

ANÁLISE DE METÁFORA EM CORPORA

Tony Berber Sardinha

Pontifícia Universidade Católica de São Paulo

Resumo

Neste trabalho, apresento maneiras de identificar metáforas em corpora. Corpora estão ficando cada vez mais fáceis de obter e propiciam uma riqueza de dados nunca vista. Porém, os corpora colocam novos desafios para o estudioso de metáfora, porque normalmente a quantidade de dados existente em um corpus supera a capacidade humana de análise de todas as ocorrências de possíveis metáforas. Proponho alguns procedimentos para análise de corpora com vistas à identificação de metáfora, focando especialmente no identificador de metáforas, um programa online dedicado exclusivamente à localização de possíveis metáforas em corpora. O uso do identificador é ilustrado por meio da análise de metáforas do presidente Lula, em um corpus contendo todos seus pronunciamentos oficiais desde sua posse, somando mais de 2 milhões de palavras.

Palavras-chave: metáfora; lingüística de corpus; identificação de metáfora.

Ilha do Desterro	Florianópolis	nº 52	p.167-199	jan./jun. 2007
------------------	---------------	-------	-----------	----------------

Abstract

In this paper, I show ways of identifying metaphors in corpora. Corpora are becoming easier to obtain, and they provide a wealth of detail that was never before seen. However, corpora also pose new challenges to metaphor researchers, since normally the amount of data in a corpus far exceeds the human capacity to pinpoint all the occurrences of possible metaphors. I suggest procedures for corpus analysis with a view to metaphor identification, focusing particularly on a metaphor identifier, an online tool especially dedicated to spotting possible metaphors in corpora. The use of this identifier is illustrated with an analysis of president Lula's metaphors, in a corpus of all his official speeches since his inauguration, totaling over 2 million words.

Keywords: metaphor; corpus linguistics; metaphor identification.

1. Introdução

Numa época em que o Presidente da República em exercício prima por suas metáforas (vestir a camisa, encher a barriga de pão e maionese numa churrascada, receber cartão vermelho, entre tantas outras), costuma-se ver e ouvir referência a metáfora constantemente na mídia. Chegaram até a dizer que o avião oficial da presidência (o 'Aero-Lula') tinha um compartimento especial para carregar as metáforas.

Uma metáfora, como o presidente Lula bem sabe, serve para dizer uma coisa por meio de outra. O que pouca gente se dá conta, entretanto, é que uma metáfora também nos faz *entender* e *conceituar* uma coisa pela outra. Quando dizemos que vamos vestir a camisa de um setor da sociedade, estamos conceituando e entendendo a sociedade como um esporte, as pessoas como jogadoras, outros grupos como adversários, a convivência como uma partida e a conduta desejada como as regras do jogo, observadas por um árbitro. Todas essas conexões, ou *mapeamentos*, não precisam ser ditas, pois estão subentendidas pelo campo conceitual de partida esportiva. Conhecemos o que é uma partida e as partes envolvidas nela, por isso podemos entender a expressão como uma metáfora (e não como uma mentira, pois afinal de contas não é preciso vestir um uniforme de um time para realizar a metáfora).

Fatos como um político de destaque ser um metaforicista ajudam a reacender o interesse num tema que vem apaixonando o ser humano por mais de um milênio, pelo menos.

Neste artigo, quero argumentar que mais do que o interesse popular circunstancial pela metáfora, há um fato novo que veio para mudar a maneira como enxergamos a metáfora, que passa bem mais despercebido: o uso de corpora eletrônicos e de ferramentas de informática para encontrá-las. Antes de entrar nesse assunto, porém, preciso fazer um prelúdio histórico-crítico.

O estudo da metáfora vem de muitos séculos; no Ocidente, data de pelo menos o século IV a.C, com Aristóteles. Podemos dizer que é tão antigo quanto a própria escrita. Por muito tempo, ele se manteve pouco alterado. Houve uma intensificação da fúria classificatória, entre os estudiosos do tema, resultando em extensas classificações de figuras de linguagem. Essa tradição atravessou a Idade Média, a Renascença, a Idade Moderna e pelo menos a metade do século XX. Na segunda metade do século XX, alguns filósofos e estudiosos voltaram-se novamente ao tema, renovando o interesse pelo assunto. Tivemos, então, as contribuições de I.A. Richards (1936), com a ênfase na expressão metafórica e na maneira de nomear os componentes da metáfora, a de Max Black (1962), com suas teorias da Substituição e da Interação, entre outros.

Porém dois acontecimentos mudaram para sempre os rumos dos estudos da metáfora. O primeiro foi a teoria cognitiva da metáfora, encapsulada na proposta de Lakoff e Johnson (1980; 1999), que discutiremos mais abaixo. Essa teoria veio mudar nossa visão de metáfora radicalmente. A metáfora deixou de ser uma figura de linguagem para ser um processo estruturador do pensamento; deixou de ser um aspecto da linguagem restrito a certos tipos de texto (ficação literária) ou de certos tipos de indivíduo (poeta) ou prática social (oratória); deixou de ser um ornamento original para ser principalmente um recurso convencional.

O outro acontecimento foi o aparecimento dos corpora eletrônicos, que vieram a mudar não somente o estudo da metáfora (Deignan) mas

também o da linguagem de modo mais geral (Biber *et al.*, 1998; Sinclair, 1991). O aparecimento de corpora não influenciou de imediato o estudo da metáfora, mas aos poucos tem alterado sobremaneira esse campo de estudo milenar. A influência da Lingüística de Corpus nos estudos da metáfora surgiu por meio da lexicografia, pois as equipes de criação de dicionário sempre tiveram de lidar com os sentidos figurados das palavras. Com a adoção de corpora eletrônicos como elemento-chave nos projetos de lexicografia (p. ex. Sinclair, 1987) as duas coisas se uniram: o dicionário precisava de análise de corpus para ser feito, e ao mesmo tempo a análise dos sentidos figurados passou a ser feita com base em evidência de corpora. Os lexicógrafos passaram a notar que os sentidos figurados, assim como os demais sentidos literais, se distinguiam um dos outros pela formação de padrões (Berber Sardinha, 2004), que são palavras que co-ocorrem habitualmente, de modo mais ou menos flexível. Daí não ser coincidência que Alice Deignan, a líder nos estudos de metáfora com corpora, mostre uma forte influência dos estudos lexicográficos na sua pesquisa. Ela privilegia uma metodologia bastante comum na lexicografia, qual seja, o uso de concordanciador, que também veremos mais a seguir.

2. Metáfora conceptual

A teoria da metáfora conceptual é uma proposta de base cognitiva formulada por George Lakoff e Mark L. Johnson, um lingüista e um filósofo americanos, respectivamente, no final da década de 1970 e divulgada, com grande sucesso, em seu livro ‘Metaphors We Live By’, de 1980 (George Lakoff & Johnson, 1980). Esse título em inglês significa algo como ‘as metáforas que nos guiam’¹. O título da obra já deixa claro o ponto principal da teoria: vivemos de acordo com as metáforas que existem na nossa cultura; praticamente não temos escolha: se quisermos fazer parte da sociedade, interagir, ser entendidos, entender o mundo, etc., precisamos seguir as metáforas que a nossa cultura nos coloca à disposição.

