Ceoil. On his fourth trial he had been awarded a bronze medal. He was extremely nervous and extremely jealous of other tenors and covered his nervous jealousy with an ebullient friendliness.”

¹⁹ “‘A beautiful, pure, sweet, mellow English tenor,’ said Aunt Kate with enthusiasm.”

²⁰ “He sang his music with great feeling and volume and was welcomed by the gallery; but unfortunately, he marred the good impression by wiping his nose in his gloved hand once or twice out of thoughtlessness.”

neste caso talvez justificada pelo predomínio do racional sobre o sensível. É em “Os mortos”, em que temos: “como presente de aniversário, sua mãe bordara para ele umas pequenas cabeças de raposa num colete de seda lilás, forrado de cetim marrom e com botões redondos em formato de amora. Era estranho que a mãe não tivesse nenhum talento musical, pois tia Kate costumava referir-se a ela como o cérebro da família Morkan” (Joyce, 1993: 187-8)²¹. Na tradução, a substituição da conjunção *though* por *pois* prejudicou a idéia de uma possível explicação.

Joyce chega mesmo a fazer ironias musicais. Em “Um caso trágico”, encontramos Mr. James Duffy, que gosta de música mas não é capaz de mostrar solidariedade à sua própria amante. Há a sugestão de que ele só admite Mozart, freqüentemente considerado pelos apreciadores de música como perfeito, e não tolera a fraqueza. Esta foi a razão de ter terminado seu caso ilícito, num aparente assomo de retidão, mas com um real descaso pelos sentimentos da companheira. Joyce ironiza sua “respeitabilidade” e enfatiza sua solidão: “À noite costumava sentar-se ao piano da senhoria ou caminhava pelos arredores da cidade. O gosto pela música de Mozart às vezes o levava à ópera ou a um concerto: as únicas extravagâncias a que se permitia” (Joyce, 1993: 113-4)²². Como se a ida à ópera ou a um concerto fosse para ele, em sua vida vazia, algo excepcional. Porém, a mesma sensibilidade que o fez apreciar Mozart o força finalmente a reconhecer sua culpa na morte (provável suicídio) da amante, ação que parece mostrar que um apreciador genuíno de música não pode permanecer, para sempre, insensível. Passa de reflexões como: “Ela não tinha apenas aviltado a si mesma mas a ele também” (Joyce, 1993: 120) a outras bem diferentes: “Um ser humano talvez o tivesse amado e ele negara-lhe vida e felicidade: condenara essa pessoa à ignomínia, a uma morte vergonhosa” (Joyce, 1993: 121-2).

Mostra também de forma irônica que nem todos os que foram expostos à música internalizaram a sensibilidade ou emotividade relativa a ela; alguns permanecem frios. Mrs. Kearney, em “Mãe”, teve uma formação musical: “Tinha estudado num excelente colégio de freiras onde aprendera francês e música. Por ser pálida de natureza e por ter maneiras um tanto rígidas fizera poucas

²¹ “His mother had worked for him as a birthday present a waistcoat of purple tabinet, with little foxes’ heads upon it, lined with brown satin and having mulberry buttons. It was strange that his mother had had no musical talent, though Aunt Kate used to call her the brains carrier of the Morkan family.”

²² “His evenings were spent either before his landlady’s piano or roaming about the outskirts of the city. His liking for Mozart’s music brought him sometimes to an opera or a concert: these were the only dissipation of his life.”

amizades na escola. Quando atingira a idade de se casar fizeram-na freqüentar diversas residências onde seus talentos musicais e suas maneiras se tornaram objeto de admiração” (Joyce, 1993: 140)²³. Isso, contudo, não evitou que ela tivesse um comportamento mercenário na negociação do pagamento dos concertos de piano da filha. É como se tivesse aprendido música por imposição e a música nunca tivesse realmente a sensibilizado. No original, percebe-se mais claramente esse aspecto, pois lamentavelmente o tradutor omitiu *ivory*, adjetivando *manners*, que pontua a frieza do personagem.

