

# Questões metodológicas de análise de metáfora na perspectiva da linguística de corpus

Tony Berber Sardinha

Recebido 28, fev. 2009/ Aprovado 1, abr. 2009

## Resumo

*Neste trabalho, são apresentadas maneiras de aplicar a Linguística de Corpus aos estudos de metáfora, com ênfase em questões metodológicas. São enfocadas duas metodologias: corpus-based e corpus-driven. Em relação à metodologia corpus-based, é mostrado o procedimento de seleção de candidatos e de uso do concordanciador. Em relação à metodologia corpus-driven, são discutidos os programas KeyWords e o Identificador de Metáforas.*

**Palavras-chave:** *Linguística de Corpus. Metáfora. Metodologias.*

## 1. Introdução

Hoje em dia, muitos já ouviram falar ou tiveram contato com a Linguística de Corpus, direta ou indiretamente, seja por meio de publicações (SINCLAIR, 1991; MCENERY e WILSON, 1996; BERBER SARDINHA, 2004), encontros científicos, livros didáticos de línguas (com suas atividades sobre ‘chunks’ e ‘collocations’) ou bons dicionários de inglês (Collins, Longman, Cambridge, Oxford, MacMillan, etc.), que via de regra são feitos com base em corpora. Segundo Berber Sardinha (2004:3):

‘A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.’

No Brasil, a Linguística de Corpus cresce cada vez mais, conforme o mapeamento apresentado por Berber Sardinha e Barcellos Almeida (2008).

Porém, a pesquisa em metáfora com base na Linguística de Corpus parece ser menos difundida, tanto no Brasil quanto no exterior. Este artigo tem como objetivo tentar preencher parte dessa lacuna, mostrando algumas maneiras pelas quais a Linguística de Corpus já se articula com os Estudos da Metáfora e relatar alguns procedimentos usados por seus praticantes para estudar metáforas em corpora eletrônicos. O foco deste trabalho é, mais precisamente, na metodologia de identificação de metáforas. A escolha desse foco se dá por duas razões principais. A primeira é que Ray Gibbs, um dos mais respeitados pesquisadores da área de metáfora, na sua introdução ao compêndio ‘The Cambridge Handbook of Metaphor and Thought’, que dá um panorama geral da área de metáfora atual, chega à seguinte conclusão:

‘Em geral, espera-se que haja um interesse crescente em questões metodológicas voltadas à definição da existência de metáfora na linguagem e no pensamento. Meu instinto diz que resoluções de alguns dos debates teóricos sobre metáfora e pensamento dependerão do modo pelo qual os estudiosos respondem a essas questões metodológicas. (...) Um desafio futuro para todos os estudiosos de metáfora é ter mais clareza sobre o tipo de evidência empírica que é necessária, e sobre como deve ser obtida e analisada, a fim de caracterizar de modo apropriado o alcance e os limites da mente metafórica.’ (GIBBS, 2008:13)<sup>1</sup>

A segunda justificativa é que a questão metodológica é chave para a Linguística de Corpus de um modo geral (MCENERY e WILSON, 1996), visto que corpora informatizados impõem a necessidade de desenvolver metodologias também informatizadas para extração de informação, caso contrário fica inviável abordar um corpus confiavelmente, por uma questão de escala dos dados. Ou seja, a Linguística de Corpus está em

<sup>1</sup> In general, there is likely to be a heightened interest in methodological questions for defining the existence of metaphor in language and thought. My hunch is that resolutions to some of the theoretical debates about metaphor and thought will partly depend on the way scholars respond to these methodological concerns. (...) A future challenge for all metaphor scholars is to have greater clarity about what kinds of empirical evidence is needed, and how it is to be obtained and analyzed, to properly characterize the reach and limits of the metaphorical mind.

constante busca por métodos de análise de corpora, para os mais diversos propósitos. Assim, ao enfocarmos questões metodológicas do estudo de metáfora em corpora, respondemos ao mesmo tempo ao desafio colocado por Gibbs e pela própria Linguística de Corpus.

Como há muitas questões metodológicas envolvidas na análise de corpora, decidimos promover um recorte na literatura e focar as seguintes questões norteadoras: (1) Quais são alguns dos procedimentos existentes para identificação de metáfora no discurso e como são implementados? (2) Quais são alguns dos procedimentos de identificação de metáfora em corpora e como são implementados? Justificamos nossa escolha por examinar procedimentos de identificação de metáfora no discurso por acreditarmos que (a) a análise de corpora muitas vezes pressupõe análise de discurso, e (b) procedimentos de identificação de metáfora no discurso auxiliam no *juízo* e na *nomeação* de metáforas, seja em corpora ou não. Esses pontos serão abordados mais detalhadamente no restante do trabalho.

## 2. Identificação de metáfora em corpus

Quando se fala em corpora eletrônicos atuais, é preciso ter em mente a questão do volume de dados, e como esse volume muda a relação do pesquisador com seus dados. Geralmente, os corpora são maiores do que a capacidade do analista de ler e julgar cada possível caso de metáfora. Hoje em dia, com a ampla disponibilidade de material textual em formato de computador, compilar corpora que ultrapassam os milhões de palavras é algo relativamente simples e rápido. Corpora de língua geral tipicamente ultrapassam as centenas de milhões de palavras; alguns já chegam a um bilhão ou mais, e a tendência é que continuem a crescer. Corpora especializados, formados por gêneros ou registros específicos, bem como corpora de fala, são geralmente menores, devido às dificuldades de coleta e/ou transcrição de dados. Desse modo, o pesquisador se vê forçado a retirar (*retrieve*) instâncias de metáfora do seu corpus sem que tenha lido todo o corpus.

Devido a isso, é preciso eleger *candidatos*<sup>2</sup>, que são palavras ou grupos de palavras que possivelmente fazem parte de uma metáfora linguística. Por exemplo, para chegar à metáfora linguística 'passou um tempo', seriam candidatos 'passou', 'um' e 'tempo', isto é, todas as palavras constantes da metáfora. Ou seja, do ponto de vista da informática, para encontrar as metáforas linguísticas não é preciso definir como candidatos palavras que venham a ser Veículos da metáfora ou Tópicos da metáfora, pois o programa de computador vai fazer apenas uma busca de caracteres no corpus, sendo assim, não importa para o programa qual a classificação da palavra de busca. Porém, para o pesquisador, talvez seja importante selecionar como candidatos apenas palavras que sejam classificadas como Tópicos ou Veículos ou alguma

<sup>2</sup> Esse termo não aparece comumente na pesquisa de metáfora com corpora. Todavia nos parece necessário usá-lo para explicitar a metodologia.

outra categoria. Isso fica a critério do pesquisador, mas em nada altera o procedimento de busca das metáforas no corpus.