Os conceitos principais dessa teoria são:

- o Metáfora conceptual² (*conceptual metaphor*): ‘Uma metáfora conceptual é uma maneira convencional de conceitualizar um domínio de experiência em termos de outro, normalmente de modo inconsciente.’³ (Lakoff, 2002, p. 4). Por exemplo, AMOR É UMA VIAGEM. A convenção geral é grafar as metáforas conceptuais em caixa alta. Uma metáfora conceptual é assim chamada porque ela conceitualiza alguma coisa. No caso acima, a metáfora fornece um conceito de amor. Segundo esse conceito, amor seria uma viagem. Esse é o conceito metafórico.
- o Expressão metafórica (*metaphorical expression*): Expressão lingüística que é uma manifestação de uma metáfora conceptual. Por exemplo, ‘nossa casamento está indo muito bem’ é uma expressão que advém da metáfora conceptual AMOR É UMA VIAGEM.
- o Domínio (*domain*): Uma área do conhecimento ou experiência humana. No exemplo acima, os dois domínios são AMOR e VIAGEM. Há dois tipos de domínio: fonte e alvo. O domínio fonte é aquele a partir do qual conceitualizamos alguma coisa metafóricamente; no exemplo, viagem; geralmente é algo concreto, advindo da experiência. O domínio alvo é aquele que desejamos conceitualizar; esse é o domínio abstrato; no exemplo, amor. Os domínios podem ser mais amplos (RELAÇÕES AMOROSAS em vez de AMOR, ou DESLOCAMENTOS em vez de VIAGEM) ou mais específicos (NAMORO ADOLESCENTE em vez de AMOR ou DIRIGIR UM CARRO NUMA ESTRADA PERIGOSA em vez de VIAGEM), dependendo da situação. Um mesmo domínio fonte por servir a vários domínios alvo; por exemplo, VIAGEM é fonte do conceito de AMOR mas também de VIDA.

- o Mapeamentos (*mappings*): As relações feitas entre os domínios. No exemplo anterior, poderíamos ter os seguintes mapeamentos, entre outros:
 - Viajantes: amantes ou marido e mulher;
 - Mapa da viagem: planos futuros da vida a dois.
 - Destino da viagem: relação feliz a dois.
 - Deslocamento tranquilo na viagem: relação sem problemas;
 - Deslocamento contínuo e previsível na viagem: relação com problemas, devido à monotonia;
 - Pegar carona na viagem: ter um caso fora do relacionamento.
- o Desdobramentos (*entailments*): as inferências que podemos fazer a partir de uma metáfora conceptual. Por exemplo, dado que AMOR É UMA VIAGEM, então:
 - Se uma viagem longa e repetitiva é monótona, então um casal que vive há muito tempo juntos pode achar o relacionamento monótono;
 - Se uma viagem longa é cansativa, então um casal que vive juntos há muitos anos pode ficar cansado do relacionamento.

Essa é uma teoria que se contrapõe à visão lógico-positivista do mundo. A Teoria da Metáfora Conceptual propõe que não há verdades absolutas, pois as metáforas são culturais e, por definição, relativas a uma dada cultura, resultantes de mapeamentos relevantes para certas civilizações ou ideologias. Por exemplo, a civilização ocidental entende

que ‘tempo é dinheiro’. Essa é uma metáfora conceptual que estrutura o pensamento daqueles que vivem em culturas ocidentais, capitalistas, mesmo que não concordem com ela; mesmo assim, ela não rege a vida de seres humanos de outras culturas (como a de povos aborígenes).

3. Desafio para a Lingüística de Corpus

A teoria da metáfora conceptual coloca desafios para a Lingüística de Corpus, principalmente porque nessa visão, metáfora é uma representação mental. Ela é cognitiva (existe na mente e atua no pensamento). Por exemplo, AMOR É UMA VIAGEM é uma representação abstrata que (a teoria prediz) reside na mente. As pessoas não dizem ‘amor é uma viagem’, com exceção dos estudiosos da metáfora! Como a Lingüística de Corpus se ocupa de dados realizados, de produção, como pode ela dar conta de encontrar as metáforas conceptuais, que residem na mente?

A própria teoria da metáfora conceptual dá as pistas para a resposta. Primeiramente, embora abstratas e mentais, se concretizam no *uso lingüístico*, pois tomam forma na fala e na escrita por meio das expressões metafóricas. Ou seja, podemos chegar a AMOR É UMA VIAGEM encontrando expressões metafóricas licenciadas por ela, como ‘estamos num beco sem saída’, ‘estamos indo muito bem’, ‘estamos sem rumo’, etc., quando usadas no contexto da vida amorosa.

A segunda pista é que as metáforas conceptuais são convencionais, quer dizer, tomam a forma de expressões recorrentes na língua. Por esse ângulo, encontrar as metáforas conceptuais envolveria encontrar padrões habituais de uso da língua. Isso a Lingüística de Corpus sabe fazer, naturalmente, pois seu maior interesse é exatamente o uso típico e habitual das formas lingüísticas. Traduzindo para o estudo da metáfora, isso significaria que encontrar formas típicas de uso pode ser uma forma de nos conduzir às metáforas conceptuais. Por exemplo, formas recorrentes como ‘desperdiçar tempo’, ‘ganhar tempo’ e ‘economizar tempo’ podem nos levar à metáfora conceptual TEMPO É DINHEIRO.

Contudo, resta ainda o problema de como sabermos se uma expressão recorrente qualquer é uma expressão metafórica. A única solução nesse caso é a interpretação do sentido da expressão, levando em conta o seu contexto. Por exemplo, ‘não estamos indo a lugar nenhum’ pode ser literal se empregada para significar deslocamento físico, ou metafórica, se significar a vida amorosa conjunta. Somente um ser humano pode dizer qual o sentido da expressão em um determinado uso. Entretanto, se precisarmos da interpretação de cada expressão para termos uma idéia de seu potencial metafórico, a pesquisa com corpora eletrônicos se inviabiliza, pois ela pressupõe a análise de muitos dados, em quantidades que em geral estão além da capacidade humana de processamento, daí o uso de computadores.

Temos aí um impasse: precisamos da interpretação para julgar o potencial metafórico, mas não podemos interpretar todas as instâncias que possuem potencial metafórico, num corpus qualquer.

Discutiremos mais adiante uma proposta para sair desse impasse, baseada na noção de probabilidade e linguagem enquanto sistema probabilístico (Berber Sardinha, 2004).

Antes disso, mostraremos algumas maneiras de identificar metáforas em corpora eletrônicos.

4. Encontrando metáforas em corpora

Para ilustrarmos os procedimentos de identificação de metáfora, utilizaremos um corpus de estudo, mais especificamente coleção de todos os pronunciamentos oficiais do presidente Luís Inácio Lula da Silva, desde sua posse até o dia 22/2/2006. O corpus contém 870 textos, somando 2.074.590 ocorrências (*tokens*) e 32.833 itens (*types*). Esses mais de dois milhões de palavras são equivalentes a 5927 páginas impressas (de 350 palavras cada), ou ainda 29 livros encadernados (de 200 páginas cada). Os arquivos foram baixados da Internet, do site <http://www.info.planalto.gov.br/download/discursos>, e o corpus foi processado pelo listador e concordanciador online do LAEL, CEPRIL⁴.

