

## TRADUZINDO O BRASIL PARA BRASILEIROS: UM ESTUDO DE LINGUÍSTICA DE CORPUS SOBRE O JORNALISMO *ON-LINE*\*

### Translating Brazil to Brazilians: A Corpus Linguistics Study about *On-line* Journalism

Maria Cecília LOPES (Faculdades Metropolitanas Unidas, São Paulo e COGEAE - Pontifícia Universidade Católica, São Paulo, Brasil)

#### Resumo:

*O objetivo deste artigo é apresentar os resultados parciais da tese defendida em 2010 que identificou imagens do Brasil veiculadas na imprensa on-line Norte-Americana e Inglesa em textos originais e suas traduções para o Português brasileiro a partir de metodologia desenvolvida para esse fim por meio da Linguística de Corpus com base no arcabouço teórico-metodológico já existente (SINCLAIR, 1987a), do Padrão Lexicogramatical (HUNSTON e FRANCIS, 2000), Pré-ativação Lexical (HOEY, 2005) e Prosódia Semântica (PARTINGTON, 1998). Os conceitos oriundos do desenvolvimento dessa metodologia são: participante-chave, colocado-chave, elemento indicador de imagem e **Pré-ativação imagética**. Apresentaremos as onze etapas metodológicas usadas no estudo.*

**Palavras-Chaves:** Imagem do Brasil, Padrões Lexicogramaticais, Tradução e Linguística de Corpus.

#### Abstract:

*This paper brings some of the results obtained in a research conducted in 2010 which aimed at identifying images of Brazil in texts published online in US and UK newspapers also translated into Portuguese. A new methodology was developed to draw on concepts from Corpus Linguistics (SINCLAIR, 1987a) and then these into original concepts related to the issue of image detection. The notions that the method started off with were lexicogrammatical patterns (HUNSTON and FRANCIS, 2000), Lexical Priming (HOEY, 2005) and Semantic Prosody (PARTINGTON, 1998). Concepts emerged throughout the development of the methodology which included: key-participant, key-collocate, image index element and **imagetic priming**.*

**Keywords:** Image of Brazil, Lexicogrammatical Patterns, Translation and Corpus Linguistics.



empresas na Europa, transformou em 1835 seu Bureau Havas de serviços de tradução na primeira agência de notícias internacional e, assim, teve início a troca de informações entre países e continentes.

O século XX testemunhou o desenvolvimento da mídia impressa pelo mundo. Com o advento e popularização da rede mundial de computadores nos anos 90, o jornalismo passou a ser um novo meio para divulgar os acontecimentos, imagens e opiniões. Logo surgiu, porém, uma dúvida sobre o futuro do jornal impresso: será que os leitores deixariam o jornalismo impresso em favor das versões virtuais? Essa questão refletia a crença de que o jornal impresso passaria a ser acessado *on-line*. O que é possível observar, até o presente momento, é que a versão virtual tem suas peculiaridades e dinâmicas próprias. Por exemplo, a atualização constante das notícias, a interação e participação instantâneas dos leitores faz do jornalismo *on-line* um universo muito mais ativo do que aquele das páginas impressas.

No mundo globalizado do século XXI, as agências de notícias investiam em suas filiais internacionais e muitos dos seus jornalistas desempenhavam também o papel de tradutores. Porém, essa notícia carrega uma marca: a visão do jornalista estrangeiro sobre o acontecimento. Será que essa visão estrangeira interessa apenas ao exterior? Será que o país fonte da notícia não se interessa por esse olhar externo? E não é difícil ter acesso pela imprensa *on-line* às notícias internacionais sobre temas nacionais.

As representações discursivas encontradas em textos sobre o Brasil publicados no e-jornalismo internacional em Inglês podem indicar a imagem do país criada a partir do ponto de vista estrangeiro. Entretanto, quando esses textos são traduzidos para o Português e publicados no e-jornalismo brasileiro, as formas de representação do Brasil parecem indicar padrões linguísticos diferentes e, conseqüentemente, imagens diferentes daquelas publicadas nos textos originais.

Assim, pretendo apresentar aqui exemplos de descrição dos padrões lexicogramaticais encontrados no *corpus* de estudo *Brasilis*<sup>2</sup> que indicam características imagéticas do Brasil na imprensa estrangeira. Tais padrões denomino como **Pré-ativação Imagética**.

---

<sup>2</sup> *Corpus* paralelo bilíngüe coletado entre 2000 e 2009 composto por 162 textos em Inglês publicados no exterior e suas traduções em Português.

## Traduzindo o Brasil para Brasileiros: Um Estudo de Caso do Jornalismo *On-line*

Foram diversos os momentos em que o Brasil ocupou a posição de protagonista na mídia internacional na primeira década do século XXI. A edição de 12 de novembro de 2009 da renomada revista sobre economia *The Economist* trouxe na capa a imagem do Cristo Redentor transformado em um foguete acompanhado do título *Brazil takes off*<sup>3</sup>. Nessa edição, quatorze páginas foram, prioritariamente, dedicadas à ascensão do país no cenário econômico mundial. As análises mostraram o sucesso das políticas adotadas pela equipe econômica do governo brasileiro, mas também faziam alertas em relação à falta de investimentos em áreas de infraestrutura, Saúde, Educação e à necessidade de combater a corrupção.

O problema geral que impulsionou este estudo gerou a seguinte questão: Qual era a imagem do Brasil no jornalismo estrangeiro veiculado pela *Internet* no início do século XXI? A maneira como o léxico é, tipicamente, usado em uma língua e a repetição desse léxico em grupos de palavras, associação semântica, podem contribuir para a criação de uma imagem daqueles que fazem parte do discurso. Há ainda a questão de como essa imagem se revela quando um texto sobre o Brasil é traduzido para o Português brasileiro e publicado na mídia brasileira *on-line* (UOL Notícias \_Jornais Internacionais<sup>4</sup>).