O processo de seleção de candidatos depende do tipo de pesquisa, que podemos classificar em dois: seletiva ou abrangente. Se for seletiva, o pesquisador seleciona os candidatos que achar melhor, de acordo com os objetivos da sua pesquisa. Se for abrangente, o pesquisador tenta abarcar todas ou pelo menos a maior quantidade possível de candidatos, a fim de dar conta da maior diversidade possível de metáforas que compõem o corpus. Há exemplos de ambos os tipos na literatura. Deignan (DEIGNAN, 1999) é um exemplo de estudo seletivo, em que a autora selecionou algumas palavras como *price*, *heated* e *blow* para investigar em um corpus de inglês geral (Bank of English). A seleção dessas palavras foi feita a partir do interesse da pesquisadora, para atingir seu objetivo, que era o de mostrar que certas colocações indicam primordialmente o sentido metafórico dessas palavras e não o sentido literal. Ela não teve a intenção de descrever o corpus inteiro, nem de mostrar palavras que tinham maior propensão, entre todas as do inglês, de exibir o ponto em questão. A sua intenção era essencialmente a de ilustrar, tomando como exemplo algumas palavras, a influência da padronização na sinalização de metáforas em inglês, um comportamento que pode ser estendido às demais palavras do vocabulário inglês. Cameron (2003), por sua vez, é um exemplo de estudo abrangente, pois nele a autora quis identificar todas as metáforas de um dado corpus, composto de transcrições de aulas de uma escola primária inglesa. A prospecção das metáforas foi feita de modo manual, isto é, por meio da leitura de todo o corpus. Mais especificamente, a autora e sua equipe leram todas as transcrições, anotaram cada candidato, compararam as anotações de cada um, discutiram os casos divergentes e chegaram a um consenso do grupo. Normalmente, as pesquisas seletivas são realizadas em corpora de língua geral, enquanto as abrangentes são efetuadas em corpora especializados, isto é, restritos a um gênero, registro, tipo textual, autor ou evento, por exemplo.

No que se segue tratamos da pesquisa do tipo abrangente, que visa à identificação da maioria das metáforas do corpus. Porém, a parte que se segue referente à especificação das palavras de busca é relevante também para a pesquisa seletiva, visto que em ambos os casos é preciso especificar as palavras de busca a serem usadas no concordanciador.

### 2.1. *Corpus de estudo*

Empregaremos um corpus para ilustrar os pontos discutidos, formado por 32 depoimentos contendo narrativas pessoais da vida de cada depoente, colhidos pelo Museu da Pessoa ([www.museudapessoa.com.br](http://www.museudapessoa.com.br)). O total de palavras ('tokens') do corpus é de 432.032, com 18.712 vocábulos ('types'). Em cada depoimento, os depoentes (todos voluntários) são entrevistados

por funcionários do Museu, que lhes fazem perguntas sobre sua vida (para mais detalhes e exemplos dos depoimentos, vide Berber Sardinha (2007c)).

## 2.2. Metodologias de identificação de metáfora

Há essencialmente duas metodologias gerais de identificação de metáfora em corpora: baseada em corpus (*corpus-based*) e movida a corpus (*corpus-driven*). Apresentaremos cada uma delas mais a seguir.

Antes disso, é preciso esclarecer três aspectos. O primeiro é que as duas metodologias enfocadas não devem ser confundidas com metodologias ascendentes (*bottom-up*) e descendentes (*top-down*), que são termos empregados na literatura. Na literatura de metáfora, uma metodologia *bottom-up* é geralmente aquela em que o analista parte da metáfora linguística e não da metáfora conceptual (por exemplo, CAMERON, 2003); por outro lado, quando o analista parte da metáfora conceptual para encontrar as metáforas linguísticas correspondente, o estudo é tido como *top-down* (por exemplo KÖVECSSES, 2002). Na nossa concepção, tanto *corpus-based* quanto *corpus-driven* referem-se a metodologias que partem da metáfora linguística e não da conceptual, sendo ambos, portanto, *bottom-up*. O segundo aspecto é que tampouco deve-se tentar distinguir as duas metodologias em relação aos paradigmas quantitativo e qualitativo, pois tanto *corpus-based* quanto *corpus-driven* possuem elementos de ambos os paradigmas. O terceiro aspecto é que em qualquer uma das metodologias é preciso *julgar* as instâncias levantadas, ou seja, o pesquisador precisa decidir se um candidato à metáfora pode de fato ser considerado uma metáfora. Para o julgamento dos candidatos, empregaremos três procedimentos: o Metaphor Identification Procedure, ou MIP (PRAGGLEJAZ GROUP, 2007), o Metaphor Identification through Vehicle Terms, ou MIV (CREET, 2006), e o MC (de 'Metáfora Conceptual', título provisório), um procedimento em fase de desenvolvimento no âmbito do Grupo de Estudos de Linguística de Corpus (GELC), cadastrado no CNPq<sup>3</sup>. Tanto o MIP quanto o MIV são destinados ao julgamento de metáforas linguísticas, isto é, unidades de metáfora expressas na fala ou na escrita (p.ex. 'estamos desperdiçando tempo'). Já o MC destina-se à categorização de metáforas conceptuais, que são construtos mentais, e podem ser definidas como maneiras convencionais de 'conceitualizar um domínio de experiência em termos de outro, normalmente de modo inconsciente'<sup>4</sup> (LAKOFF, 2002:4), como TEMPO É DINHEIRO, que se manifesta na linguagem por meio da metáfora linguística 'estamos desperdiçando tempo', entre outras. Devido a restrições de espaço, não podemos detalhar os procedimentos neste trabalho; para mais detalhes, o leitor pode se referir a Berber Sardinha (2009b).

<sup>3</sup> <http://dgp.cnpq.br/buscaoperacional/detalhegrupo.jsp?grupo=071801AMV3COY>

<sup>4</sup> 'A conceptual metaphor is a conventional way of conceptualizing one domain of experience in terms of another, often unconsciously.'

### 2.2.1. Metodologia baseada em corpus (corpus-based)

A metodologia baseada em corpus tem como característica principal o fato de que o pesquisador delimita os candidatos de antemão. Essencialmente, o pesquisador pode delimitar dois tipos de candidatos: (1) Relacionados ao Tópicos: palavras referentes aos temas encontrados no corpus. (2) Relacionados aos Veículos: palavras usadas metaforicamente no corpus. Os candidatos podem ser delimitados de diversas maneiras, incluindo:

- leitura parcial do corpus;
- consulta à literatura prévia;
- intuição e/ou conhecimento prévio do pesquisador.

Em seguida, o pesquisador faz buscas no corpus para encontrar os candidatos que eleger. Essas buscas geralmente são feitas com um concordanciador.