Vejamos quais seriam as opções de análise desse corpus.

4.1. Opção 1: Ler o corpus

A primeira opção (por ser a mais tradicional) seria ler todo o corpus e anotar as metáforas. Isso tem uma desvantagem séria que é a exigência de muita paciência (para ler uma biblioteca de 29 livros!) e muita disciplina para não deixar escapar ocorrências de metáfora e manter consistência na análise. Muitas vezes, o analista deixa passar uma ocorrência de uso de metáfora para somente mais tarde, quando encontra outra ocorrência, lembrar que já havia visto tal uso antes. Nesse momento, ele deve voltar, procurar as ocorrências anteriores e marcá-las. Ou, pelo contrário, muitas vezes anotamos uma metáfora mas depois, mais à frente, mudamos de idéia. Nesse caso, é também preciso voltar atrás e desmarcar as ocorrências prévias. Por tudo isso, é imprescindível que se leia o corpus pelo menos duas vezes, para garantir fidedignidade no processo, o que significaria, no caso do corpus de pronunciamentos do presidente, ler o equivalente a 58 livros! Se o analista tiver um bom ritmo e conseguir dar conta de 100 páginas por dia (um ritmo alucinante, sem descanso de final de semana!), levará 59 dias para fazer a primeira leitura e mais 59 para a segunda, totalizando praticamente quatro meses de trabalho árduo!

Afora todos esses problemas, quando se trata de um corpus grande como esse, é preciso utilizar algum tipo de marcação (*markup*), de preferência numa versão eletrônica do corpus (por exemplo, num processador de texto) para permitir localizar as ocorrências de metáfora no corpus. Uma marcação feita em papel traz muitas dificuldades para contabilizar as ocorrências quando a fase de identificação estiver terminada.

4.2. Opção 2: Fazer buscas a partir da intuição e conhecimento prévio

A segunda alternativa seria partir de um ‘palpite’, seja motivado pela intuição ou por conhecimento das metáforas que o presidente já

utilizou. Por exemplo, poderíamos buscar num concordanciador palavras como ‘churrascaria’, ‘gol contra⁵’ e ‘time⁶’, que fizeram parte de metáforas já empregadas pelo presidente Lula e que foram veiculadas na imprensa. No caso de ‘churrascaria’, ele disse em agosto de 2003 que:

‘Quando vocês vão numa churrascaria comer um rodízio, tem aquele apressado que tudo que é maionese que vem ele come, que tudo que é lingüiça que vem ele come. Aí, quando chega a hora da coisa nobre, ele já está com o bucho cheio, não quer mais a picanha, a costela, a chamada carne nobre. Pagou por ela e não comeu porque foi apressado.’⁷

Ele respondia às críticas de que seu governo não estava produzindo resultados, dizendo que ainda era muito cedo para cobrar. Assim, os que cobravam eram ‘apressados’ como aqueles que não têm paciência para esperar o prato principal numa churrascaria e se saciam comendo o *couvert*. Uma metáfora conceptual que subjaz a esse uso é GOVERNAR É COMER EM CHURRASCARIA.

Algumas palavras importantes que constam nessa metáfora são (juntamente com suas freqüências no corpus):

Tabela 1: Freqüência de palavras usadas na metáfora da churrascaria

Palavra	Freq.	%
buchó	2	0,00010%
churrascaria	1	0,00005%
costela	7	0,00034%
costelas	1	0,00005%
maionese	1	0,00005%
picanha	6	0,00029%
picanhas	1	0,00005%
rodízio	2	0,00010%

Fazendo busca dessas palavras no corpus, obtemos a concordância a seguir.

1 o chega a hora da coisa nobre, ele já está com o "bucho cheio", ele não quer mais a picanha, a coste
2 na mesa e não valorizamos aquela que enchemos o "bucho". No mais eu quero agradecer a vocês, porque
3 sei se vocês já perceberam, quando vocês vão numa churrascaria comer um rodízio, tem aquele apressad
4 eu já vi mulher com uma estaca de pau grudada na costela, que perfurou o seu útero e matou o seu fi
5 m o "bucho cheio", ele não quer mais a picanha, a costela, a chamada carne nobre. Ele já não consegui
6 jeito que nós sabemos assar ou nunca comeram uma costela bem feita. Nós temos que fazer. O Presiden
7 tão podemos tratá-los com uma boa picanha, uma boa costela, quem sabe uma buchada, isso não pode ofer
8 em maio. Vou chegar com uma caixa de picanha e de costela. O embaixador brasileiro sabe que tem que
9 s para comerem um belo churrasco; vamos fazer uma costela, não queimada como vocês faziam para mim q
10 cês faziam para mim quando eu vinha aqui, mas uma costela de qualidade, uma picanha de qualidade, pa
11 vamo levar na nossa bagagem umas picanhas e umas costelas. E ao chegar vamos chamar o primeiro-minis
12 r um rodízio, tem aquele apressado que tudo que é maiôneze que vem ele come, tudo que é linguiça que
13 já está com o "bucho cheio", ele não quer mais a picanha, a costela, a chamada carne nobre. Ele já
14 possível, porque nunca comeram uma picanha do jeito que nós sabemos assar ou nunca co
15 aís. E nós, que não podemos tratá-los com uma boa picanha, uma boa costela, quem sabe uma buchada, i
16 ara o Japão, em maio. Vou chegar com uma caixa de picanha e de costela. O embaixador brasileiro sabe
17 ver na Embaixada. Vamos convidá-lo para comer uma picanha e vamos ver se depois ele vai ter problema
18 eu vinha aqui, mas uma costela de qualidade, uma picanha de qualidade, para ver se, depois disso, o
19 sta nada fazer, vamos levar na nossa bagagem umas picanhas e umas costelas. E ao chegar vamos chamar
20 eram, quando vocês vão numa churrascaria comer um rodízio, tem aquele apressado que tudo que é maiô
21 te talhões, para explorar a madeira em sistema de rodízio, permitindo assim a regeneração da mata. O

Figura 1: Concordância de palavras envolvidas com a metáfora da churrascaria

Analisando a concordância (e algumas vezes os textos completos, para ver um contexto maior e poder decidir com mais segurança), verificamos sete usos metafóricos, nas linhas: 1, 2, 3, 5, 12, 13 e 20. Desses sete, seis fazem parte da mesma passagem citada acima, associando-se assim à metáfora conceptual GOVERNAR É COMER EM CHURRASCARIA. Apenas a linha de número 2 traz uma citação de um outro contexto. Nele, o presidente usa 'bucho' da seguinte maneira:

"Então, eu penso que isso tem que ser discutido com os companheiros, porque senão nós ficamos discutindo sempre a comida que falta na mesa e não valorizamos aquela que [sic] enchemos o 'bucho'" (29/11/2005). Aqui, o presidente Lula quer dizer que 'em política, por mais que você faça, sempre estará faltando alguma coisa, e por mais que as pessoas conquistem sempre vai precisar mais um

pouco' (29/11/2005). Ou seja, ele reclama que os políticos não valorizam a situação atual do país ('a comida que enche o bucho') e desejam sempre mais ('a comida que falta na mesa'). Parece subjazer aí metáforas conceptuais como:

- A SITUAÇÃO DO PAÍS / GOVERNO É UMA REFEIÇÃO À MESA.
- DAR VALOR À SITUAÇÃO DO PAÍS / GOVERNO É APRECIAR UMA PRATO QUE SUSTENTA.

