

# METÁFORAS E LINGUÍSTICA DE CORPUS: METODOLOGIA DE ANÁLISE APLICADA A UM GÊNERO DE NEGÓCIOS

## (Metaphors and Corpus Linguistics: A method for finding metaphors in a business genre)

Tony BERBER SARDINHA\*

**Resumo:** *O presente trabalho visa a relatar o desenvolvimento de uma metodologia de identificação de metáforas em corpora eletrônicos. Como exemplo, foi tomado o gênero teleconferências de apresentação de resultados financeiros. A metodologia é do tipo “bottom-up” / “corpus-driven” e se baseia na identificação de palavras com frequência marcante (palavras-chave) e de seus padrões de co-ocorrência, seguido do cálculo de similaridade semântica entre essas palavras. Com isso, chega-se a um conjunto de palavras que são então interpretadas em seu co-texto, por meio de concordâncias.*

**Palavras-chave:** *Linguística de Corpus; metáfora; negócios; teleconferências.*

**Abstract:** *This paper aims at reporting the development of a method for metaphor identification in computer corpora. The method was tested on a particular corpus, namely of investment conference calls, and comprises procedures that work from the bottom up, and rely on marked frequency, collocation and semantic similarity as signalling devices for metaphor. As such, the method is an example of corpus-driven research into metaphor. The application of these procedures yields a number of metaphor candidates, which are then checked manually through concordances.*

**Key-words:** *Corpus Linguistics; metaphor; business; conference calls.*

### 1. INTRODUÇÃO

O presente trabalho visa a relatar o desenvolvimento e a aplicação de uma metodologia de identificação de metáforas em corpora eletrônicos. Tal metodologia foi testada em um corpus de um gênero específico do âmbito dos negócios, qual seja, o de teleconferências de apresentação de

---

\* O autor agradece os auxílios fornecidos pelo CNPq (Brasília, Brasil) e pela CAPES. Também é grato aos pareceristas pela leitura cuidadosa e sugestões.

resultados financeiros. Tais teleconferências são realizadas por bancos da área de investimento, e dela participam investidores que possuem fundos aplicados naquele banco. Os participantes interagem a distância, por meio do telefone. As teleconferências estão estruturadas em dois blocos principais: no primeiro, há um monólogo, ou uma série deles, feitos por diretores do banco, que estão em geral na sede da empresa; no segundo, acontecem perguntas e respostas, dos vários participantes, espalhados pelo mundo, aos diretores do banco. As transcrições dessas teleconferências são, a seguir, disponibilizadas na Internet, para registro e referência futura.

Há alguns trabalhos voltados para o exame de metáforas no âmbito empresarial. O mais célebre é Morgan {, 1996 #7274} que discute as maneiras como as empresas são geridas a partir de metáforas organizacionais. Segundo o autor, ao longo dos anos, certas metáforas têm dominado as maneiras segundo as quais as empresas têm criado administradas, tais como empresa como máquina, empresa como organismo e empresa como cérebro. Moon et al. (2003) enfocam a problemática da adoção do lema de cidadania pelas empresas, a partir da perspectiva da metáfora empresa como cidadão. Os autores questionam essa prática, refutando a legitimidade da adoção da metáfora do cidadão quando ela se dá nos planos legal e estatutário. Milne et al. (2003) investigam a metáfora da jornada (journey), utilizada em relatórios ambientais (sustainability reports). Neles, as empresas evitam referir à questão da degradação do meio-ambiente decorrente de suas práticas convidando a sociedade a se juntarem a elas na jornada em direção à sustentabilidade. Esses trabalhos são voltados à administração de empresa e à teoria organizacional. Nesse sentido, seu foco principal é a análise de uma metáfora em particular sem preocupação em mostrar como elas se mostram lingüisticamente. Tais estudos tendem a tomar como base de suas conclusões um exame parcial de exemplares de textos, sem levar em conta as especificidades dos gêneros estudados.