Assim, apresento a seguir um recorte do estudo sobre a imagem do Brasil na mídia *on-line* em Inglês e suas traduções para o Português. Nele, partirei de escolhas lexicogramaticais (Berber Sardinha, 2013) frequentes encontradas em uma dada coletânea de textos, *corpus*, na Língua Inglesa, comparadas com aquelas encontradas nas suas traduções para o Português, para apontar imagens do Brasil no jornalismo estrangeiro. A proposta é fazer um estudo metodológico à luz da Linguística de Corpus para analisar os padrões lexicogramaticais que possam sugerir como a imagem dos participantes do discurso se realiza, linguisticamente, na LF<sup>5</sup> e na LA<sup>6</sup> em textos jornalísticos. Ao conjunto desses padrões denominei **Pré-ativação Imagética**. Ao sugerir neste estudo a investigação de traços típicos indicativos da imagem do Brasil nas duas línguas, pretendi demonstrar que nessa comparação a imagem torna-se saliente e pode revelar como os estrangeiros retratam ou imaginam o Brasil quando reportam alguns temas como política, economia e meio-ambiente. Por imagem, entendo um conjunto de características externas e públicas que representam uma determinada ideia que se tem de algo ou de alguém.

<sup>3</sup> Disponível em < [http://www.economist.com/node/14845197?story\\_id=14845197](http://www.economist.com/node/14845197?story_id=14845197) >. Acesso em: 21 nov. 2009.

<sup>4</sup> Disponível em < <http://noticias.uol.com.br/jornais/> >. Acesso em: 12 out. 2011.

<sup>5</sup> Língua-fonte: doravante LF.

<sup>6</sup> Língua-alvo: doravante LA.

## 2. Fundamentação Teórica

Primeiramente, cabe esclarecer que, por não haver uma fundamentação teórica que alie estudos de imagem e *corpus*, apresento nesta etapa temas provenientes dos pressupostos da Linguística de Corpus<sup>7</sup>, com especial atenção à noção de padronização lexicogramatical, fundamental para a identificação dos elementos indicadores de imagem de extrema relevância para a metodologia que desenvolvi.

A LC pode ser definida como uma área que tem como cerne a exploração da linguagem baseando-se em relatos empíricos que trazem à tona aquilo que é, tipicamente, usado em uma determinada língua. Assim, é possível estudar-se as regularidades e padronizações da língua em uso indicativas da presença de características que não ocorrem de forma aleatória.

Todavia, tais ocorrências podem, de maneira sutil ou não, frequentar a língua em uso reiteradas vezes, fixando determinados atributos aos itens com os quais mais se associam.

A pesquisa aqui apresentada partiu de um conjunto de dados linguísticos textuais, coletados, criteriosamente, em duas línguas (LF e LA), e segue os parâmetros empíricos da LC, ou seja, a detecção de padrões linguísticos, suas probabilidades de usos e ocorrências, entre outros e compreendeu a análise de duas línguas: Inglês e Português. A Língua Inglesa é apresentada como LF e as traduções em Português como LA, lembrando que as possibilidades de ocorrência de traços linguísticos são infinitas, mas a relevância encontrou-se na ocorrência desses traços na língua em uso.

Há controvérsias quanto a LC: alguns teóricos a consideram uma metodologia enquanto outros a entendem como uma abordagem de apoio voltada ao desenvolvimento de pesquisas.

Segundo Hoey (1997 *apud* Berber Sardinha, 2004, p. 37), a LC “não é um ramo da linguística, mas a rota para a linguística”. Essa afirmação significa que a LC deve ser vista além de uma possibilidade metodológica de lidar com dados linguísticos, elevando-a à condição de abordagem.

Assim sendo, por meio da LC é possível desenvolver estudos que dizem respeito a diversos aspectos da linguagem como a sintaxe, o léxico ou o gênero textual e ainda associá-la a outros segmentos dos estudos da linguagem como a Análise do Discurso, a Pragmática e os Estudos de Tradução, por exemplo.

---

<sup>7</sup> Doravante LC.

De acordo com Biber *et. al* (1999, p. 5), pesquisas conduzidas a partir da abordagem de *corpus* possibilitam “[...] identificar e analisar “padrões de associação” complexos que são formas sistemáticas nas quais características linguísticas são usadas em conjunto com outras de caráter linguístico ou não linguístico”.<sup>8</sup>

Hunston e Francis (2000, p. 37) postulam que os padrões de uma palavra:

“[...] podem ser definidos como todas as palavras e estruturas com as quais são associados regularmente e que contribuam para seu significado. Se uma combinação de palavras ocorrer com relativa frequência, se for dependente de uma escolha específica e se houver um significado claro associado então, um padrão pode ser identificado.”<sup>9</sup>

Um dos primeiros estudiosos a definir um dos tipos de padrão, a colocação, foi o linguista britânico J. R. Firth. Com sua frase “deve-se julgar uma palavra por sua companhia”<sup>10</sup> (1957 *apud* Partington, 1998, p. 15), a colocação surge como uma nova maneira de estudar e entender os fenômenos lexicais.

A colocação é definida por Sinclair (1991, p. 170) como sendo “a ocorrência de duas ou mais palavras distantes num pequeno espaço de texto entre si.”<sup>11</sup> Para Hoey (2005), os estudos de Sinclair foram fundamentais para estudar-se a língua em uso considerando-se léxico e gramática em conjunto a partir do fenômeno denominado colocação.

Sinclair (1991) aplicou o conceito da Prosódia Semântica, usado, primeiramente, pela Escola Firthiana na descrição de fenômenos fonológicos, para explicar padrões de associações entre itens lexicais e aquilo que elas criam com relação à conotação, positiva, negativa ou neutra. Segundo Partington (1998, p. 66), a Prosódia Semântica está relacionada com o padrão formado entre itens lexicais e a conotação produzida, que pode ser positiva, negativa ou neutra.

Outro exemplo de estudo que segue a vertente dos padrões linguísticos: a Pré-ativação Lexical elaborada por Hoey (2005). De forma sucinta, é possível explicá-la como uma tendência que o usuário da língua tem de associar e guardar itens lexicogramaticais em sua mente a partir de encontros frequentes com eles. Assim, de acordo com o autor, cada usuário teria um concordanciador mental de cada item com o qual se encontra.

<sup>8</sup> “[...] to identify and analyze complex “association patterns”: the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features.”

<sup>9</sup> “[...] can be identified as all words and structures which are regularly associated with the word and which contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it.”

<sup>10</sup> “you shall judge a word by the company it keeps.”

<sup>11</sup> “[...] the occurrence of two or more words within a short space of each other in a text.”

Ainda sobre a Pré-ativação Lexical, procurei descobrir na pesquisa que relato se a recorrência de determinados itens em padrões lexicogramaticais constroem e mantêm a imagem não apenas de maneira individual e natural, mas de forma sistemática em uma população representada pelo *corpus*.