Poderíamos fazer o mesmo com outros 'palpites' ou lembranças de usos passados de metáforas do presidente para formular buscas no corpus. Mas esse exemplo da busca por palavras associadas à famosa metáfora da churrascaria já ilustra aspectos importantes desse procedimento:

- É preciso recorrer à memória diretamente ou pelo menos como meio de encontrar textos que tenham metáforas já usadas, para daí retirar termos para efetuar a busca no corpus.
- As buscas podem retornar poucas ocorrências, relativas ao tamanho do corpus. No exemplo mostrado acima, foram um total de 21 ocorrências, o que representa apenas 0,001% do corpus.
- Pode haver poucas instâncias de uso metafórico dentre os resultados da busca. No exemplo, foram sete usos metafóricos de um total de 21 ocorrências, o que soma apenas 33% de palavras empregadas com sentido metafórico (uma probabilidade de 1 em 3).
- As palavras selecionadas podem não ser as que têm mais potencial metafórico. Podemos ver isso na variação da probabilidade de uso metafórico entre as palavras buscadas.

Como se vê abaixo, a probabilidade de uso metafórico entre elas variou de 12,5% a 100%. Ou seja, o sentido metafórico de algumas delas não é o sentido mais comumente empregado pelo presidente:

- o Bucho: 100% (2 usos metafóricos em 2 usos no total)
- o Churrascaria: 100% (1 em 1)
- o Costela/Costelas: 12,5% (1 em 8)
- o Maionese: 100% (1 em 1)
- o Picanha: 14,3% (1 em 7)
- o Rodízio: 50% (1 em 2)

4.3. Opção 3: Fazer buscas a partir da lista de palavras

A terceira alternativa é inspecionar a lista de palavras e retirar de lá termos de busca. A inspeção em si é trabalhosa, pois são mais de 30 mil palavras. Se decidirmos olhar as palavras mais freqüentes, veríamos a parte superior da lista de freqüência, que aparece abaixo.

1 que (77141)	51 anos (4000)
2 de (67386)	52 temos (3974)
3 a (55749)	53 à (3934)
4 e (52865)	54 vamos (3905)
5 o (50298)	55 são (3900)
6 do (32634)	56 estamos (3773)
7 para (27687)	57 mundo (3601)
8 da (26246)	58 foi (3572)
9 não (23472)	59 todos (3550)
10 é (21773)	60 quero (3470)

- | | |
|----------------------|---------------------------|
| 11 um (19764) | 61 ele (3343) |
| 12 eu (18397) | 62 países (3275) |
| 13 com (16991) | 63 pessoas (3232) |
| 14 em (16704) | 64 companheiro (3217) |
| 15 nós (16252) | 65 nossa (3119) |
| 16 uma (15682) | 66 dizer (3085) |
| 17 no (13648) | 67 hoje (3061) |
| 18 os (12916) | 68 nos (3042) |
| 19 mais (12126) | 69 pode (3027) |
| 20 se (11788) | 70 coisa (2938) |
| 21 na (11402) | 71 também (2848) |
| 22 brasil (10273) | 72 essa (2835) |
| 23 as (9753) | 73 estado (2834) |
| 24 porque (9108) | 74 ano (2826) |
| 25 tem (7839) | 75 política (2745) |
| 26 gente (7734) | 76 você (2651) |
| 27 país (7471) | 77 este (2592) |
| 28 por (7222) | 78 estão (2585) |
| 29 fazer (6976) | 79 sua (2570) |
| 30 vocês (6799) | 80 ministro (2555) |
| 31 presidente (6701) | 81 coisas (2527) |
| 32 mas (6414) | 82 era (2482) |
| 33 como (6309) | 83 apenas (2460) |
| 34 muito (6170) | 84 agora (2449) |
| 35 ao (6064) | 85 cada (2419) |
| 36 aqui (6020) | 86 meus (2412) |
| 37 isso (5909) | 87 esse (2408) |
| 38 dos (5866) | 88 vezes (2373) |
| 39 ou (5122) | 89 quem (2367) |
| 40 ser (4879) | 90 pelo (2367) |
| 41 meu (4814) | 91 seja (2365) |
| 42 vai (4754) | 92 desenvolvimento (2364) |
| 43 governo (4565) | 93 me (2324) |

44 das (4541)	94 até (2321)
45 ter (4405)	95 povo (2312)
46 está (4402)	96 dia (2279)
47 nosso (4343)	97 entre (2213)
48 já (4324)	98 grande (2203)
49 república (4189)	99 tempo (2189)
50 quando (4145)	100 importante (2186)

Quadro 1: Lista de freqüência do corpus de pronunciamentos do presidente Lula; os números entre parênteses indicam a freqüência da palavra no corpus.

Dentre essas palavras, poderíamos talvez escolher ‘vai’ (posição 42; poderia indicar uma metáfora de viagem), ‘temos’ (52, podendo revelar metáfora ontológica), ‘vamos’ (54, talvez indicativa de metáfora de viagem) e ‘tempo’ (99, possivelmente envolvendo metáfora do TEMPO É DINHEIRO). O próximo passo seria efetuar buscas dessas palavras com o concordanciador.

Tal exercício sugere que essa estratégia de análise:

- Exige intuição e experiência por parte do analista para selecionar palavras que tenham potencial metafórico.
- Para que a análise não seja parcial, seria preciso verificar desse modo todas as mais de 31 mil palavras (*types*) constantes na lista.
- A proporção de formas (*types*) com algum uso metafórico gira em torno de 3% (Cameron, 2003) a 8% (Berber Sardinha, 2006a) do total de formas de um corpus. Se extrapolarmos para o corpus de pronunciamentos do presidente, teríamos 2626 palavras com potencial de uso metafórico. O problema para quem tenta achar tais palavras numa lista de palavras é que elas podem estar em qualquer posição da lista. Uma pesquisa sugere que, quando as

pessoas se defrontam com essa tarefa, tendem a escolher uma proporção grande de candidatos (13 a 17% da lista), mas acertam pouco (14%) (Berber Sardinha, 2006b).

4.4. Opção 4: Empregar um programa identificador de metáforas

Nesta seção apresentamos uma implementação de um programa voltado especificamente para a identificação de metáforas. Esse identificador é o único disponível publicamente no momento e funciona com corpora de língua portuguesa. Trata-se do identificador de metáforas online do LAEL, CEPRIL, que é acessado pelo endereço http://www2.lael.pucsp.br/corpora/metaphor_tagger.

O identificador de metáforas é um *software* específico para identificar metáforas em arquivos de computador. Programas como esse são raros. Há menção na literatura de apenas um outro extrator de metáforas (Z. Mason, 2004), que é de acesso restrito; por isso trataremos do identificador do LAEL, CEPRIL apenas.

A seção se divide em duas partes. A primeira mostra o funcionamento e o desenvolvimento do programa. A segunda ilustra o uso do identificador com uma análise do corpus de pronunciamentos do presidente Lula.