Para ilustrar, vamos tomar o corpus de estudo e retomar cada um dos critérios acima. Primeiramente, pela leitura parcial do corpus, pude encontrar Veículos como ‘combativa’, ‘lutando’ e ‘baixava’. Em seguida, em consulta à literatura prévia, pude selecionar candidatos como ‘tive uma vida cheia’ (metáfora linguística ligada à metáfora conceptual VIDA É UM RECIPIENTE); ‘tirei o máximo da vida’ (VIDA É UM RECIPIENTE); ‘aproveitar a oportunidade’ (VIDA É UM JOGO DE AZAR); ‘estou me sentindo para cima’ (FELIZ É PARA CIMA); ‘ele está mesmo para baixo’ (TRISTE É PARA BAIXO); ‘ele caiu doente’ (DOENÇA E MORTE SÃO PARA BAIXO); ‘já havia passado muito tempo quando’, ‘o tempo para agir chegou’, ‘nas semanas seguintes’, ‘nas semanas precedentes’ (TEMPO É UM OBJETO EM MOVIMENTO); ‘avancarmos através dos anos’, ‘aproximando do fim do ano’ (TEMPO É UM OBJETO ESTACIONÁRIO), todas de Lakoff e Johnson (1980/2002:61, 62, 100, 101, 118). E ainda ‘sem direção na vida’, ‘onde eu quero estar na vida’, ‘estar numa encruzilhada na vida’, ‘deixar alguém atrapalhar seu caminho’, ‘passar por muita coisa na vida’ (A VIDA É UMA VIAGEM) (KÖVECSSES, 2002:3). Por fim, por meio da minha intuição, pensei em candidatos como ‘vida pesada’, ‘vida toda’, ‘vida de luta’ e ‘levar a vida’. Também formulei um grupo de palavras relacionadas, a meu ver, aos acontecimentos da vida, incluindo palavras como ‘vida, morte, melhorar, piorar, amigos, família, pai, mãe, irmão, irmã, tio, tia, passado, presente, futuro, cidade, bairro, rua’.

Tendo feito esse tipo de levantamento dos candidatos, o próximo passo consiste em fazer concordâncias. Para efetuar as buscas, é preciso selecionar e isolar as palavras que serão os termos de busca, de modo a otimizar os achados. Por exemplo, em relação ao primeiro candidato, que é ‘tive uma vida cheia’, o pesquisador pode optar por buscar as palavras ‘tive’, ‘vida’ e

‘cheia’, separadamente, a fim de obter o maior número possível de resultados da busca. Caso contrário, se optar por buscar exatamente a expressão completa, pode obter poucos ou nenhum resultados. Contudo, como não há regras a respeito, o pesquisador pode experimentar buscas com palavras isoladas ou com grupos de palavras e avaliar os resultados. Ainda em relação às concordâncias, é preciso buscar as formas flexionadas desses candidatos. Por exemplo, para o candidato ‘amigo’, as formas flexionadas seriam amigos, amigas, amigão, amigona, amiguinho, etc. O pesquisador pode especificar cada uma, ou usar um caracter coringa, que nos concordanciadores geralmente é o asterisco. Assim, para encontrar as formas flexionadas de ‘amigo’<sup>5</sup>, a palavra de busca seria ‘amig\*’, na qual o asterisco substitui as diversas letras que compõem a porção variável de cada forma flexionada. Mas é preciso lembrar que o caracter coringa pode encontrar palavras indesejadas; por exemplo, ‘agi\*’: a intenção com esse caracter coringa é capturar algumas formas conjugadas do verbo ‘agir’, como ‘agindo’, ‘agimos’, etc. Mas ela também pode encontrar, caso existam no corpus, palavras como ‘agitação’, ‘agilizar’ e ‘agitado’, que não são derivadas do verbo ‘agir’. Dessa forma, uma possível lista de candidatos resultante da aplicação desses princípios seria<sup>6</sup>:

agi*	combativa*	onde
amig*	direç*	oportunidade*
ano*	doente*	pai*
aproveit*	encruzilhada*	pass*
aproxim*	est*	pesada*
atrapalh*	família*	pior*
através	fim	precedente*
avanç*	futuro*	presente*
bairro*	irmã*	rua*
baix*	irmão*	seguinte*
baixo	lev*	semana*
cai*	lut*	sent*
caminh*	luta*	tempo*
cheg*	melhor*	tia*
cheia*	morte*	tio*
cidade*	máximo*	tir*
cidade*	mãe*	vida*
cima		

Tabela 1: Candidatos a Veículo

<sup>5</sup> As aspas não devem ser digitadas no concordanciador.

<sup>6</sup> Devido a limitações de espaço, não apresentamos a lista definitiva. É recomendável alterá-la substituindo alguns caracteres coringa pela formas flexionadas individuais, como as do verbo ‘agir’, que são ‘ajo’, ‘ages’, ‘age’ etc.

Em seguida, passamos ao concordanciador, onde digitamos cada um dos candidatos. Para ilustrar, mostramos abaixo uma parte da concordância de 'vida', feita com o WordSmith Tools Concord:

79 P1- A senhora tem uma história de vida bonita.  
 80 es viu? Que mensagem que eles tem que enfrentar a vida como eu enfrentei a minha. (risos)  
 81 ntei a minha. (risos) A gente tem que enfrentar a vida conforme ela vem. Então, não sei o  
 82 R- Não falei tudo já? A minha vida. (risos)  
 83 R- Então, eu acho que na vida, sabe do que eu senti muita falta?

**Figura 1:** Concordância de 'vida'

A largura e a ordenação da concordância podem ser alteradas pelo usuário<sup>7</sup>. O analista faria agora o julgamento de todas as linhas de concordância, que no corpus de estudo somaram 628. O número de ocorrências (ou 'citações') de uma palavra recomendada para identificar seus sentidos é mil (DEIGNAN, 2005). Esse número de corte pode ser aplicado quando o pesquisador tiver mais de mil ocorrências de uma dada palavra no seu corpus, limitando os casos a analisar a mil. Como aqui as 628 ocorrências correspondem a todas as ocorrências do corpus, esse número de corte não precisa ser aplicado. Para exemplificar, tomaremos o caso da linha 81, que contém a frase 'A gente tem que enfrentar a vida conforme ela vem.' Aplicando os procedimentos de análise MIV, MIP e MC, temos:

### MIV

Candidato a Veículo: enfrentar.

Tópico: necessidade de lidar com acontecimentos difíceis da vida de alguém.

Sentido básico: opor-se, encarar fisicamente.

Sentido contextual: lidar com os problemas.

Incongruência entre os sentidos: Sim.

Conexão ou transferência de sentido entre Tópico e Veículo? Sim.

Lidar com os problemas está ligado a opor-se a esses problemas fisicamente.