Um caráter duvidoso também pode ser denunciado através da música. Em “A pensão”, o traço agressivo e grosseiro de Jack, filho de Mrs. Mooney, a dona da pensão, se evidencia pelo seu gosto por “músicas cômicas”, vulgares: “Sempre que encontrava os amigos tinha uma boa história para lhes contar e estava sempre no faro de alguma grande jogada, fosse com respeito a algum cavalo ou alguma artista. Era também hábil com os punhos e gostava de cantar canções picantes” (Joyce, 1993: 71)²⁴.

Polly, filha de Mrs. Mooney, mostra seu lado vulgar e provocador pela música que escolhe para cantar: “também cantava:

*Sou uma ... garota levada.
Por que então esconder:
Todos já devem saber.*

Polly era esbelta e tinha dezenove anos; seus cabelos eram claros e macios e sua boca, pequena de lábios carnudos. Os olhos, cinzentos e levemente esverdeados, tinham o hábito de virar para cima sempre que ela falava com alguém, o que lhe dava um ar de madona perversa” (Joyce, 1993: 71)²⁵.

A atmosfera freqüentemente se instala por meio do clima musical que se faz sentir. Em “Depois da corrida”, a corrida vertiginosa dos acontecimentos,

²³ “She had been educated in a high-convent, where she had learned French and music. As she was naturally pale and unbending in manner she made few friends at school. When she came to the age of marriage she was sent out to many houses where her playing and ivory manners were much admired.”

²⁴ “When he met his friends he had always a good one to tell them, and he was always sure to be on to a good thing – that is to say, a likely horse or a likely artiste. He was also handy with the mits and sang comic songs.”

²⁵ “She sang: *I’m... a naughty girl... You needn’t sham: you know I am.* Polly was a slim girl of nineteen; she had light soft hair and a small full mouth. Her eyes, which were grey with a shade of green through them, had a habit of glancing upwards when she spoke with anyone, which made her look like a perverse madonna.”

GONÇALVES, Lourdes Bernardes. *Linguística de Corpus e Análise Literária: uma Aplicação a Dubliners de Joyce*.

que termina no luxo do iate do americano e no jogo de cartas alucinante, é toda pontuada pela música, como fica evidente nos trechos:

“Villona estava mesmo de ótimo humor; por quilômetros e quilômetros viera entoando com uma voz profunda, de baixo, uma melodia a meia voz” (Joyce, 1993: 53)²⁶.

“O veículo corria ao lado da multidão, agora diluída em cores suaves, em meio à música de sinos jubilosos” (Joyce, 1993: 56)²⁷.

“A bordo haveria comida, música, carteadado. Villona disse com entusiasmo:

– Que beleza!

Na cabine havia um pequeno piano. Villona tocou uma valsa para Farley e Rivière [...]” (Joyce, 1993: 56)²⁸.

Em “Os mortos”, lembrando os tempos passados de Dublin, há uma evocação de sua qualidade musical, ainda superior à atual, dourando as lembranças: “[Mr. Browne] Contou, também, como as galerias do velho Theatre Royal estavam sempre lotadas, noite após noite e como, certa noite, um tenor italiano bisou cinco vezes o ‘Deixe-me tombar como soldado’, sem omitir o dó de peito uma única vez, e, ainda, como, em seu entusiasmo, os rapazes que freqüentavam as galerias às vezes desatrelavam os cavalos do coche de uma *prima donna* e puxavam eles mesmos o veículo até o hotel” (Joyce, 1993: 199)²⁹. Esta passagem mostra como uma grande emoção pode ser narrada por Joyce sem que ele se desvie de seu estilo contido. No mesmo conto, ainda é a música que dá o tom de festividade à reunião: “Todos os convidados levantaram-se, de taça na mão e, virando-se para as três senhoras, cantaram em uníssono, sob o comando de Mr. Browne:

²⁶ “Decidedly Villona was in excellent spirits; he kept up a deep bass hum of melody for miles of the road.”

²⁷ “They drove by the crowd, blended now into soft colours, to a music of merry bells.”

²⁸ “There was to be supper, music, cards. Villona said with conviction: ‘It’s delightful’. There was a yacht piano in the cabin. Villona played a waltz for Farley and Rivière [...]”

²⁹ “He told too of how the top gallery of the old Royal used to be packed night after night, of how one night an Italian tenor had sung five encores to ‘Let me like a soldier fall’, introducing a high C everytime, and of how the gallery boys would sometimes in their enthusiasm unyoke the horses from the carriage of some great *prima donna* and pull her themselves through the streets to her hotel.”