4.4.1. O identificador de metáforas

O programa se chama ‘identificador’ porque identifica candidatos a metáfora. Ele faz uma triagem do corpus, mostrando uma seleção prévia das palavras com maior probabilidade de terem sido empregadas metafóricamente. Ele identifica os candidatos a metáfora colocando ‘etiquetas’ em cada palavra do corpus. Em Lingüística de Corpus, uma etiqueta (*tag*) é um código que carrega alguma informação sobre uma palavra. No caso desse identificador, a informação trazida pela etiqueta é a probabilidade de uso metafórico. Essa probabilidade varia de 0,0001 a 1, sendo que 0,0001 (ou 0,01%) significa ‘praticamente nenhuma probabilidade’ de uso metafórico, e 1 (100%) indica ‘certeza de uso metafórico’.

O programa coloca em prática a noção de probabilidade e, de modo mais amplo, a de linguagem enquanto sistema probabilístico. Essa noções querem dizer que a linguagem é usada seguindo certas probabilidades: dependendo do contexto, algumas formas de expressão (*wordings*) são mais prováveis do que outros e, do mesmo modo, alguns sentidos são mais prováveis do que outros. Essa noção faz parte do senso comum de qualquer falante nativo, mas é muito rara na teoria lingüística. Poucos são os autores que incluem a probabilidade de uso como um aspecto teórico da competência lingüística (Biber, 1988; Bod *et al.*, 2003; Halliday, 1994).

Essa noção de probabilidade de uso está muito presente na Lingüística de Corpus. Como seus praticantes têm mostrado (Hunston, 2002), o uso lingüístico vem associado a certas probabilidades, bem como o sentido das palavras e expressões. Nos estudos da metáfora, Deignan (2005) mostra que o sentido metafórico das palavras varia de acordo com a maneira pela qual essas palavras são usadas.

Podemos partir desse conceito de probabilidade para o problema da localização das metáforas: se pudermos saber qual a probabilidade de uma palavra ou expressão ser metafórica, teremos condições de instruir o computador para localizar as palavras que possuam certa probabilidade de metaforicidade em um corpus. De posse dessas palavras, podemos então analisá-las mais detalhadamente, interpretando-as para saber se foram ou não de fato usadas metafóricamente e qual a metáfora conceptual subjacente a ela. Ou seja, falando metafóricamente, o computador passa a funcionar como um filtro: ele ‘filtra’ um corpus, retirando as ‘impurezas’ e mostrando somente as ocorrências mais desejáveis.

Em termos mais técnicos, podemos pensar nessa ferramenta como um identificador. Ele etiquetaria (isto é, colocaria códigos) em cada palavra de um corpus dizendo qual sua probabilidade de ter sido usada metafóricamente. O analista partiria então dos dados processados pelo identificador, sabendo de antemão quais probabilidades de metaforicidade. Isso possivelmente tornaria seu trabalho mais eficiente e abrangente, pois não precisaria decidir de antemão quais palavras

enfocar, a partir das muitas que seu corpus contém. Com base nas palavras etiquetadas, o analista pode identificar as expressões metafóricas; por exemplo, se o identificador indicasse que ‘economiza’ possui alta probabilidade de metaforicidade (digamos 90%), o analista iria ao corpus e veria como essa palavra foi usada. Digamos que tenha sido numa expressão como ‘ele economiza muito tempo comendo fora’. Nesse caso, ele julgaria o uso de ‘economiza’ como metafórico e decidiria, baseando-se em seus critérios, como delimitar a expressão metafórica que engloba a palavra metaforizada ‘economiza’ (podendo ser toda a frase ou apenas ‘ele economiza muito tempo’).

Já que estamos tratando da metáfora em uso, é útil distinguiirmos, neste ponto, as partes de uma expressão metafórica:

- Tópico: É a porção *não* metafórica de uma expressão metafórica. Em ‘ele economiza muito tempo’, o Tópico compreende ‘ele’ e ‘tempo’. Não há consenso sobre quais partes de um enunciado formam o Tópico de uma expressão. Ele pode ser apenas um termo (‘tempo’), mais de um (‘ele’ e ‘tempo’), ou mesmo nenhum. Por exemplo, um provérbio como ‘mais vale um pássaro na mão do que dois voando’, quando usada no seu sentido primeiro de dizer que alguém é precavido, não possui Tópico. Toda a expressão é metafórica.
- Veículo: É a porção *metafórica* de uma expressão metafórica. Em ‘ele economiza muito tempo comendo fora’, o Veículo é ‘economiza’. Em ‘mais vale um pássaro na mão do que dois voando’, todas as palavras fazem parte do Veículo.

Essa terminologia foi introduzida por Richards (1936) e continua em voga hoje em dia nos estudos da metáfora lingüística (p.ex Cameron, 2002).

Em suma, o identificador:

- Faz uma triagem do vocabulário do corpus, mostrando uma seleção prévia das palavras com maior potencial de uso metafórico.

- Tem a função de indicar a probabilidade de palavras fazerem parte do Veículo.
- Não visa a mostrar qual a extensão do Veículo ou da expressão metafórica.
- Não substitui o analista humano. O identificador não faz o julgamento da metaforicidade, o que só pode ser feito por um ser humano.
- Não dispensa o concordanciador. O analista precisa de um concordanciador para buscar no corpus as formas sinalizadas pelo identificador.

A Figura 2 mostra a tela de entrada do identificador.

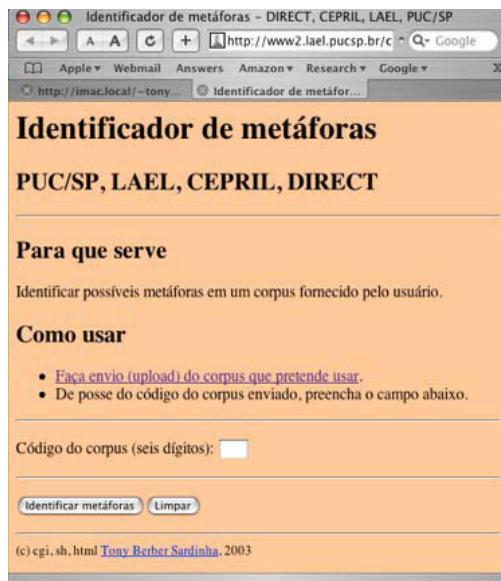


Figura 2: Tela de entrada do identificador online de metáforas

Como se nota na figura, o identificador é muito simples de usar; não há ajustes a fazer. É importante frisar que, ao contrário do

concordanciador, que aceita corpora de várias línguas, o identificador funciona apenas com textos escritos em português.

A tela de resultados do programa aparece na Figura 3.