**Julgamento: uso metafórico.**

### MIP

Candidato a Veículo: enfrentar.

Sentido básico: opor-se, encarar fisicamente.

Sentido contextual: lidar com os problemas.

Contraste entre os sentidos: O sentido básico contrasta com o sentido contextual.

Há relação entre os sentidos? Sim. O sentido de lidar com os problemas está ligado a opor-se a esses problemas fisicamente, no contexto.

**Julgamento: uso metafórico.**

<sup>7</sup> Vide Berber Sardinha (1996b; 2009a) para maiores detalhes sobre como fazer isso no WordSmith Tools.

## MC

Tópico de superfície/Termos do Tópico: vida

Tópico conceptual/Domínio-alvo: vida

Candidato a Veículo de superfície: enfrentar

Veículo conceptual/Domínio-fonte: adversário

Proposta de Metáfora conceptual [Domínio-alvo é Domínio-fonte]: VIDA É UM ADVERSÁRIO

Testes da metáfora proposta:

Há mapeamentos possíveis? Sim.

Liga um domínio abstrato a um concreto? Sim.

**Julgamento: metáfora conceptual é plausível.**

### 2.2.2. Metodologia movida<sup>8</sup> a corpus (corpus-driven)

Na metodologia movida a corpus, o pesquisador começa por uma prospecção do corpus, sem delimitar de imediato quais palavras seriam candidatas. Esta metodologia envolve também a análise de concordâncias, como na metodologia baseada em corpus. A diferença está na seleção dos candidatos, que é feita com base em uma prospecção feita por programas de computador.

Essa prospecção pode ser feita de várias maneiras, usando ferramentas computacionais como o KeyWords, que faz parte da suíte WordSmith Tools (SCOTT, 1997), o Identificador de Metáforas do CEPRIL (BERBER SARDINHA, 2006), o WMatrix (RAYSON, 2002) e o Semantic Distance (PEDERSEN e PATWARDHAN, 2002). O KeyWords retira do corpus palavras com frequência marcada, isto é, que ocorrem relativamente mais vezes no corpus de estudo em comparação a um corpus de referência. O pesquisador pode olhar as palavras-chave encontradas pelo programa e selecionar aquelas que mais caracterizam seu corpus. Ele foi usado para encontrar candidatos a metáforas por Partington (2006) e por Berber Sardinha (2002b). O Identificador de Metáforas, por sua vez, é um programa destinado exclusivamente à localização de metáforas em corpora e mostra a probabilidade de cada palavra do corpus ser usada metaforicamente. O pesquisador pode ter uma visão instantânea de quais palavras foram possivelmente usadas metaforicamente no seu corpus. O programa já foi empregado em diversas pesquisas e continua em desenvolvimento. Por fim, o WMatrix faz, entre outras coisas, uma análise semântica do corpus, classificando cada palavra em uma categoria de conteúdo (p.ex. 'love' [amor] é classificado como pertencente à categoria de 'relationships' [relacionamentos]) (CULPEPPER, 2009). Com base nessa classificação, o pesquisador pode selecionar campos semânticos e as palavras associadas a esses campos semânticos para efetuar a análise. O WMatrix funciona apenas com dados em inglês. O

<sup>8</sup> O termo 'movido a corpus' aparece em Berber Sardinha (2002a), mas há outras alternativas na literatura.

Semantic Similarity (também conhecido por Semantic Distance) (PEDERSEN e PATWARDHAN, 2002) calcula a proximidade entre os sentidos de pares de palavras e exibe essa proximidade na forma de um escore (por exemplo, 'dignity' [dignidade] e 'respect' [respeito], que possuem um escore de similaridade de 8,8, de um máximo possível de 12,5). Sua aplicação na localização de metáfora ainda é experimental. Berber Sardinha (2007a) verificou o escore de similaridade de vários pares de palavras de metáforas linguísticas em um corpus e notou que muitas delas tinham um escore baixo, denotando distância semântica. Essa distância pode ser interpretada como uma manifestação da incongruência ou anomalia que é considerada um dos requisitos para a existência de metáfora (CAMERON, 2003; PRAGGLEJAZ GROUP, 2007; STEEN, 2007). O calculador de Semantic Similarity funciona apenas com dados do inglês.<sup>9</sup> Como os dados usados aqui são em português, nem o WMatrix nem o Semantic Similarity serão apresentados. Consequentemente, enfocaremos, a seguir, o KeyWords e o Identificador de Metáforas, ilustrando o funcionamento de cada um com análises de metáfora com base nos procedimentos MIP, MIV e MC discutidos acima.

#### 2.2.2.1. O programa KeyWords do WordSmith Tools

Palavras-chave são palavras que possuem uma frequência relativa estatisticamente diferente (isto é, maior ou menor) no corpus de estudo em comparação a um corpus de referência (SCOTT e TRIBBLE, 2006; BERBER SARDINHA, 2009a). Palavras-chave podem ser calculadas por meio de alguns programas de computador, como o WordSmith Tools (SCOTT, 1997) e o Reading Class Builder (MOREIRA FILHO, 2007). Trataremos aqui do WordSmith Tools, que é o programa que difundiu esse procedimento<sup>10</sup>. O componente do WordSmith Tools que calcula as palavras-chave é o KeyWords. Devido ao escopo deste trabalho, não é possível explicar o modo de utilização do programa KeyWords; para uma explicação detalhada, o leitor pode se referir à ajuda que acompanha o programa, que está em inglês, ou a Berber Sardinha (2009a), que mostra todos os comandos do programa WordSmith Tools, ou ainda a Berber Sardinha (1996a, 1999d, 2001), que traz um resumo dos comandos e da utilização do programa e a Berber Sardinha (2005) que discute critérios para seleção de palavras-chave.

O corpus de estudo é aquele em que se pretende encontrar as metáforas. Pode ser variado ou restrito a um gênero, registro, autor, etc. Já o corpus de referência é o termo de comparação, sendo geralmente maior e mais variado do que o corpus de estudo. Normalmente espera-se que o corpus de referência seja

<sup>9</sup> O programa está disponível no sítio do CEPRIIL (<http://www2.lael.pucsp.br/corpora>).

<sup>10</sup> O WordSmith Tools está na versão 5, mas empregamos aqui a versão 3, por preferência pessoal. Não há diferenças no cálculo das palavras-chave de uma versão para outra.

pelo menos cinco vezes maior do que o de estudo (BERBER SARDINHA, 1999a).