Portanto, na pesquisa, a Pré-ativação Lexical seria um processo de naturalização em que os encontros recorrentes com determinados itens lexicogramaticais tomam aquilo que não é óbvio em óbvio ou usual em um determinado período. Isso pode ser importante, por exemplo, para textos jornalísticos e publicitários, como ferramenta para instituir uma ideia ou imagem.

### 3. Metodologia

Primeiramente, apresento a definição para *corpus* adotada neste trabalho:

“Um conjunto de dados linguísticos (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 linguístico 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.” (Sanchez 1996, p. 8-9 *apud* Berber Sardinha, 2004, p. 18).

Além das características acima elencadas, o *corpus* da pesquisa foi classificado: especializado, por ser composto por um tipo específico de texto (Hunston, 2005), o jornalístico. Em segundo lugar, com relação a sua disposição interna, esse *corpus* apresenta-se organizado da seguinte forma: paralelo e alinhado. Diante disso, passo a denominar o *corpus* de estudo na LF e na LA como *corpus* de estudo paralelo. Quanto ao seu tamanho, se considerado o número total de palavras (291.643), ele pode ser classificado como médio, seguindo a classificação sugerida por Berber Sardinha (2004, p. 26). Além disso, é formado de dois *subcorpora*: um na LF e outro na LA. Finalmente, quanto à descrição do *corpus* de estudo paralelo, esclareço que as características dos dados são pertinentes a esse *corpus* e, portanto, não devem ser entendidas como uma generalização, aplicável a qualquer tipo de texto.

Os passos para a coleta foram:

1. Determinação do sítio que forneceria os textos na LA.;

2. Escolhido o sítio, passei a coletar os textos e armazená-los em dois formatos: html e txt;
3. Terminada a coleta dos textos na LA, busquei os textos na LF que se encontravam em diversos sítios internacionais como New York Times, The Financial Times, Boston Globe, USA Today. Os textos escritos na LF também foram armazenados nos formatos html e txt.

O sítio UOL Mídia Global foi escolhido por se apresentar, no período da coleta do *corpus*, como o único provedor que tornava possível o acesso aos textos jornalísticos assinados como traduções, por vezes acompanhadas do nome do tradutor ou como Tradução UOL, e pelo texto disponibilizado pelo provedor que aponta para a missão desse serviço:

“Em UOL Mídia Global o internauta pode ler diariamente textos de alguns dos mais importantes jornais dos EUA, como o "The New York Times", "USA Today", "Boston Globe", "Los Angeles Daily News", além dos europeus "Financial Times" (Inglaterra) [...]

As reportagens sobre o Brasil recebem atenção especial no UOL Mídia Global. Tudo para que os internautas conheçam a visão estrangeira sobre as questões nacionais, além de saber o que os formadores de opinião da Europa e Estados Unidos estão lendo sobre nosso país.”<sup>12</sup> (Da Redação UOL, 2001).

Assim, é possível concluir que o objetivo proposto pelo sítio colabora com o proposto na pesquisa: revelar a imagem do Brasil no exterior por meio de textos que tratassem de assuntos como economia, política, sociedade, entre outros.

Ressalto que o período inicial da obtenção dos textos apresentou alguns obstáculos, dentre eles o fato de que os textos escritos na LF e traduzidos na LA não apresentavam, necessariamente, a mesma estruturação. Sobre algumas dessas mudanças específicas por vezes encontradas na tradução jornalística, Bielsa e Bassnett (2009, p. 64) apontam para as seguintes categorias que as caracterizam:

- “Alteração de título e lead;
- Eliminação de informação desnecessária;
- Adição de informação secundária importante;
- Alteração na ordem de parágrafos;
- Resumo de informação.”

---

<sup>12</sup> Meu grifo.



Segundo as autoras, a justificativa para tais intervenções é a necessidade de fazer com que o texto jornalístico traduzido possa ser relevante e adequado ao contexto do público-alvo. No caso dos dados da pesquisa as categorias acima podem ser atestadas, em especial quanto a mudança de título e eliminação ou adição de informação na LA. A meu ver, porém, essas alterações podem não apenas representar características específicas do texto jornalístico traduzido, como também tornar evidentes as diferenças na maneira de retratar os assuntos na LF e LA.

Com tais temas em mente e com a expectativa de encontrar nessas traduções elementos lexicogramaticais que pudessem favorecer o estudo das escolhas feitas na LA, iniciei a coleta para a formação do *corpus* de estudo descrito da seguinte maneira:

Tabela 1: Dados estatísticos do corpus de estudo paralelo.

	Textos LF	Textos LA
Tokens (itens)	145.507	146.136
Types (formas)	13.109	15.263
Type/TokenRatio (razão forma/item)	9	10
Standardized Type/TokenRatio (razão forma/item padronizada)	46,61	46,78
Sentences (sentenças)	6.812	6.228

Fonte: Tabela extraída da tese da autora deste artigo - Lopes (2010, p. 100)<sup>13</sup>.

### Corpora de Referência

Com o propósito de comparar os dados do *corpus* de estudo paralelo da LF, recorri aos seguintes *corpora* de referência: British National Corpus<sup>14</sup> e Corpus of Contemporary American English<sup>15</sup> (Davies, 2008). A opção por dois *corpora* de referência deve-se ao fato de que são *corpora* complementares.

Com relação à descrição das características do primeiro *corpus* de referência:

-BNC: em CD-ROM; 100 milhões de palavras (Inglês Britânico, escrito e falado);

-COCA: disponível, gratuitamente, na *Internet*, 385 milhões de palavras (Inglês Americano, escrito e falado).

Para o cotejo dos dados na LA do *corpus* de estudo paralelo, fiz uso do Banco de Português v.2<sup>16</sup>: versão em CD-ROM<sup>17</sup> (Projeto DIRECT)<sup>18</sup>; 700 milhões palavras (Português Brasileiro escrito e falado).

<sup>13</sup> Disponível em: < [http://www.sapientia.pucsp.br/tde\\_busca/arquivo.php?codArquivo=12198](http://www.sapientia.pucsp.br/tde_busca/arquivo.php?codArquivo=12198) >.

<sup>14</sup> Doravante BNC.

<sup>15</sup> Doravante COCA.

<sup>16</sup> Doravante BP2.

<sup>17</sup> Essa versão é a mais atualizada do Banco de Português e está disponível, por enquanto, apenas para a consulta dos alunos do programa em LAEL da PUC de São Paulo.

<sup>18</sup> Direct\_Em Direção à Linguagem do Trabalho. PUCSP Disponível em: < <http://www2.lael.pucsp.br/direct/> >. Acesso em 3 jan. 2010.