Crop, 10, 2004

As três são boas companheiras,
As três são boas companheiras,
As três são boas companheiras,
Ninguém pode negar” (Joyce, 1993: 205)³⁰.

A festa musical atrai até os que passam na rua, apesar do frio:

“O piano tocava uma valsa e Gabriel ouvia o farfalhar das saias roçando na porta do salão. Talvez houvesse pessoas no cais lá fora, na neve, olhando para as janelas iluminadas e ouvindo a valsa” (Joyce, 1993: 202)³¹.

Essa análise poderia se prolongar, mas não há espaço nem razão para tal, pois o propósito deste trabalho foi o de apontar modos de como se pode explorar um texto literário usando a Lingüística de Corpus.

5. Conclusão

O objetivo deste trabalho foi demonstrar o amplo leque de possibilidades do uso da Lingüística de Corpus para a análise literária. É uma pesquisa que vai fornecendo novo material à medida em que progride por mostrar, muitas vezes, aspectos que não haviam ocorrido ao investigador, ou para justificar afirmações decorrentes do processo de interpretação. A Lista de Palavras, obtida com o software WordSmith Tools, por si só já fornece uma variedade de informações preciosas. As Palavras-Chave, também fornecidas pelo programa, vão revelar as áreas semânticas mais importantes do texto em estudo, uma vez que elas representam palavras que foram usadas com frequência maior que no Corpus de Referência. No caso em estudo, tivemos a primeira surpresa com *Mr*, de chavidade 922,3, especialmente alta. É importante salientar que a Lista de Palavras deve ser estudada em todos os seus detalhes, pois qualquer forma morfológica pode ser relevante na análise literária. Por exemplo, a exclusão do título *Mr*, como vimos, implicaria uma perda significativa de informação. Essa

³⁰ “All the guests stood up, glass in hand, and turning towards the three seated ladies, sang in unison, with Mr. Browne as leader: ‘For they are jolly gay fellows, For they are jolly gay fellows, For they are jolly gay fellows, Which nobody can deny’.”

³¹ “The piano was playing a waltz tune and he would hear the skirts sweeping against the drawing-room door. People, perhaps, were standing in the snow on the quay outside, gazing up at the lighted windows and listening to the waltz music.”

GONÇALVES, Lourdes Bernardes. *Linguística de Corpus e Análise Literária: uma Aplicação a Dubliners de Joyce*.

observação nos levou a considerações sobre a formalidade do estilo de Joyce, a conseqüente especulação sobre como a emoção se manifestaria em seus contos e a surpreendente descoberta de toda uma área semântica musical usada por Joyce. É interessante também lembrar que, nesse tipo de análise, é imprescindível que o Corpus de Referência seja de fato comparável com o Corpus de Estudo para que fatores discrepantes como gêneros ou épocas diferentes não venham a influenciar a pesquisa.

A Linguística de Corpus provou ser de grande valia neste estudo, principalmente na identificação da área semântica musical e na observação das Concordâncias das palavras dessa área. Já houve estudos sobre a utilização da música em *Ulysses* e *Finnegans Wake*, e mesmo em *Dubliners*, mas geralmente colocando-a como pano de fundo ou como um reforço metafórico. As Concordâncias, no entanto, mostraram claramente como Joyce usou um vocabulário da área de música no próprio delineamento de seus personagens. *Insights* como esse certamente fazem da Linguística de Corpus um poderoso instrumento para Análise Literária.

Bibliografia

Livros e artigos:

- BERBER-SARDINHA, Tony. 2004. *Linguística de Corpus*. Editora Manole Ltda. São Paulo.
- BOSINELLI, Rosa M. B. & MOSHER Jr., Harold F. (Ed.). 1998. *ReJoycing: New readings of Dubliners*. The University Press of Kentucky. Lexington.
- BOWKER, Lynne. 2001. "Towards a methodology for a corpus-based approach to translation studies", in *META XLVI* 2, 2001, 345-364.
- BURGESS, Anthony. 1965. *ReJoyce*. W. W. Norton & Co. New York.
- DRABBLE, Margaret. 1996. *The Oxford Companion to English Literature*. Oxford University Press. 5th edition. Oxford.
- JOYCE, James. 1993. *Dublinenses*. Editora Siciliano. São Paulo. (Trad. José Roberto O'Shea)
- KENNEDY, Graeme. 1998. *An introduction to Corpus Linguistics*. Addison Wesley Longman Ltd. Harlowe.
- TOGNINI-BONELLI, Elena. 2001. *Corpus Linguistics at work*. John Benjamin. Amsterdam/Philadelphia.