As palavras-chave são de dois tipos: positivas ou negativas. As positivas são aquelas cuja frequência no corpus de estudo é maior do que no corpus de referência, ao passo que as negativas são aquelas cuja frequência é menor no corpus de estudo. Para efeitos de seleção de candidatos à metáfora, parece-nos que as palavras-chave positivas sejam mais úteis do que as negativas, pois o foco da análise é o corpus de estudo e não de referência. Com as palavras-chave positivas, o pesquisador pode encontrar aquelas palavras que mais caracterizam seu corpus, ao passo que com as negativas ele encontraria aquelas que mais 'faltam' no seu corpus. Isso não significa que as negativas não sejam potenciais bons candidatos; alguns pesquisadores podem decidir incluí-las também. O cálculo do que chamamos de 'maior' e 'menor' é feito pelo programa por meio de testes estatísticos como o qui-quadrado e o log-likelihood, que comparam a frequência de cada palavra no corpus com sua frequência no corpus de referência (caso a palavra em questão não exista no corpus de referência sua frequência será zero).

O pesquisador pode alterar parâmetros para a análise de palavras-chave. Podem ser alterados elementos como a frequência mínima que a palavra deve ter para ser levada em conta (o default é 2), a quantidade máxima de palavras-chave a serem exibidas (o default é 500) e o nível de significância do teste estatístico (o default é 0,000001). Qualquer mudança desses valores pode alterar a composição do conjunto de palavras-chave resultante e, portanto, afetar quais palavras são consideradas como candidatas à metáfora. A composição do corpus de referência também influencia os tipos de palavra que podem se tornar chave. Um corpus de características semelhantes ao corpus de estudo tende a 'filtrar', ou seja, eliminar, os elementos semelhantes, porque muitas palavras que ocorreram no corpus de estudo também podem ter ocorrido no corpus de referência com a mesma intensidade, resultando em uma lista de palavras-chave que não inclui essas palavras. Como dissemos, extraímos as palavras-chave do corpus de estudo, usando como corpus de referência o Banco de Português versão 1, com 223 milhões de palavras, de vários gêneros e registros, orais e escritos. As primeiras 20 palavras-chave aparecem abaixo.

N	WORD	FREQ	VIDA.LST %	FREQ	TUDO.LST %	KEYNESS	P
1	EU	11,439	2.67	248,174	0.11	51,276.8	0.0000000
2	P	4,163	0.97	12,282		33,834.9	0.0000000
3	ERA	5,594	1.30	220,212	0.10	18,815.2	0.0000000
4	TINHA	3,772	0.88	93,487	0.04	15,905.5	0.0000000
5	PRA	2,859	0.67	48,410	0.02	14,093.8	0.0000000
6	AÍ	2,898	0.68	51,773	0.02	13,994.2	0.0000000
7	LÁ	2,944	0.69	70,894	0.03	12,567.4	0.0000000
8	SENHORA	1,632	0.38	12,644		10,429.6	0.0000000
9	GENTE	2,382	0.56	70,789	0.03	9,231.4	0.0000000
10	ENTÃO	2,545	0.59	89,223	0.04	9,085.9	0.0000000
11	MINHA	2,285	0.53	70,989	0.03	8,666.8	0.0000000
12	MEU	2,192	0.51	74,920	0.03	7,926.7	0.0000000
13	MUITO	3,404	0.79	244,815	0.11	7,779.1	0.0000000
14	PORQUE	2,873	0.67	172,549	0.07	7,467.7	0.0000000
16	R	4,453	1.04	467,176	0.20	7,382.8	0.0000000
16	ELA	2,943	0.69	188,953	0.06	7,302.4	0.0000000
17	ELE	4,570	1.07	524,670	0.23	6,924.3	0.0000000
18	VOCÊ	2,347	0.55	130,921	0.06	6,402.4	0.0000000
19	ME	2,151	0.50	131,587	0.06	5,461.3	0.0000000
20	IA	1,055	0.25	18,339		5,144.5	0.0000000

Figura 2: Parte superior da lista de palavras-chave do corpus de estudo

O *display* das palavras-chave contém as seguintes informações, na ordem da esquerda para a direita – Word: cada palavra-chave; Freq: frequência da palavra-chave no corpus de estudo; Vida.lst %: frequência da palavra-chave exibida no corpus de estudo, na forma de porcentagem, em relação ao total de palavras (*tokens*) do corpus de estudo; Freq: frequência da palavra-chave no corpus de referência; Udo.lst %: frequência da palavra-chave no corpus de referência, na forma de porcentagem, em relação ao total de palavras (*tokens*) do corpus de estudo; Keyness: resultado da prova estatística (log-likelihood); P: valor de p associado ao resultado da prova estatística (para explicações detalhadas sobre o cálculo desses números, vide Berber Sardinha (2009a)). A lista é ordenada por *default* por Keyness, em ordem decrescente. Desse modo, as palavras-chave mais distintivas aparecem no topo da lista. As mais distintivas são aquelas cujas frequências no corpus de estudo são mais díspares em relação ao corpus de referência. A palavra ‘mais chave’ do corpus é, portanto, ‘eu’, que corresponde a 2,67% das ocorrências no corpus de estudo, em comparação a apenas 0,11% no corpus de referência; a disparidade é grande, pois ‘eu’ é 24 vezes mais comum no corpus de estudo do que no de referência. De modo geral, essas principais palavras-chave referem-se à oralidade (pronomes pessoais, dêiticos) e ao gênero depoimento (verbos no passado, operadores argumentativos).

Em seguida, o pesquisador deve selecionar algumas palavras-chave no corpus para analisar. Não há critérios consensuais sobre como selecionar um número ideal de palavras-chave a serem buscadas, embora haja algumas propostas na literatura. Berber Sardinha (1999b, a, c, 2005) discute algumas maneiras de selecionar palavras-chave, a fim de que o pesquisador possa se concentrar nas mais relevantes do seu corpus. No caso da pes-

quisa em metáfora, a experiência nos mostra que a análise de algumas palavras-chave (até dez) pode revelar uma variedade de metáforas. Sendo assim, a princípio, mesmo selecionando poucas palavras podemos chegar a um conjunto expressivo de metáforas. Podemos começar, então, pela análise da palavra-chave principal do corpus, que é 'eu'. Como ela ocorreu mais de 11 mil vezes no corpus, é recomendável que seja feito um recorte de mil linhas para serem analisadas, conforme discutido acima.

A figura abaixo traz um trecho da concordância de 'eu'.

914 " Agora, a responsabilidade é da mãe."(risos). Aí eu fui morar com mamãe. Aí já comecei a trabalhar  
915 . Era na rua 15 de novembro. Bem pertinho. Porque eu trabalhava numa loja na esquina da rua Direita  
916 R - Aí, quando eu cheguei na adolescência...  
917 costura, essas coisas de doméstica. Aí depois que eu entrei pro comércio é que eu fui fazer o curso  
918 stica. Aí depois que eu entrei pro comércio é que eu fui fazer o curso de Comércio pra me especializ

**Figura 3:** Concordância de 'eu'

Examinamos a linha 916, que contém a frase 'Aí, quando eu cheguei na adolescência...', utilizando os procedimentos MIV, MIP e MC:

#### **MIV**

Candidato a Veículo: cheguei.