## Ferramentas

O programa WordSmith<sup>®</sup> 4 (Scott, 2004) foi usado para processar os dados do *corpus* paralelo de estudo e, por vezes, os dos *corpora* de referência. Fiz uso das seguintes ferramentas do programa: WordList<sup>®</sup>, KeyWords<sup>®</sup> e Concord<sup>®</sup>.

A primeira ferramenta, WordList<sup>®</sup>, foi utilizada para processar a lista de palavras do *corpus* paralelo de estudo. O programa gera uma lista na qual cada palavra é apresentada, separadamente, ou em agrupamentos de duas ou mais palavras, *clusters*. Escolhi a primeira possibilidade, palavras apresentadas em separado, para a análise linguística.

A segunda ferramenta, KeyWords<sup>®</sup>, produz listas de palavras-chave do *corpus* de estudo. Palavras-chave são palavras relevantes no *corpus* que se pretende estudar, processadas a partir de critérios de significância estatística aplicados pelo programa, que compara as palavras do *corpus* de estudo com as de um *corpus* de referência. As palavras-chave são calculadas pela comparação entre a frequência de cada item na lista de palavras do *corpus* de estudo e a frequência desse item na lista de palavras do *corpus* de referência.

A lista obtida apresentou 500 palavras-chave na LF e, como salienta Berber Sardinha (1999), faz-se necessária uma seleção das palavras por parte do pesquisador que torne viável seu estudo. Ademais, como o objetivo da pesquisa era a descrição de padrões lexicogramaticais que levem à imagem do Brasil retratada nos textos jornalísticos do *corpus* de estudo, fiz um recorte nos dados obtidos na lista de palavras-chave que pudesse indicar essa imagem. Assim, extraí da lista de palavras-chave os dez (10) primeiros participantes do discurso (substantivos próprios e comuns) na LF, que chamo de **participantes-chave**, para iniciar o estudo das características dos padrões lexicogramaticais que envolvem esses atores sociais.

A terceira e última ferramenta utilizada na pesquisa, Concord<sup>®</sup>, foi escolhida por processar listas de concordâncias. Para tanto, é necessário escolher-se um termo de busca a ser localizado nos dados, que é apresentado de forma centralizada, cercado pelo seu contexto. Essa lista de concordâncias é essencial para o estudo dos padrões lexicogramaticais, pois permite a leitura do termo de busca em todas as suas ocorrências e também daquilo que ocorre em suas cercanias, os chamados Collocates (colocados). O Concord<sup>®</sup> apresenta a lista de concordâncias e a lista dos colocados da palavra de busca, com suas frequências e indicação da posição desse colocado (1<sup>a</sup>. à direita, 1<sup>a</sup>. à esquerda, e assim por diante).

A aplicação das ferramentas Concord<sup>®</sup> e Collocates na pesquisa teve por objetivo viabilizar o estudo dos padrões lexicogramaticais de determinados itens tanto no *corpus* paralelo de estudo quanto nos *corpora* de referência. É importante frisar que tanto o *corpus* de estudo paralelo quanto os *corpora* de referência (BNC e BP2) estão armazenados em formatos compatíveis com o programa WordSmith<sup>®</sup> 4.

Escolhidas as ferramentas computacionais, faz-se necessária uma síntese dos termos utilizados na pesquisa por serem inéditos e por fazerem parte essencial da metodologia sugerida para a descrição das características imagéticas.

O primeiro deles é o **participante-chave** (substantivo próprio ou comum): representa um ator social do discurso cuja frequência no *corpus* de estudo justifica-se pelas características da ferramenta KeyWords<sup>®</sup>.

O segundo, os **colocados-chave**: itens que atribuem aos participantes-chave alguma característica e ocorrem em uma janela 5:5.

Os elementos indicadores de imagem são representados na pesquisa pelos colocados-chave que atribuem ao **participante-chave** algum tipo de característica específica. Contudo, outros recortes são possíveis a partir de critérios adotados pelo analista interessado na descrição da **Pré-ativação Imagética**. Por exemplo, análise apenas de substantivos comuns da lista de palavras, WordList<sup>®</sup>, e de colocados representados apenas por adjetivos.

## Metodologia Para Análise da PI

A elaboração de uma metodologia para análise de dados não se constitui em tarefa fácil. Em estudos anteriores pude testar diversos caminhos até chegar à sistematização das etapas descritas abaixo:

- Formação de um *corpus* paralelo alinhado:
  - 1- Coleta e armazenamento os dados paralelos (LF e LA);
  - 2- Alinhamento do *corpus* paralelo (LF e LA).
- Processamento dos dados:
  - 3- Processamento do corpus na LF via WordSmith Tools<sup>®</sup> 4 (Scott, M. 2004) da lista de palavras-chave: participantes-chave<sup>19</sup>;
  - 4- Seleção dos PC: os PC são os atores sociais do discurso cuja PI deseja-se descrever;
  - 5- Processamento de elementos indicadores de imagem<sup>20</sup> de cada PC, concordâncias e colocados: os colocados-chave<sup>21</sup>;

---

<sup>19</sup> Doravante PC.

<sup>20</sup> Doravante EIIM.

<sup>21</sup> Doravante CC.

6- Seleção dos CC válidos que evidenciam a PI do PC: CC válidos, atribuem ao PC alguma característica específica. Caso algum dos CC válidos seja também um PC, ele será excluído da lista de CC para que não ocorra circularidade nos dados. Por ser uma pesquisa de cunho metodológico, toda palavra é candidata a análise dependendo do analista; algumas candidatas a PC e CC foram escolhidas aqui para demonstração do método; para se ter uma visão rica não é necessária uma análise completa, toda a lista, mas sim profunda;

7- Agrupamento dos CC válidos em relação ao PC na LF: nessa fase, a lista de CC já pode indicar os grupos de associações semânticas indicativas da PI. Se esse for o caso passa-se à etapa dez (10) abaixo;

8- Colocados do CC no *Corpus* de Referência<sup>22</sup> na LF para desambiguação de sentido, caso não esteja claro: se a lista de CC válidos não deixar clara a imagem, buscar pelos colocados do CC no CR na LF para chegar às associações semânticas indicativas da PI. Por vezes, a busca no CR dos CC na LF não ocorre nessa etapa, mas apenas no confronto com a LA, etapa onze (11).