METAPHOR TAGGER OUTPUT							
v0.4 (Feb 22, 2006)							
(c) Tony Berber Sardinha, 2005							
http://imac20.local/~tony/metaphor_tagger/							
Nome: Tony Berber Sardinha							
Descrição: 870 pronunciamentos e discursos do Presidente Lula							
Obs: Fonte: http://www.info.planalto.gov.br/download/discursos							
*** SHOWING NO MORE THAN THE 3000 MOST FREQUENT WORDS IN THE CORPUS ***							
Sorted by probability							
Thu Feb 23 16:57:39 BRT 2006							
#	Word	Tag(Prob)	Vehicle	Left Bndl	Right Bndl	Framewk	Wordclass
000001	conquistar	.9250	1.0000	1.0000	1.0000	.9210	.7041
000002	dar	.9166	.9583	1.0000	1.0000	.9210	.7041
000003	niveis	.9068	1.0000	1.0000	1.0000	.8085	.7259
000004	crescer	.9033	.9625	1.0000	1.0000	.8500	.7041
000005	segmentos	.9024	1.0000	1.0000	1.0000	.8235	.6887
000006	fazer	.9022	.8863	1.0000	1.0000	.9210	.7041
000007	passar	.9011	.8807	1.0000	1.0000	.9210	.7041
000008	passou	.8981	.9531	1.0000	1.0000	.8333	.7041
000009	caminho	.8980	.8593	1.0000	1.0000	.9423	.6887
000010	longo	.8925	.9513	1.0000	1.0000	.7857	.7259
000011	cresceu	.8924	.9890	1.0000	1.0000	.7692	.7041
000012	crescendo	.8916	.9915	1.0000	1.0000	.7627	.7041
000013	colocar	.8889	.8196	1.0000	1.0000	.9210	.7041
000014	passo	.8807	.8461	1.0000	1.0000	.8690	.6887
000015	feito	.8769	.8591	1.0000	1.0000	.8217	.7041

Figura 3: Tela de resultados de análise do identificador de metáforas

Essa tela de resultados contém as seguintes informações:

- #: número seqüencial.
- Word: palavra do corpus.
- Tag(Prob): etiqueta indicando a probabilidade de uso metafórico. É a média aritmética das probabilidades individuais seguintes:
- Vehicle: probabilidade de uso metafórico segundo o banco de dados de Veículos.

- Left Bndl: probabilidade de uso metafórico segundo o banco de dados de pacotes lexicais⁸ à esquerda da palavra.
- Right Bndl: probabilidade de uso metafórico segundo o banco de dados de pacotes lexicais à direita da palavra.
- Frame: probabilidade de uso metafórico segundo o banco de dados de enquadramentos colocacionais.
- Wordclass: probabilidade de uso metafórico segundo o banco de dados de classes gramaticais.

De posse dessas informações, já é possível usar o identificador. A seguir apresentamos maiores detalhes sobre sua criação e seu funcionamento interno. Essas informações não são, portanto, essenciais para fazer funcionar e começar a interpretar os resultados do identificador.

Conforme mencionado acima, o identificador faz uso de vários bancos de dados contendo informações sobre Veículos e as palavras que os acompanham. Esses bancos de dados foram criados a partir de uma pesquisa (Berber Sardinha, 2005) que se destina a saber quais palavras são metafóricas no português, qual seu grau de metaforicidade, quais palavras ocorrem perto delas e qual o grau de metaforicidade das classes gramaticais (substantivo, verbo, adjetivo, etc.). Essas informações foram obtidas a partir de anotação manual de metáforas em corpora (segundo o método 2, descrito neste capítulo). Cada banco de dados contém um tipo de informação específica:

- Veículos: Palavras com sentido metafórico. Esse banco de dados possui um registro de cada palavra que possuía uso metafórico nos corpora pesquisados. Pode ser visto como um dicionário de palavras metafóricas. Cada palavra aparece associada com sua probabilidade de uso metafórico. Por exemplo, a palavra ‘crescimento’ aparece nesse banco com a

probabilidade de 0,99 (isto é, 99% de chance de uso metafórico). Significa que essa palavra é praticamente sempre usada metaforicamente nos dados pesquisados, como em ‘crescimento econômico’, que exprime a metáfora conceptual A ECONOMIA É UM SER VIVO.

- Pacotes lexicais (*bundle*) à esquerda da palavra: uma seqüência fixa de três palavras imediatamente à esquerda de cada palavra do corpus (com exceção das duas primeiras, porque estas não possuem três palavras à sua esquerda). Por exemplo, ‘a necessidade de’ tem probabilidade de 0,86. Significa que essa expressão aparece muitas vezes à esquerda de um Veículo, como em ‘a necessidade de crescimento’.
- Pacotes lexicais (*bundle*) à direita da palavra: uma seqüência fixa de três palavras imediatamente à direita de cada palavra do corpus (com exceção da última e da penúltima, porque estas não possuem três palavras além delas). Por exemplo, ‘a palavra ao’ tem probabilidade de 1 (100%). Significa que essa expressão sempre apareceu logo após um Veículo nos dados pesquisados, como em ‘passar a palavra ao [fulano]’. Essa expressão está ligada à metáfora do CANAL, visto que conceitualiza a linguagem como algo que se passa e se transmite de uma pessoa a outra.
- Enquadramentos colocacionais. São seqüências fixas de três palavras que incluem um Veículo, na segunda posição. Por exemplo, o enquadramento ‘o ... de’ possui probabilidade de 0,85 (85%) e aparece com Veículos como ‘crescimento’, em expressões do tipo ‘o crescimento de [70% nos lucros]’, que evoca a metáfora conceptual UMA EMPRESA É UM SER VIVO.
- Classes gramaticais. A informação desse banco de dados consiste da probabilidade de metaforicidade de cada classe de palavra. São elas:

Adjetivo	0,73
Advérbio	0,63
Artigo	0,00
Conjunção	0,00
Numeral	0,00
Pronome	0,00
Preposição	0,61
Substantivo	0,69
Verbo	0,70

O identificador interage com esses bancos de dados da maneira descrita a seguir. Para cada palavra do corpus, ele vai:

- Extrair colocações à esquerda.
- Extrair colocações à direita.
- Extrair quadros colocacionais.
- Identificar a classe gramatical (com um etiquetador gramatical à parte, chamado de Qtag (O. Mason, 1997).
- Fazer busca de cada item no banco de dados correspondente.
- Se encontrar o item, atribuir à palavra a probabilidade constante no banco de dados.
- Se não encontrar o item, atribuir à palavra a probabilidade de .0001.

O principal banco de dados é o de Veículos, pois ele contém palavras que já foram usadas metaforicamente antes em outros textos. Mas esse banco apenas não é suficiente, pois ele é incompleto. Há

sempre mais palavras na língua do que qualquer banco de dados lexical pode abranger. Por isso, o identificador deve estar preparado para ‘adivinar’ palavras que não estejam nesse dicionário de Veículos. Além disso, esse processo de adivinhação também é importante porque mesmo palavras que já tenho uso metafórico atestado podem ser usadas não metaforicamente. O identificador deve reconhecer essa possibilidade e ajustar a probabilidade de metaforicidade de uma palavra de acordo com seu contexto, podendo esse ajuste ser feito para mais (caso o contexto seja muito favorável ao uso metafórico) quanto para menos (caso o contexto não favoreça o uso metafórico).