Tópico: adolescência.

Sentido básico: chegar fisicamente a algum lugar.

Sentido contextual: ter uma idade considerada como sendo parte da adolescência.

Incongruência entre os sentidos: Sim.

Conexão ou transferência de sentido entre Tópico e Veículo? Sim.

Ter uma idade determinada pode ser entendida como chegar a essa idade.

**Julgamento: uso metafórico.**

#### **MIP**

Candidato a Veículo: cheguei.

Sentido básico: chegar fisicamente a algum lugar.

Sentido contextual: ter uma idade considerada como sendo parte da adolescência.

Contraste entre os sentidos: o sentido básico contrasta com o sentido contextual.

Há relação entre os sentidos? Sim. Ter uma idade determinada pode ser entendida como chegar a essa idade.

**Julgamento: uso metafórico.**

#### **MC**

Tópico de superfície/Termos do Tópico: adolescência

Tópico conceptual/Domínio-alvo: vida

Candidato a Veículo de superfície: chegar

Veículo conceptual/Domínio-fonte: viagem  
 Proposta de Metáfora conceptual [Domínio-alvo é Domínio-fonte]: VIDA É UMA VIAGEM

Testes da metáfora proposta:

Há mapeamentos possíveis? Sim.

Liga um domínio abstrato a um concreto? Sim.

**Julgamento: metáfora conceptual é plausível.**

#### 2.2.2.2. O identificador de metáforas

Uma outra opção para prospecção de metáforas dentro da metodologia *corpus-driven* é o Identificador de Metáforas (também conhecido por Etiquetador de Metáforas ou Metaphor Candidate Identifier, MCI). Este é um programa online, disponível gratuitamente na rede (<http://www2.lael.pucsp.br> e <http://www.corpuslg.org/tools>). Já foi testado em diversas pesquisas de metáfora (p.ex. BERBER SARDINHA, 2007b; RODRIGUES, 2007; CANCIAN, 2008; MELLO, 2008; MESTRINER, no prelo). Devido ao escopo deste trabalho, não será possível oferecer um tutorial do programa; para isso o leitor pode se referir às referências citadas anteriormente. O Identificador de Metáforas funciona com duas línguas: português e inglês, cada uma correspondendo a uma versão diferente do programa. Cada versão foi criada com elementos diferentes, como descrito mais abaixo.

O Identificador de Metáforas tem o objetivo de examinar cada palavra de um corpus, aplicar os mesmos critérios de identificação a cada palavra e produzir uma lista de candidatos, no final. O programa não pressupõe que as metáforas sejam palavras e não quer circunscrever as metáforas a unidades ortográficas; apenas o processamento é feito por palavras, por razões computacionais. O pesquisador, em seguida, examina essa lista de candidatos e decide qual ou quais candidatos pretende estudar. Partindo dos candidatos, que são palavras, o pesquisador pode chegar a metáforas que são palavras, expressões, orações, frases, enfim, unidades praticamente de qualquer extensão. Normalmente, a quantidade de candidatos é muito grande para que se possa estudar todos eles. Dessa forma, algum tipo de seleção subsequente à análise do Identificador ainda é aplicada aos dados. Os candidatos são apresentados em uma lista ordenada pela probabilidade de uso metafórico de cada candidato, de tal modo que os candidatos com maior probabilidade aparecem no topo do lista e os demais em ordem decrescente de probabilidade. O pesquisador pode decidir investigar os candidatos com maior probabilidade de uso metafórico, que estão no topo da lista, os de média probabilidade, que estão no meio da lista, ou os de menor probabilidade, que estão no final da lista, por exemplo, ou ainda aplicar algum outro critério de escolha.

O Identificador de Metáforas tem como base as seguintes premissas: (1) visão probabilística da língua. Segundo essa visão, os sistemas linguísticos são postos em uso com maior ou menor probabilidade (HALLIDAY, 1993). Essa variação de probabilidade é o fruto ao mesmo tempo da variação individual de uso da língua (HOEY, 2005) e de aspectos sociais (HALLIDAY, 1993). Berber Sardinha (2008) sugere que a probabilidade de uso metafórico muda quando uma mesma palavra ou expressão é usada em um gênero específico e na língua em geral. Essa propriedade é chamada de *resetting* ou reajuste. (2) Padronização da linguagem. A linguagem é composta de padrões, que são formações compostas por palavras utilizadas tipicamente em conjunto (SINCLAIR, 1970/2004; SINCLAIR, 1991). Essas formações estão associadas a sentidos específicos, de tal modo que dois padrões diferentes normalmente exprimem dois sentidos diferentes. Na medida em que metáforas são recursos para exprimir sentidos, pode-se esperar que haja padrões associados a usos metafóricos e padrões associados a usos não metafóricos. Há estudos que dão respaldo a essa hipótese, como Deignan (2005, 2006), que mostrou a existência de padrões que estão fortemente associados a sentidos metafóricos.

O objetivo do programa não é o de substituir o pesquisador. O programa tem a função apenas de ser um coadjuvante na pesquisa, fazendo uma análise compreensiva e fidedigna do corpus. É compreensiva porque o programa examina cada uma das palavras do corpus. E é fidedigna porque os critérios de identificação são aplicados exatamente da mesma forma a cada uma das palavras.

O Identificador de Metáforas foi preparado a partir da análise manual de corpora. No caso da versão em português, foi analisado manualmente um corpus especializado composto por transcrições de teleconferências de bancos de investimento, o que foi complementado pela análise de um corpus geral do português (Banco de Português). Na versão em inglês, foi feita a análise das 500 palavras mais frequentes do inglês, conforme as frequências do British National Corpus. A qualidade do Identificador de Metáforas ainda não foi medida formalmente. Não sabemos ainda qual é a sua precisão na indicação de candidatos que sejam verdadeiramente parte de metáforas linguísticas. Tampouco sabemos sua capacidade de abrangência, isto é, quantas das metáforas de um corpus são sinalizadas pela lista de candidatos apresentada pelo programa. Contudo, de modo informal, já pudemos aferir que o Identificador de Metáforas funciona melhor com corpora de composição similar àqueles com que o programa foi treinado. No caso do português, com corpora de textos de investimentos ou economia; no caso do inglês, com corpora gerais. Para ilustrar, submetemos o corpus

de estudo ao identificador, que retornou uma listagem exibida parcialmente na figura abaixo.