- Contraste entre LF e LA para:

9- Buscar os pares tradutórios, metodologicamente válidos, de cada CC no Concordanciador Paralelo: o CC é a palavra de busca e o PC é o termo de seleção, poderá estar em um horizonte maior ou menor que cinco (5). Assim, as concordâncias paralelas do CC em que o PC ocorre são obtidas. Porém, deve-se ainda dessa lista selecionar as ocorrências em que o PC e o CC aparecem em um horizonte de cinco (5) à esquerda e cinco (5) à direita, característica para colocados;

10- Analisar os pares tradutórios válidos de cada CC e buscar os CC no CR na LA, se necessário (e no CR na LF se não tiver sido feito na etapa nove (9): no contraste entre LF e LA, iniciado na etapa anterior, já é possível definir os grupos semânticos que levam à PI do PC. A etapa nove (9) acima é repetida para as traduções dos CC válidos na LA, caso a imagem não esteja clara;

11- Analisar a PI do PC na LF e LA, a partir das associações semânticas obtidas.

A seguir, para exemplificar a aplicação das etapas acima, apresento um recorte da análise dos dados com a análise de um dos participantes-chave (PC), *Amazon*.

---

<sup>22</sup> Doravante CR.

#### 4. Análise

As quatro primeiras etapas foram aplicadas considerando-se todo o *corpus* de estudo paralelo.

##### **Etapa 1: Coleta e Armazenamento dos Dados Paralelos (LF e LA)**

Como descrito anteriormente, foram coletados dados de sítios jornalísticos que disponibilizam textos em Inglês (LF) sobre o Brasil e, a partir do sítio UOL Mídia Global, as traduções desses textos para o Português (LA).

##### **Etapa 2: Alinhamento do Corpus Paralelo (LF e LA)**

As etapas foram: a codificação do *corpus* de estudo paralelo anterior ao envio ao Alinhador *on-line* de textos paralelos (Vanilla Aligner) (CEPRIL, LAEL, PUCSP)<sup>23</sup>, o tratamento dos arquivos na LF e na LA e a posterior inserção no sítio do alinhador.

##### **Etapas 3 e 4: Processamento da Lista de Palavras-Chave, Doravante Participantes-Chave (PC) e Seleção dos PC**

Apresento na **tabela 2** abaixo os participantes-chave na LF, fruto do processamento feito com a ferramenta WordSmith Tools<sup>®</sup> 4. Os participantes-chave na LA também se encontram presentes, pois serviram como suporte para as etapas de contraste entre LF e LA para a análise da PI.

Tabela 2: Participantes-chave na LF e LA.

N	Participante-Chave (LF)	Freq.	Participante-Chave (LA)	Freq.
1	BRAZIL	738	BRASIL	956
3	LULA/ LUIZ/ INÁCIO/ SILVA	379	LULA/ LUIZ/ INÁCIO/ SILVA	439
9	SÃO (PAULO)	124	-	
12	BRAZILIANS	121	BRASILEIROS/ BRASILEIRAS	214/66
13	RIO (de JANEIRO)	103	-	
14	AMAZON	145	AMAZÔNIA	93
18	ALCKMIN	75	ALCKMIN	86

Fonte: Tabela extraída da tese da autora deste artigo - Lopes (2010, p. 103 e 119).

Após definir o padrão lexicogramatical a partir das colocações e das concordâncias do item pesquisado, o sentido é dado pela ideia a que o padrão remete. Apresento exemplos dos colocados do CC.

<sup>23</sup> Disponível em< <http://www2.lael.pucsp.br/corpora/alinhador/index.html>>. Acesso em 28 abr. 2007.

A partir da etapa cinco, a análise é focada em cada PC. A título de ilustração, apresentarei os padrões lexicogramaticais encontrados junto a alguns dos itens estudados na LF e na LA para o PC *Amazon* indicativos de características imagéticas no *corpus* deste estudo.

**Etapa 5: Processamento da Lista de Concordâncias e de Colocados de Cada PC (Colocados-Chave)**

A partir do Concordancer, acessei a lista de 144 colocados de *Amazon*. Alguns desses colocados-chave (CC) são:

Tabela 3: Exemplo da lista de CC do WordSmith® 4 para Amazon.

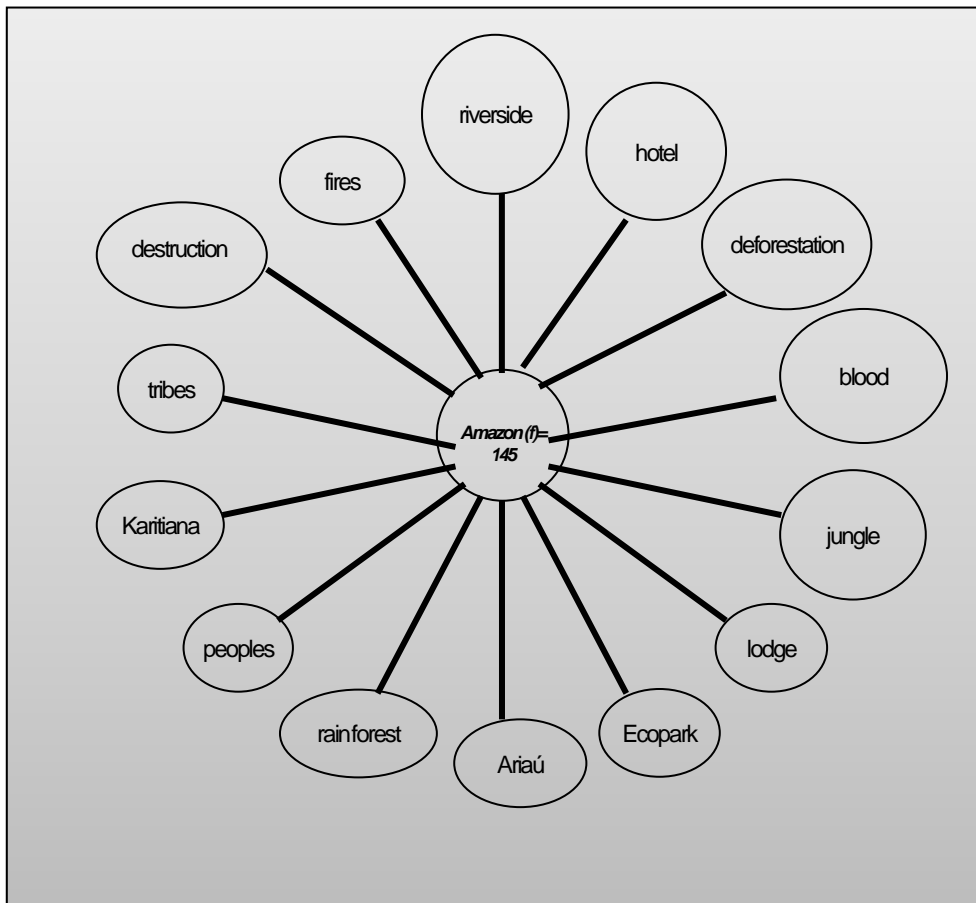
	Word
1	JUNGLE
4	RAINFOREST
5	FOREST
5	DEFORESTATION
1	BLOOD

Fonte: Lopes (2010).