Para ilustrar, suponhamos que o identificador encontre a palavra ‘crescimento’ num texto. Segundo o banco de dados de Veículos, a probabilidade dessa palavra é 0,9966. Mas suponha ainda que essa palavra esteja sendo usada assim: ‘o crescimento das raízes foi prejudicado pelo solo infértil.’ Nesse caso, o uso não é metafórico. Se o programa confiasse apenas no dicionário de Veículos, diria ao usuário que essa palavra possui alta probabilidade de metaforicidade, o que seria um engano. Mas felizmente o identificador ‘olha’ para o contexto imediato da palavra e detecta que ‘crescimento’ é seguido do pacote ‘das raízes foi’. Esse pacote não faz parte do banco de dados de pacotes à direita, o que faz com que o identificador atribua valor 0,0001 para a palavra ‘crescimento’ nesse item. Isso, por sua vez, causa um ajuste do valor final da probabilidade, para baixo. Se parássemos aqui, a etiqueta final da palavra ‘crescimento’ já teria caído, dos originais 0,9966 para 0,4984 (isto é, $0,9966 \times 1$, dividido por 2). Observando a tela de resultados, o usuário poderia perceber isso, pois notaria que a coluna ‘Vehicle’ aponta um valor alto para ‘crescimento’ enquanto ‘Right Bndl’ mostra um valor perto de zero. Quando isso acontece, é sinal de que uma palavra tipicamente metafórica pode ter sido usada de modo literal.

4.4.2. Análise do corpus de pronunciamentos do presidente Lula

Aplicamos o etiquetador na análise do corpus de pronunciamentos do presidente e obtivemos uma lista com as probabilidades de uso

metafórico das palavras ocorrentes no corpus. Parte dessa lista apareceu acima na Figura 3.

Olhando a lista, o analista já tem uma visão mais apurada da presença de metáforas no corpus, do que, por exemplo, vendo a lista de freqüência de palavras (opção 3, discutida acima). As probabilidades são um bom guia inicial para o analista, muito embora, conforme explicamos, o identificador não mostre com certeza quais são as metáforas.

Tomando as palavras com alta probabilidade de uso metafórico (com 80% ou mais), temos a seguinte lista:

1	conquistar	93%	20	papel	86%
2	dar	92%	21	passando	86%
3	níveis	91%	22	abrir	86%
4	crescer	90%	23	instrumento	86%
5	segmentos	90%	24	levar	86%
6	fazer	90%	25	passos	86%
7	passar	90%	26	ganhando	85%
8	passou	90%	27	movimento	85%
9	caminho	90%	28	ganham	85%
10	longo	89%	29	instrumentos	84%
11	cresceu	89%	30	claro	84%
12	crescendo	89%	31	colocado	84%
13	colocar	89%	32	competitivo	83%
14	passo	88%	33	levando	83%
15	feito	88%	34	tirar	83%
16	baixo	88%	35	mexer	81%
17	alto	88%	36	salto	80%
18	ganhar	87%	37	peso	80%
19	tenho	86%	38	ver	80%

Quadro 2: Palavras com maior potencial de uso metafórico no corpus de pronunciamentos do presidente Lula

Essas 38 palavras estão distribuídas entre as posições 29 ('fazer') a 2594 ('competitivo') da lista de palavras. A distância entre a mais e a menos freqüente desse grupo é de 2565 posições. Há, desse modo, apenas 1,5%⁹ de palavras com alto potencial metafórico entre essas 2565 posições da lista de palavras. Isso demonstra a grande dificuldade que um analista teria para chegar a essa seleção de 38 palavras inspecionando a lista de palavras.

Para ilustrar a seqüência do procedimento, veremos as ocorrências de 'conquistar', a palavra com maior probabilidade de uso metafórico no corpus.

Das 152 ocorrências de 'conquistar', apenas três (1%) são literais (conquistar prêmio, medalhas e coisas). As demais (99%) são metafóricas, referindo-se em sua maior parte a conquistar:

- Cidadania, liberdade, democracia, espaço (num mundo globalizado), direito, credibilidade, confiança, oportunidades, felicidade, aliados.

Abaixo aparece uma concordância com citações de 'conquistar cidadania':

3 175 milhões de seres humanos que têm urgência em conquistar a sua cidadania. As reformas estruturais
 5 que o Brasil não vai perder essa oportunidade de conquistar a cidadania para todo o nosso povo. Outras
 6 essas, no tempo necessário, as condições para ela conquistar a cidadania no campo. É para isso que se
 15 empregos para que a gente possa definitivamente conquistar a cidadania do povo, a gente não pode a
 19 lembrar que o álcool estava esquecido e voltou a conquistar cidadania, não apenas internamente, mas

Figura 4: Concordância de 'conquistar cidadania'

O domínio fonte de 'conquistar' é a guerra ou a competição. Assim, nas citações acima, a metáfora conceptual envolvida pode ser (em ordem crescente de generalização):

- para ter cidadania é preciso lutar numa guerra

- para ter direitos na sociedade é preciso lutar numa guerra
- vida em sociedade é guerra

Em suma, esses resultados indicam:

- As palavras com maior probabilidade de uso metafórico parecem indicar campos fontes importantes para a construção da mensagem presidencial: GUERRA (conquistar), ESPORTE (ganhar, competitivo, colocado, salto), VIAGEM (caminho, passos), ORGANISMO (crescer), RECEPTÁCULO (níveis) e da MÁQUINA (instrumento, peso).
- Sozinhas, essas 38 palavras levantadas pelo identificador respondem por uma boa parcela do potencial metafórico do vocabulário do presidente: elas somam 19092 ocorrências, o que representa 1% do total de ocorrências de palavras do corpus. Considerando que a densidade de metáforas em corpora tem sido estimada em torno de 3% (Cameron, 2003) a 8% (Berber Sardinha, 2006b), esse 1% responderia por 13%¹⁰ a 33%¹¹ da quantidade esperada de metáforas no corpus.
- As candidatas são de alta probabilidade metafórica (acima de 80%), indicando possíveis usos metafóricos consistentes do falante.

É preciso frisar que as palavras candidatas indicadas pelo programa precisam ser analisadas uma a uma para se comprovar seu uso metafórico. A triagem do identificador serve de ponto de partida para a análise, que deve incluir o componente qualitativo.

Nesta seção, mostramos um exemplo do uso do identificador para identificar metáforas em um corpus eletrônico. Nossa intenção não é apresentar os resultados da análise do corpus, mas sim enfocar o processo de uso do identificador na análise. Para uma apresentação dos resultados da pesquisa com esse corpus, vide Berber Sardinha (em preparação).

5. Conclusão

O estudioso da metáfora, quando se defronta com um corpus eletrônico, enfrenta o desafio de identificar metáforas em grandes quantidades de dados, muito mais do que normalmente se tem feito na área. A literatura sobre metáfora demonstra claramente esse problema. Via de regra, os trabalhos usam dados inventados ou de origem desconhecida (como George Lakoff & Johnson, 1980), analisam intensamente poucos textos (como Cameron, 2003) ou poucas metáforas em muito textos (como Deignan, 2005).

Além disso, quando o referencial teórico é a metáfora conceptual (George Lakoff & Johnson, 1980), esse desafio se potencializa, pois a meta da pesquisa com metáfora conceptual é mostrar metáforas abstratas, que não possuem realização direta na língua (e.g. O AMOR É UMA VIAGEM) e, por conseguinte, no corpus. O analista precisa normalmente usar a intuição ou a memória para concretizar as metáforas conceptuais em termos de busca para inspecionar o corpus (por exemplo, no caso de AMOR É UMA VIAGEM, ele pode tentar ‘estamos indo bem’, ‘de vento em popa’, ‘estamos num beco sem saída’, etc.).