Sites Eletrônicos:677

Textos de J. Joyce: <http://www.indbazaar.com/soulkurry> (download em 23/03/04)

Crop, 10, 2004

Textos de D. H. Lawrence: <http://www.gutenberg.net.au/ebooks03> (download em 31/03/04)

Textos de K. Mansfield: <http://www.gutenberg.net/etext/1429> (download em 31/03/04)
<http://www.digital.library.upenn.edu/women/mansfield.html> (download em 26/04/04)

Textos de V. Woolf: <http://www.gutenberg.net.au/ebooks02> (download em 31/03/04)
Site sobre Música e Joyce: www.james-joyce-music.com

NORMAS DE APRESENTAÇÃO DOS TRABALHOS ENVIADOS À CROP

1. Serão submetidos à aprovação da Comissão Editorial artigos sobre temas no campo de literatura, língua, tradução e estudos culturais de expressão inglesa. Os trabalhos podem ser redigidos em português ou inglês e enviados em disquetes digitados em Word for Windows 2.0, 6.0, 7.0, ou Word 97 acompanhados de três cópias impressas, sendo duas anônimas;
2. Os originais deverão se apresentados sem formatação especial. O texto corrido, sem recuos, deve estar marginado à esquerda e digitado em espaço duplo, sem divisão silábica. Usar o tabulador para os recuos inevitáveis; empregar a barra de espaços entre duas palavras, e apenas uma vez; usar a tecla <ENTER> apenas para terminar um parágrafo.
3. Todos os textos devem conter resumo e palavras-chave, *abstract* e *keywords*. Os resumos não devem ultrapassar 250 palavras. As palavras-chave devem ser limitadas ao máximo de 6.
4. As notas devem vir ao pé de página;
5. Pede-se especial atenção para a seguinte formatação:
 - **Fonte:** Times New Roman, corpo 12;
 - título: centralizado, com as primeiras letras em caixa alta, corpo 14;
 - nome do autor: caixa alta e baixa, em *italico*, alinhado à direita, com asterisco indicando para o pé de página a instituição a que está filiado;
 - empregue *italico* para palavras estrangeiras e neologismos;
 - empregue **negrito** para destaques, por exemplo, de termos técnicos;
 - evite **grifos**;
 - evite LETRAS MAIÚSCULAS, a não ser no início de palavras;
 - empregue “aspas” para citações (trechos mais extensos em parágrafos separados);
 - empregue ‘apóstrofos’ para citações dentro de citações;
 - para os nomes de autores citados, empregue caracteres normais, caixa alta e baixa. (não use LETRAS MAIÚSCULAS);
 - bibliografia: caixa alta, corpo 10, alinhada à esquerda. Entradas com edentação de 1,27.
6. Enviar os trabalhos para o Departamento de Letras Modernas, USP, LILINA:
Av. Prof. Luciano Gualberto 403 – CEP 05508-010 – São Paulo – SP.
Fax: 3091-5041.

LIVRARIA HUMANITAS
Av. Prof. Luciano Gualberto, 315 – Cid. Universitária
05508-900 – São Paulo – SP – Brasil
Tel: 3091-3728 / Telefax: 3091-3796
e-mail: livrariahumanitas@usp.br

HUMANITAS – DISTRIBUIÇÃO
Rua do Lago, 717 – Cid. Universitária
05508-900 – São Paulo – SP – Brasil
Telefax: 3091-4589
e-mail: pubfflch@edu.usp.br
<http://www.flch.usp.br/humanitas>

Ficha técnica

Mancha 11,5 x 19 cm

Formato 16 x 22 cm

Tipologia Garamond 12/15 e Coronet

Papel miolo: pólen soft 80g/m²

capa: color plus marfim 180 g/m²

Impressão da capa marron fotográfico

Impressão e acabamento PROVO GRÁFICA

Número de páginas 330

Tiragem 500 exemplares