Por sua vez, a escolha do gênero teleconferências de apresentação de resultados financeiros se deve a alguns fatores. Primeiramente, trata-se de um gênero que começa a ser usado no país, por grandes empresas, a respeito do que sabemos muito pouco, do ponto de vista discursivo. Ainda não há pesquisas disponíveis sobre a linguagem empregada nesse gênero. Em segundo lugar, as teleconferências são importantes para a comunidade em que está inserida, notadamente investidores, acionista, diretores e gestores

de carteiras de investimento. As teleconferências executam a importante tarefa de prover um canal de comunicação em tempo real entre o público interessado e a empresa, coisa que outros gêneros, como os relatórios de resultados (trimestrais, semestrais e anuais), impressos, não proporcionam. Por fim, as teleconferências se apresentam como um gênero híbrido, mediado pela tecnologia, que acontece essencialmente pelo canal acústico, mas que engloba elementos visuais: trata-se de uma interação levada a cabo via contato telefônico múltiplo entre as partes envolvidas, às vezes com o apoio audiovisual de transparências e apresentações multimídia (em PowerPoint, por exemplo), disponibilizadas via Internet. Além disso, muitas vezes os apresentadores trazem textos prontos que são lidos em voz alta. Há, ainda, muitas vezes, a participação do público por meio de perguntas que são respondidas pelos representantes da empresa. É, portanto, um gênero interessante, pois mistura elementos escritos e falados, monológicos e dialógicos, preparados e espontâneos.

Desde a publicação de *Metaphors we live by*, por George Lakoff e Mark Johnson (1980), há mais de 20 anos, tem-se visto um interesse crescente e contínuo nas questões relacionadas à metáfora. Metáfora é entendida aqui como a expressão de um domínio conceptual em termos de outro, de tal modo que (1) um domínio é compreendido em termos do outro, e (2) os dois domínios são diferentes. Daí o nome de metáfora conceptual<sup>1</sup>. Assim, por exemplo, TEMPO É DINHEIRO é uma metáfora conceptual, pois há interligação de dois domínios diferentes, de tempo e de dinheiro, sendo tempo (o tópico, ou alvo) entendido como dinheiro (o veículo, ou fonte). Numa metáfora conceptual distinguem-se dois domínios, o fonte (source domain) e o alvo (target domain), que são unidos metaforicamente por meio de mapeamentos (correspondências). O domínio alvo é aquele que se deseja entender, enquanto o fonte é aquele a partir do qual se entende o alvo. No caso de TEMPO É DINHEIRO, o domínio referente a tempo é o alvo que é entendido em termos do domínio referente a dinheiro, que é a fonte. Os mapeamentos indicam, por exemplo, que tempo é algo que se economiza, perde, ganha, sobre o qual se tem posse, etc.

---

1. Adotaremos a forma “conceptual” em vez de “conceitual”, para seguirmos o uso estabelecido na tradução do volume de Lakoff e Johnson para o português do Brasil, por Mara Sofia Zanotto e equipe (Lakoff & Johnson, 2002).

Um dos problemas com os estudos de metáfora que seguem a orientação de Lakoff e Johnson (1980) reside na metodologia de identificação das metáforas. Os próprios Lakoff e Johnson (1980) não deixam claro quais os procedimentos que utilizaram para encontrar as metáforas que apresentam.

Numa tentativa de melhorar esse problema, Steen (1999) apresenta uma lista de procedimentos (checklist) para identificação de metáforas. Essa lista inclui uma ampla gama de elementos a serem analisados, divididos em blocos destinados à análise conceitual, lingüística e comunicativa. Em cada bloco, o pesquisador deve empreender uma análise pautando-se por critérios formais, que levam em conta o vocabulário, a sintaxe, a função pragmática e a forma retórica, entre outros.

Uma abordagem que crie condições para a exploração de metáforas usando corpora eletrônicos e computador já foi sugerida por alguns pesquisadores. Conforme defende Steen (1999, 81):

“if one wants to discriminate between types of metaphor embodying specific configurations of metaphor features of all kinds, corpus research is crucial. Corpus work is also imperative if the object of investigation is the difference between literary and non-literary metaphor... Moreover, corpus research can yield realistic materials for rating studies of metaphor, offering an opportunity to establish the desirable connection between analytic metaphor properties produced in linguistic research, on the one hand, and informants’ judgements on the other.”