Essa situação restringe o escopo da análise de metáfora. Aparte a invenção de dados, o que tem sido feito na literatura é analisar muitas metáforas em poucos textos ou poucas metáforas em muitos textos. A opção de análise 1, mostrada acima, é o mais adequado para a primeira situação, enquanto as opções 2 e 3 são mais recomendadas para a segunda situação.

Contudo, há uma terceira situação: a necessidade de analisar muitas metáforas em muitos textos. Essa situação tende a se tornar cada vez mais comum devido à maior disponibilidade de textos em formato eletrônico, como por exemplo textos retirados da Internet. Com essa maior disponibilidade de corpus, vem a pressão para se identificar o maior número possível de metáforas no maior número possível de textos. Para que isso seja viável, é preciso usar a informática, pois os computadores são feitos justamente para desempenhar tarefas

repetidas sem se cansar. O identificador de metáforas vem justamente cumprir essa função.

O identificador é útil em trabalhos em que se tenha a preocupação de verificar todas as palavras de um corpus. Conforme dito acima, o identificador faz isso, avaliando a probabilidade de uso metafórico de cada uma das muitas palavras que existam no corpus e retornando uma lista encabeçada pelas palavras de maior probabilidade. De posse desses resultados, o analista pode:

- Enfocar as palavras com maior probabilidade. De posse de um concordanciador (conforme mostrado na seção anterior), ele pode fazer uma busca por essas palavras e julgar seu uso metafórico. Nesse caso, o etiquetador funciona como um ‘filtro’ para indicar ao analista quais palavras ele poderia dar prioridade na análise do corpus.
- Enfocar as palavras com menor probabilidade. Nesse caso, o analista tentaria descobrir as metáforas novas ou mais criativas do corpus. Para isso, ele precisaria também de um concordanciador, para localizar tais palavras no corpus e julgar sua metaforicidade.

Independente do uso que se faça dos resultados do identificador, a sua função principal é a de garantir abrangência na pesquisa em metáfora. O problema de abrangência é sério e pode ser esquematizado da seguinte maneira:

1. A metáfora é uma característica comum e natural da língua; há uma grande profusão delas ao nosso redor.
2. O processo de julgamento de metaforicidade depende da interpretação do sentido do candidato à metáfora (uma palavra, expressão ou frase identificada como possivelmente

metafórica, segundo as condições necessárias discutidas no método 2, neste capítulo).

3. Esse processo de julgamento somente pode ser feito por seres humanos. Nenhuma máquina (até hoje) conseguiu entender a linguagem humana natural de modo satisfatório, para decidir se um trecho de língua foi empregado metaforicamente ou não.
4. O ser humano, por mais bem intencionado e trabalhador que for, não consegue fazer esse julgamento muitas e muitas vezes, de modo confiável. O cansaço afeta o desempenho do ser humano em tarefas intelectuais cansativas, prejudicando a qualidade de seu julgamento.
5. Em conclusão, a interpretação humana, exigida em tarefas de análise de metáfora, é limitada, podendo ser efetuada de modo satisfatório em uma quantidade limitada de casos. Mesmo que se estenda o tempo disponível indefinidamente, isso não resolve o problema, pois os critérios, sendo subjetivos, passariam a ser aplicados de modo inconsistente à medida que o tempo passa.

Em conclusão, neste artigo argumentamos que o uso de corpora na pesquisa em metáfora coloca desafios para seus praticantes. Esses desafios não eram encontrados antes, quando a análise se restringia a apenas um ou alguns textos, ou quando os dados do analista eram inventados. Na perspectiva da teoria contemporânea dominante nos estudos da metáfora (metáfora conceptual), o uso de corpora coloca ainda mais desafios, pois não é possível localizar metáforas conceptuais diretamente em corpora. Elas precisam ser concretizadas em expressões metafóricas, o que normalmente envolve o uso da intuição e da memória, a fim de que se tornem termos de busca no concordanciador. Urge, portanto, que se encontrem maneiras de garantir abrangência e

eficiência nos estudos da metáfora com corpora, para que a riqueza de dados que qualquer corpus oferece (e que sempre nos surpreende) não seja desperdiçada.

Para terminar, é preciso ressaltar que há várias maneiras de analisar corpora eletrônicos no campo dos estudos da metáfora, conforme mostramos, e que essas maneiras podem ser combinadas. Nenhuma delas em si é suficiente para dar conta desse fenômeno tão fascinante e importante do domínio da língua quanto é a metáfora.

Notas

1. A tradução brasileira da obra intitula-se ‘Metáforas da vida cotidiana’ (Zanotto et al., 2002).
2. O termo em inglês ‘conceptual’ foi traduzido para o português por ‘conceptual’ e não por ‘conceitual’, em Zanotto et al., 2002, e tem se firmado como a tradução aceita pela comunidade.
3. ‘A conceptual metaphor is a conventional way of conceptualizing one domain of experience in terms of another, often unconsciously.’
4. <http://lael.pucsp.br/corpora>
5. <http://www2.uol.com.br/pagina20/04022004/painel.htm>
6. <http://clipping.planejamento.gov.br/Noticias.asp?NOTCod=182756>
7. http://www2.correioweb.com.br/cw/EDICAO_20030822/pri_pol_220803_134.htm
8. Este e os demais termos desta lista são explicados a seguir.
8. Isto é, 38/2565.
10. Isto é, 1 / 8.
11. Isto é, 1 / 3.

Referências

- Berber Sardinha, T. (2004). *Lingüística de corpus*. São Paulo: Manole.
- Berber Sardinha, T. (2005). O léxico metafórico do português, *V Encontro de Corpora*. UFSCar.
- Berber Sardinha, T. (2006a). A corpus-based tagger for metaphors.
- Berber Sardinha, T. (2006b). Collocation lists as instruments for metaphor detection in corpora.
- Berber Sardinha, T. (em preparação). Metáforas do presidente lula.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics - investigating language structure and use*. Cambridge: Cambridge University Press.
- Black, M. (1962). *Models and metaphors; studies in language and philosophy*. Ithaca, N.Y.: Cornell University Press.
- Bod, R., Hay, J., & Jannedy, S. (2003). *Probabilistic linguistics*. Cambridge, Mass: MIT Press.
- Cameron, L. (2002). Metaphors in the learning of science: A discourse focus. *British Educational Research Journal*, 28(5).
- Cameron, L. (2003). *Metaphor in educational discourse*. London: Continuum.
- Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Halliday, M. A. K. (1994). *An introduction to functional grammar*. London: Edward Arnold.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh : The embodied mind and its challenge to western thought*. New York: Basic Books.
- Mason, O. (1997). Qtag-a portable probabilistic tagger. Online document, corpus research, University of Birmingham, UK, available at <http://www-clg.Bham.Ac.Uk>.
- Mason, Z. (2004). Cormet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1), 23-44.
- Richards, I. A. (1936). *The philosophy of rhetoric*. Oxford: Oxford University Press.
- Sinclair, J. (1987). *Looking up: An account of the Cobuild project in lexical computing and the development of the Collins Cobuild English language dictionary*. London: Collins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.