#	Word	Tag(Prob)	Vehicle	Left Bndl	Right Bndl	Framework	WordClass
000001	ver	.7635	.3950	1.0000	1.0000	.7187	.7041
000002	pessoal	.7190	.0001	1.0000	1.0000	.8690	.7259
000003	brasil	.7115	.0001	1.0000	1.0000	.8690	.6887
000004	grupo	.7115	.0001	1.0000	1.0000	.8690	.6887
000005	nome	.7115	.0001	1.0000	1.0000	.8690	.6887
000006	ano	.7078	.0001	1.0000	1.0000	.8503	.6887
000007	livro	.7078	.0001	1.0000	1.0000	.8503	.6887
000008	era	.7075	.0001	1.0000	1.0000	.8333	.7041
000009	falar	.7075	.0001	1.0000	1.0000	.8333	.7041
000010	casa	.7061	.0001	1.0000	1.0000	.8421	.6887

**Figura 4:** Parte superior da tela de resultados do Identificador de Metáforas do CEPRIIL

A listagem é exibida em ordem decrescente pela coluna Tag(Prob), que indica a probabilidade de uso metafórico medida pelo identificador. A palavra com maior probabilidade de uso metafórico, segundo o programa, é 'ver'. O pesquisador pode selecionar quais palavras desejar, não sendo necessário seguir a ordem apresentada. No nosso caso, selecionamos a palavra 'ver' e fizemos uma concordância dela no WordSmith Tools. Foram encontradas 415 ocorrências, o que não exigiu a seleção das linhas para nos mantermos dentro das mil ocorrências. Parte da concordância é exibida na figura a seguir. O padrão (BERBER SARDINHA, 2004) exibido na concordância é 'ver se + lembrar', que está nas linhas 362 a 364 e 366.

```

361 um dia o zelador - porque havia fiscalização para ver se estava tudo fechado - disse assim: "olha es
362 utra. Se você perguntasse três, quatro , deixa eu ver se me lembro: a série das bonecas. Que meu fil
363 o aviso pelo correio a gente ficava torcendo para ver se o dinheiro estava lá. Eu e a minha mãe, peg
364 utra. Se você perguntasse três, quatro , deixa eu ver se me lembro: a série das bonecas. Que meu fil
365 R - Deixe eu ver se me lembro. Vendi muito, muito, muito. Sabe,
366 R - Era da... Deixa ver se eu me lembro.(pausa) Não me lembro. Aí depo

```

**Figura 5:** Concordância de 'ver'

Esse padrão pode ser analisado deste modo:

### MIV

Candidato a Veículo: ver.

Tópico: lembranças.

Sentido básico: usar os olhos para enxergar alguma coisa.

Sentido contextual: Acessar a memória para tentar recuperar uma lembrança.

Incongruência entre os sentidos: Sim.

Conexão ou transferência de sentido entre Tópico e Veículo?

Sim. Usar o sentido da visão para enxergar as lembranças pode ser entendido como acessar a memória para tentar recuperar as lembranças.

**Julgamento: uso metafórico.**

### MIP

Candidato a Veículo: ver.

Sentido básico: usar os olhos para enxergar alguma coisa.

Sentido contextual: Acessar a memória para tentar recuperar uma lembrança.

Contraste entre os sentidos: O sentido básico contrasta com o sentido contextual.

Há relação entre os sentidos? Sim. Usar o sentido da visão para enxergar as lembranças pode ser entendido como acessar a memória para tentar recuperar as lembranças.

**Julgamento: uso metafórico.**

**MC**

Tópico de superfície/Termos do Tópico: lembrar.

Tópico conceptual/Domínio-alvo: lembrar.

Candidato a Veículo de superfície: ver.

Veículo conceptual/Domínio-fonte: ver.

Proposta de Metáfora conceptual [Domínio-alvo é Domínio-fonte]: LEMBRAR É VER

Testes da metáfora proposta:

Há mapeamentos possíveis? Sim.

Liga um domínio abstrato a um concreto? Sim.

**Julgamento: metáfora conceptual é plausível.**

### **3. Comentários finais**

Neste artigo, foram apresentadas metodologias, procedimentos e programas de computador que podem ser usados na análise de metáforas em corpora. A intenção foi a de mostrar um leque de opções contendo as principais opções de procedimentos e ferramentas utilizadas nas pesquisas de metáfora com corpus, sem necessariamente advogar a primazia de nenhuma delas. Cada pesquisador deve procurar, dentro desse leque, a(s) mais relevante(s) tendo em vista seus objetivos de pesquisa. Esperamos com isso ter avançado um pouco no desafio de desenvolver metodologias mais explícitas e abrangentes de identificação de metáfora. Como a Linguística de Corpus é uma área que se desenvolve rapidamente, e o interesse dos linguistas de corpus pela área de metáfora vem aumentando, assim como o dos analistas de metáfora pela Linguística de Corpus, podemos esperar que esse leque de opções se amplie cada vez mais.

### Abstract

*In this paper, I look at ways in which we can apply Corpus Linguistics principles to research in metaphor, with a focus on methodological issues. Two methodologies are presented: corpus-based and corpus-driven. As far as corpus-based methods, I show a procedure for selecting metaphor candidates as well as how to use a concordancer to query a corpus. As far as corpus-driven methodologies, I present two pieces of software: KeyWords and the Metaphor Candidate Identifier.*

**Keywords:** *Corpus Linguistics. Metaphor. Methods.*

### Referências

BERBER SARDINHA, T. Applications of WordSmith KeyWords. *LWPAL (Liverpool Working Papers in Applied Linguistics)*, v. 2, n. 1, 1996a. p. 81-90.

\_\_\_\_\_. WordSmith tools. *Computers & Texts*, v. 12, 1996b. p. 19-21.

\_\_\_\_\_. A influência do tamanho do corpus de referência na obtenção de palavras-chave. *DIRECT Papers*, 38. LAEL, PUC/SP, São Paulo / AELSU, University of Liverpool, 1999a. Disponível em <<http://www2.lael.pucsp.br/direct>>.

\_\_\_\_\_. O banco de palavras-chave. *DIRECT Papers*, 39. LAEL, PUC/SP, São Paulo / AELSU, University of Liverpool, 1999b. Disponível em <<http://www2.lael.pucsp.br/direct>>.

\_\_\_\_\_. Um ponte de corte generalizado para listas de palavras-chave. *DIRECT Papers*, 41. LAEL, PUC/SP, São Paulo / AELSU, University of Liverpool, 1999c. Disponível em <<http://www2.lael.pucsp.br/direct>>.

\_\_\_\_\_. Using KeyWords in text analysis: Practical aspects. *DIRECT Papers*, 42. LAEL, PUC/SP, São Paulo / AELSU, University of Liverpool, 1999d. Disponível em <<http://www2.lael.pucsp.br/direct>>.

\_\_\_\_\_. Comparing corpora with WordSmith Keywords. *The ESPecialist*, v. 22, n. 1, 2001. p. 87-99.