**Etapa 6: Seleção dos CC Válidos que Evidenciam a PI do PC *Amazon***

O diagrama 1 traz ao centro o PC *Amazon*, sua frequência (f) no CE na LF e, ao seu redor, os CC. A análise dos padrões lexicogramaticais formados pelo PC e CC é indicativa da imagem desse ator social no *corpus* da pesquisa.

Diagrama 1: Seleção dos CC válidos que possam indicar a PI do PC Amazon.

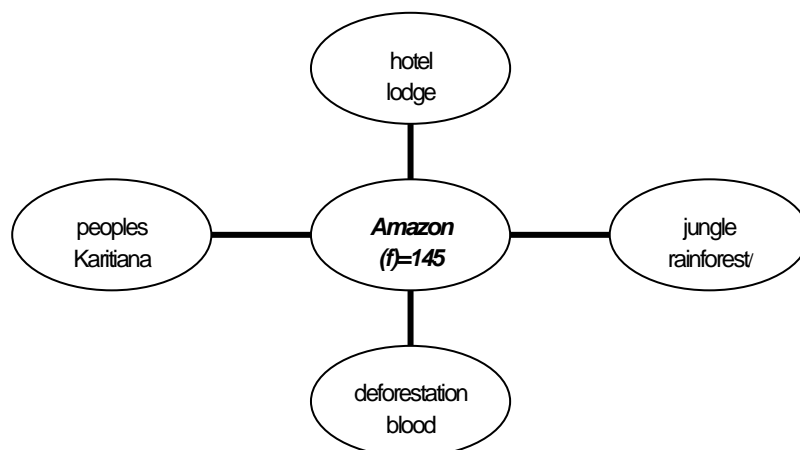


Fonte: Lopes (2010, p. 128).

### **Etapa 7: Agrupamentos dos CC Referentes ao PC na LF**

Para agrupar os CC válidos em relação ao PC na LF, escolhi separá-los em grupos por fazerem parte do mesmo grupo semântico. Obtive a configuração representada no diagrama 2:

Diagrama 2: Agrupamentos dos CC referentes ao PC Amazon.



Fonte: Lopes (2010, p. 128).

### Etapa 8: Colocados do CC no CR na LF para Especificação de Sentido

Foi necessário verificar no CR os CC *jungle*, *rainforest* e *rain forest*, pois parecem ter sido usados como formas sinônimas para o PC *Amazon*. Os resultados para *jungle* foram:

Tabela 4: Associação semântica para jungle no Corpus de Referência BNC.

Associação semântica CC jungle (925 vezes no Corpus de Referência BNC)			
Exemplos de colocados do CC no BNC	No. de ocorrências	Sentido do grupo semântico (CC e seus colocados)	Grupo semântico (CC e seus colocados)
trees	12	natureza	nomes
river(s)	7		
wild	10	natureza	atributos
law (of the)	24	metáfora da dificuldade (prosódia semântica negativa)	nome
concrete	16	metáfora da cidade (prosódia semântica negativa)	atributo
film	12	cultura	nomes

Fonte: Lopes (2010, p. 129).

Cada um desses grupos foi estudado a partir de suas concordâncias. Alguns exemplos:

jungle + natureza (nomes) / natureza (nomes) + jungle

1. skyscrapers like banded snakes between jungle trees. On Eighth Street the wind was N

natureza (atributos) + jungle

1. country, exotic plants and wild jungle still unspoilt by the

metáforas da dificuldade (prosódia semântica negativa) (nome) + jungle

1. Similarly, "the law of the jungle", meaning, in fact, the absence of any

atributo (metáforas da cidade com prosódia semântica negativa) + jungle

1. of life in the concrete jungle, and its jaundiced appraisal of

cultura (nomes) + jungle / jungle + cultura (nomes)

1. the film about combat in the jungle, which they left before



Em um primeiro momento, é possível constatar que *jungle* se associa a itens lexicais que expressam características, a outros que refletem uma imagem ligada a elementos culturais e, por último, àqueles que indicam um sentido figurado, metáfora, imprimindo em *jungle* um caráter romântico e subjetivo.

### **Etapa 9: Busca Pelos Pares Tradutórios, Metodologicamente Válidos, de cada CC no Concordanciador Paralelo**

Os CC me levaram à busca pelos pares tradutórios, metodologicamente, válidos no Concordanciador Paralelo, ocorrências de cada CC na LF e suas traduções para a LA. Os resultados obtidos foram:

Tabela 5: Exemplos de pares tradutórios metodologicamente válidos para Amazon.

Pares tradutórios metodologicamente válidos				
CC (LF)	No. de Ocorrências	CC (LA)	No. de Ocorrências	Grupos semânticos indicam sentido de
jungle	5	floresta	4	natureza
		selva	1	
rainforest	7	floresta	3	
		Amazônia	2	
		floresta tropical	1	
rain forest	3	sem tradução	1	
		floresta	3	

Fonte: Lopes (2010, p. 131).

### **Etapa 10: Análise dos Pares Tradutórios Válidos de Cada CC e Busca dos CC no CR na LA, Se Necessário (e no CR na LF Se Não Tiver Sido Feito na Etapa 8)**

*Jungle* foi traduzido de diferentes formas, Por exemplo:

LF	LA
You are, after all, in the heart of the Amazon jungle, and	Afinal, você está no meio da selva amazônica, e
attempt to conserve the Amazon jungle — or the leading	tentativa de conservar a Floresta Amazônica ou em um

Fonte: Lopes (2010).

Busquei, então, descobrir se floresta e selva indicam o mesmo sentido na LA e, comparativamente, se são itens que satisfazem o sentido da LF. Para floresta:

Tabela 6: Associação semântica para floresta no Corpus de Referência Banco de Português V. 2.