\_\_\_\_\_. Corpora eletrônicos na pesquisa em tradução. *Cadernos de Tradução*, v. 9, n. 1, 2002a. p. 15-59.

\_\_\_\_\_. Metaphor in early applied linguistics writing: A corpus-based analysis of lexis in dissertations. In: I Conference on Metaphor in Language and Thought, 2002b, PUCSP.

\_\_\_\_\_. *Linguística de Corpus*. São Paulo: Manole, 2004.

- \_\_\_\_\_. Como encontrar as palavras-chave mais importantes de um corpus com WordSmith Tools. *Delta*, v. 21, n. 2, 2005. p. 237-250.
- \_\_\_\_\_. A tagger for metaphors. In: RaAM - Researching and Applying Metaphor 6, 2006, Leeds, UK.
- \_\_\_\_\_. Finding metaphors with the help of the computer. In: RaAM Workshop 'Issues in Researching Metaphor in Discourse', 2007a, Universidad de Castilla - La Mancha, Spain.
- \_\_\_\_\_. *Metáfora*. São Paulo: Parábola, 2007b.
- \_\_\_\_\_. Recontando a vida em narrativas pessoais: Um estudo de metáforas na perspectiva da Linguística de Corpus. *Organon*, v. 21, 2007c. p. 143-160.
- \_\_\_\_\_. Metaphor probabilities in corpora. In: ZANOTTO, M. S. et al (Orgs.). *Confronting Metaphor in Use: An Applied Linguistic Approach*. Amsterdam/Atlanta, GA: Benjamins, 2008. p. 127-148.
- \_\_\_\_\_. *Pesquisa em Linguística de Corpus com WordSmith Tools*. Campinas: Mercado de Letras, 2009a.
- \_\_\_\_\_. *Procedimentos de identificação de metáfora*. Trabalho inédito, 2009b. Disponível em: <<http://www.corpuslg.org/gelc/pub>>.
- BERBER SARDINHA, T.; BARCELLOS ALMEIDA, G. M. A Linguística de Corpus no Brasil. In: TAGNIN, S.; VALE, O. A. (Orgs.). *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, 2008. p. 17-40.
- CAMERON, L. *Metaphor in Educational Discourse*. London: Continuum, 2003.
- CANCIAN, E. P. F. *O discurso de uma revista especializada em educação: um olhar sobre a construção metafórica do professor*. (Dissertação de mestrado) - LAEL, PUCSP, São Paulo, 2008.
- CREET. *Metaphor Analysis Project*. Trabalho inédito, 2006. Disponível em: <<http://creet.open.ac.uk/projects/metaphor-analysis/index.cfm>>.
- CULPEPPER, J. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics*, v. 14, n. 1, 2009. p. 29-59.
- DEIGNAN, A. Linguistic metaphor and collocation in nonliterary corpus data. *Metaphor and Symbol*, v. 14, n. 1, 1999. p. 19-36.
- \_\_\_\_\_. *Metaphor and Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins, 2005.
- \_\_\_\_\_. A grammar of linguistic metaphors. In: STEFANOWITSCH, A.; GRIES, S. T. (Orgs.). *Corpus-based Approaches to Metaphor and Metonymy*. Berlin; New York: M. de Gruyter, 2006. p. 106-122.

- GIBBS, R. W. (Org.) *The Cambridge Handbook of Metaphor and Thought*. New York: Cambridge University Press, 2008.
- HALLIDAY, M. A. K. Quantitative studies and probabilities in grammar. In: HOEY, M. (Org.) *Data Description Discourse -- Papers on the English Language in Honour of John McH Sinclair on his Sixtieth Birthday*. London: HarperCollins, 1993. p. 1-25.
- HOEY, M. *Lexical Priming: A New Theory of Words and Language*. London, New York: Routledge, 2005.
- KÖVECSES, Z. *Metaphor - A Practical Introduction*. New York: Oxford University Press, 2002.
- LAKOFF, G. *Moral Politics: How Liberals and Conservatives Think* 2nd. ed. Chicago: University of Chicago Press, 2002.
- LAKOFF, G.; JOHNSON, M. L. *As Metáforas da Vida Cotidiana*. Campinas: Mercado de Letras / EDUC, 1980/2002.
- MCENERY, T.; WILSON, A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.
- MELLO, L. *Identificação e tradução de metáforas linguísticas e conceituais em abstracts da esfera acadêmica: Uma análise baseada em Linguística de Corpus*. (Dissertação de mestrado) - LAEL, PUCSP, São Paulo, 2008.
- MESTRINER, V. *Metáforas de Luís Inácio Lula da Silva e de George W. Bush numa perspectiva da Linguística de Corpus*. (Dissertação de mestrado) - LAEL, PUCSP, São Paulo, no prelo.
- MOREIRA FILHO, J. L. Reading Class Builder. Versão 1. PUCSP, 2007. Disponível em: <<http://www.corpuslg.org/software>>.
- PARTINGTON, A. Metaphors, motifs and similes across discourse types: Corpus-Assisted Discourse Studies (CADS) at work. In: STEFANOWITSCH, A.; GRIES, S. T. (Orgs.). *Corpus-based Approaches to Metaphor and Metonymy*. Berlin; New York: M. de Gruyter, 2006. p. 267-304.
- PEDERSEN, T.; PATWARDHAN, S. Distance Perl Package. Versão 0.1. University of Minnesota, Duluth, 2002. Disponível em: <[www.d.umn.edu/~tpederse/Code](http://www.d.umn.edu/~tpederse/Code)>.
- PRAGGLEJAZ GROUP. MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, v. 22, n. 1, 2007. p. 1-39.
- RAYSON, P. *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. (PhD Dissertation) - Computing Dept., University of Lancaster, Lancaster, 2002.
- RODRIGUES, A. D. S. S. *Metáforas do líder empresarial e histórico: uma abordagem baseada em corpus*. (Dissertação de mestrado) - LAEL, PUCSP, São Paulo, 2007.
- SCOTT, M. WordSmith Tools. Version 3. Versão. Oxford University Press, 1997. Disponível em: <[www.lexically.net/wordsmith](http://www.lexically.net/wordsmith)>

SCOTT, M.; TRIBBLE, C. *Textual Patterns - Key Words and Corpus Analysis in Language Education*. Amsterdam/Philadelphia: John Benjamins, 2006.

SINCLAIR, J. M. *English Collocation Studies: The OSTI Report*. London ; New York: Continuum, 1970/2004.

———. *Corpus, Concordance, Collocation*. Oxford, New York: Oxford University Press, 1991.

STEEN, G. *Finding Metaphor in Grammar and Usage*. Amsterdam / Philadelphia: John Benjamins, 2007.