Associação semântica CC floresta (4.000 vezes no Corpus de Referência Banco de Português V.2.)			
Exemplos de colocados do CC no BP2	No. de ocorrências	Sentido do grupo semântico (CC e seus colocados)	Grupo semântico (CC e seus colocados)
atlântica Amazônica tropical amazônica	397 221 194 260	características da floresta	atributos
Brasil Amazônia sul	96 73 22	localização	nomes

Fonte: Lopes (2010).

Alguns exemplos são:

floresta + características da floresta (atributos)

1. deste estudo indicam que a floresta Atlântica montana regenera-se dentro dos

floresta +localização (nomes)

- 1.composição florística em remanescentes de floresta decídua no Brasil. O presente

Os exemplos extraídos do CR\_BP2 mostram que floresta associa-se a itens lexicais cujos grupos semânticos descrevem a natureza de determinado lugar.

Alguns exemplos para cada grupo semântico das linhas de concordância:

selva + características da selva (atributos)

1. as dificuldades de trabalhar na selva amazônica ou a epidemia de

presença militar (nomes) + selva

1. A Brigada de Infantaria de Selva é comandada a partir de do

Tabela 7: Associação semântica para selva no Corpus de Referência Banco de Português v. 2.

Associação semântica CC selva (648 vezes no Corpus de Referência Banco de Português v.2)			
Exemplos de colocados do CC no BP2	No. de ocorrências	Sentido do grupo semântico (CC e seus colocados)	Grupo semântico formado pelo CC e seus colocados
Amazônica/amazônica	77	características da selva	atributos
Infantaria	31	presença militar	nomes

lei (da)	27	metáforas da dificuldade	nome
(de) pedra	7	metáfora da cidade	atributo
guerra	21	ameaças à vida humana	nomes
sobrevivência	8	condição de vida	nome

Fonte: Lopes (2010, p. 133).

Alguns exemplos para cada grupo semântico das linhas de concordância são:

selva + metáfora da cidade (prosódia semântica negativa) (atributo)

1.tê -lo em São Paulo, na selva de pedra, deixando de comer o seu feijão

ameaças à vida humana (prosódia semântica negativa) (nomes) + selva /

selva + ameaças à vida humana (prosódia semântica negativa) (nomes)

1.destreza e habilidade na guerra na selva, em todas as suas

condição de vida (prosódia semântica negativa) (nome) + selva

1.particularidades como sobrevivência no mar e na selva, prática de tiro e viagens

Os exemplos acima demonstram que selva associa-se a itens lexicais que evidenciam características do ponto de vista da sua descrição, da presença do homem e sua consequência desfavorável, elementos militares, doenças e sobrevivência, e de metáforas que retratam imagens áridas.

### **Etapa 11: Análise da PI do PC Amazon na LF e LA**

A PI de *Amazon* na LF e LA revela a imagem igual quando trata de povos nativos e turismo, fazendo uma descrição de caráter informativo.

Há uma diferença quando trata de lugar/meio-ambiente. Na LF, *jungle* imprime uma imagem às vezes real, às vezes subjetiva. Por um lado, *jungle* é um lugar da natureza com atributos que a descrevem; por outro, é uma criação humana usada para expressar uma metáfora de sobrevivência, bem como um lugar literário.

Já com *rainforest/rain forest*, *Amazon* revela uma dicotomia: PI descritiva (tipo e localização) e negativa (alerta dos riscos da destruição). Cabe ressaltar que a diferença ortográfica pode produzir diferentes

associações, de acordo com Hoey (2005, p. 27). No caso dos itens em questão, todavia, os grupos são os mesmos e muitos dos colocados se repetem.

Na LA, tanto *jungle* quanto *rainforest/rain forest* foram traduzidas na maioria das vezes por floresta, e isso implica em uma PI diferente por ser, puramente, descritiva na LA, em especial associado a uma descrição botânica, sem a imagem metafórica de *jungle* nem a de alerta aos danos de *rainforest/rain forest*.

Portanto, concluo que selva é o melhor par tradutório para *jungle*, já que ambos remetem a dois dos três grupos semânticos encontrados, natureza e metáfora.

Para *rainforest/ rain forest*, o par tradutório floresta indica a descrição da natureza, mas não o alerta aos danos. Em uma pesquisa no *Corpus* de Referência Banco de Português v. 2, o item Amazônia (4.000 ocorrências) coloca-se com itens negativos como desmatamento, problemas e outros. Sugiro que o par tradutório para a LF seja Amazônia para alcançar a imagem de *rainforest/ rain forest*.

De uma maneira geral, fica claro que alguns aspectos da imagem do estrangeiro sobre *Amazon* traz consigo padrões lexicogramaticais diferentes. Acredito que tal fato ocorra em decorrência do distanciamento do estrangeiro e da proximidade geográfica de muitos brasileiros a esse ambiente, o que o faz parte do imaginário e da realidade nacional desde o início da colonização do território brasileiro.

A ausência de tradução de uma frase ou palavra pode, mesmo que de maneira sutil, alterar a PI pretendida na LF.

Outras ausências na LA, como *Ariaú* e *lodge*, podem ter decorrido do fato de que o texto traduzido omite, aproximadamente, onze parágrafos escritos no original na LF, todos eles relativos a informações turísticas como nomes de hotéis, pousadas, vacinação e o melhor período para se visitar a Amazônia. A omissão das informações não permitirá ao leitor da LA ter acesso aos mesmos dados turísticos que o leitor da LF. Por outro lado, também é possível supor que a linha editorial da LA tenha optado por suprimir tais informações por considerá-las um tipo de propaganda para os estabelecimentos mencionados na LF.

Finalmente, concluo que a imagem de *Amazon* nas duas línguas é semelhante, mas não igual, uma vez que os nomes usados para descrevê-la têm diferentes nuances de sentido e imagem que devem ser consideradas quando da descrição dos padrões, como no caso desta pesquisa bem como quando das escolhas tradutórias.

## 5. Considerações Finais

Investigar a imagem do Brasil em textos jornalísticos escritos na LF e traduzidos para a LA possibilitou descobrir como o Brasil se revela para os leitores de textos traduzidos.

O aporte teórico principal na pesquisa, a LC, a partir da concepção de linguagem como sistema probabilístico, mostrou-se essencial para que eu investigasse e interpretasse os padrões lexicogramaticais dos itens propostos com dados autênticos usados tanto na língua-fonte como na língua alvo. A presença constante de outros tantos itens que caracterizam, linguisticamente, a forma como determinado ator social é entendido e representado, socialmente, apoiou-se nos padrões lexicogramaticais a partir do conceito da colocação e das associações semânticas por meio da Pré-ativação Lexical. Esse último deu origem ao conceito que sugiro chamar de **Pré-ativação Imagética** (PI) e defino como um fenômeno linguístico explícito ou não, mas observável a partir da análise em *corpora* de palavras que, frequentemente, se encontram com outras, culminando em uma imagem representada, socialmente, pelos textos veiculados em comunidades discursivas. Assim, procurei apresentar no transcorrer da investigação que marcas lexicogramaticais indicam imagens identificáveis em estudos sistemáticos com *corpora* comparáveis e/ou paralelos, pois no contraste entre língua-fonte e língua-alvo as mudanças dessas imagens tornam-se visíveis, mesmo que sutis.

Os dados revelados pela pesquisa completa conduzida em quatro anos auxiliaram a descobrir se os padrões lexicogramaticais encontrados junto a alguns dos itens estudados nos textos originais em Inglês desvelavam características da imagem do Brasil na imprensa internacional. As imagens foram reveladas pelas categorias obtidas a partir dos conjuntos temáticos formados pelos participantes-chave, os colocados-chave e, por vezes, pelos colocados dos colocados-chave nos estudos das associações semânticas. Portanto, o estudo por meio da **Pré-ativação Imagética** possibilita dizer que a imagem é de um país naquela época (2006-2010):

- Cujas economia estava em crescimento com indícios de estabilidade;
- Visto de forma superlativa pelo estrangeiro que, até então, parecia desconhecer seu potencial e dimensão;
- Que precisava estar alerta quanto aos problemas encontrados na natureza, em especial na região Amazônica, inclusive com seus habitantes nativos, que, ao mesmo tempo, era apresentada a partir do seu potencial para o turismo;
- Um país cujos habitantes que imigraram em busca de oportunidades no exterior buscavam retornar graças às condições econômicas estáveis;

- Cujas grandes cidades eram descritas de modo superlativo em relação a aspectos socioeconômico e demográfico;
- Cujas correntes políticas representadas por determinados atores sociais indicavam ideologias políticas nem sempre consistentes;
- Cujos políticos eram apresentados a partir de determinados aspectos pessoais, políticos e de forma de governança.

Além das contribuições pretendidas, sem dúvida há outras questões que ficarão para as próximas etapas e que poderão contribuir para o aprimoramento da pesquisa. Por exemplo, estudos comparativos entre *corpora* de textos jornalísticos sobre o Brasil advindos de veículos de comunicação com linhas ideológicas diferentes. Tais estudos poderiam verificar como os padrões lexicogramaticais realizam a linha editorial dos meios de comunicação pesquisados. Ou ainda, a análise da imagem do Brasil por meio da **Pré-ativação Imagética** em um *corpus* jornalístico diacrônico monolíngue em Inglês. Esse estudo teria como objetivo descobrir se a metodologia do *corpus* paralelo poderia ser adaptada para identificar Elementos Identificadores de Imagem em *corpora* não paralelos, mais especificamente, em um *corpus* de textos sobre o Brasil publicados no exterior.

Recebido em: 03/2015; Aceito em 06/2015.

### Referências Bibliográficas

- BERBER SARDINHA, T. 1999. *O banco de palavras chave*. DIRECT Papers 39. Disponível em < <http://www2.lael.pucsp.br/direct/DirectPapers39.pdf> >. Acesso em: 10 jan. 2002.
- \_\_\_\_\_. 2004. *Linguística de Corpus*. São Paulo: Editora Manole.
- \_\_\_\_\_. 2013. *Lexicogrammar*. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 3365–3370). Hoboken, NJ: Wiley.
- BIELSA, E. BASSNETT S. 2009. *Translation in Global News*. London/New York: Routledge.
- DAVIES, M. *Corpus of Contemporary American English: 360 million words, 1990-present*. Disponível em < <http://www.americancorpus.org> >. Acesso em: 10 abr. 2008.
- DAVIES, M. FERREIRA, M. *Corpus do Português: 45 milhões de palavras, 1300-1900*. Disponível em < <http://www.corpusdoportugues.org/> >. Acesso em: 15 jul. 2008.
- HALL, J. In FRANKLIN, B. (ed.) Pulling Newspapers Apart: *Analysing Print Journalism*. p. 204-211.
- HOEY, M. 2005. *Lexical Priming: a new theory of words and language*. New York: Routledge.

- HUNSTON, S. FRANCIS, G. 2000. *Pattern Grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam e Philadelphia: John Benjamins.
- OLOHAN, M. 2004. *Introducing Corpora in Translation Studies*. Oxfordshire: Routledge.
- PARTINGTON, A. 1998. *Patterns and Meanings - Using Corpora for English Language Research and Teaching*. Studies in Corpus Linguistics 2. Amsterdam/Philadelphia: John Benjamins.
- SINCLAIR, J.M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- SCOTT, M. 2004. *WordSmith Tools version 4*. Oxford: Oxford University Press.

**Referências Bibliográficas *On-line*:**

- ALINHADOR ONLINE DE TEXTOS PARALELOS** (Vanilla aligner). \_CEPRIL, LAEL, PUC/SP. Disponível em < <http://www2.lael.pucsp.br/corpora/alinhador/index.html> > Acesso em: 28 abr. 2007.
- CONCORDANCIADOR PARALELO PUC/SP, LAEL, CEPRIL, DIRECT**. Disponível em < <http://www2.lael.pucsp.br/corpora/parallelconc/index.html> >. Acesso em: 28 nov. 2007.
- UOL Mídia Global**. Disponível em <<http://noticias.uol.com.br/>> Acesso em: 09 jan. 2008.
- \_\_\_\_\_. **UOL Mídia Global traz em português o melhor do jornalismo internacional**, 2001. Disponível em < <http://sobre.uol.com.br/ultnot/ novidade/ noticias/ ult98u231.jhtm> > Acesso em: 09 jan. 2008.

*Maria Cecília Lopes has an MA and a PhD in Applied Linguistics from PUC-SP. She teaches Translation at FMU; teaches at COGEAE-PUC-SP. She is the author of Compreensão Oral em Língua Inglesa (2008, IESDE) and Tecnologias e Mídias no Ensino de Inglês: o corpus nas receitas. (2012, Macmillan). E-mail: [cecilialopes@uol.com.br](mailto:cecilialopes@uol.com.br)*