Com um corpus eletrônico é possível fazer um levantamento mais abrangente da presença e da ausência de metáforas:

“the study of large corpora can give information about the frequency and use of linguistic metaphors which is otherwise difficult to access. ... A computerized corpus can enable the researcher to detect patterns of usage more quickly than either the use of intuition or the analysis of individual texts, as words or expressions are automatically retrieved from the corpus and sorted.” (Deignan, 1999, 177-178)

É desejável que tal abordagem de identificação de metáforas seja, portanto, baseada em corpora eletrônicos (Deignan, 2005). Isso se constitui numa abordagem recente; tanto assim, que uma introdução atual, abrangente e extensa aos estudos da metáfora (Kövecses, 2002) sequer menciona o uso de corpora.

## 2. METÁFORA

A publicação de *Metaphors we live by* (Lakoff e Johnson, 1980), suscitou um interesse crescente nos estudos da metáfora da comunicação humana, especialmente com uma orientação cognitiva nas questões relacionadas às metáforas conceituais.

Um conceito central numa visão cognitiva da metáfora é o de metáfora conceitual. Metáforas conceituais definem maneiras como pensamos e experienciamos a realidade. Elas são formadas pelo mapeamento de um domínio (ou campo conceitual) em outro, de tal forma que um é entendido por meio de outro. Um domínio conceitual é definido como um conjunto que organiza a experiência de modo coerente (Kövecses, 2002: 4). Por exemplo, AMOR É UMA VIAGEM (LOVE IS A JOURNEY), sendo uma metáfora conceitual, relaciona o domínio de AMOR com o de VIAGEM, de tal modo que a experiência de amar seja entendida de acordo com a experiência de viajar.

É preciso fazer uma distinção entre metáfora conceitual e metáfora lingüística (ou expressão lingüística metafórica). Uma metáfora conceitual indica as maneiras pelas quais o pensamento se organiza e como as pessoas interagem, enquanto uma metáfora lingüística representa a realização, na fala ou escrita, de uma metáfora conceitual. Em outras palavras, a expressão lingüística remete à representação metafórica conceitual. Normalmente, as metáforas conceituais não possuem correlatos lingüísticos diretos. Por exemplo, as pessoas geralmente não dizem ou escrevem DISCUSSÃO É GUERRA. Uma das exceções é TEMPO É DINHEIRO, que é uma metáfora conceitual e lingüística ao mesmo tempo.

Há três tipos básicos de metáforas conceituais. O primeiro são as estruturais, em que os domínios presentes são bem estruturados e permitem o estabelecimento de um conjunto grande de mapeamentos entre os domínios. Por exemplo, na metáfora TEMPO É MOVIMENTO (TIME IS MOTION), podemos estabelecer vários mapeamentos entre o domínio fonte (TEMPO) e o alvo (MOVIMENTO): “tempo é uma coisa, a passagem do tempo é um movimento, o tempo futuro está à frente do observador” (Kövecses, 2002 :33). Esses mapeamentos nos levam a compreender o tempo em termos de elementos como objetos, em relação à sua localização e seu deslocamento, de tal modo que o tempo presente situa-se no mesmo local em que o observador, ou

falante. Assim, uma expressão como o tempo não pára faz referência ao deslocamento do tempo, como se ele estivesse em movimento contínuo.

O segundo tipo de metáfora é a ontológica, em que os domínios envolvidos são menos estruturados. Metáforas desse tipo envolvem operações em que algo abstrato é visto como concreto, ou em que algo concreto mas inanimado é representado como algo animado (personificação).

Por exemplo, na metáfora INFLAÇÃO É UMA ENTIDADE (Lakoff e Johnson, 1980: 26), a inflação, que é um conceito, portanto abstrato, é caracterizado como algo concreto. Assim, temos expressões do tipo “a inflação está correndo os salários”. A personificação é um recurso comum para a metáfora ontológica, sendo a computação um campo fértil para esse tipo de metáfora. Expressões do tipo “o computador comeu meus arquivos” baseiam-se na atribuição de qualidades humanas a uma máquina, de tal modo que o computador assume qualidades normalmente atribuídas a pessoas.

O terceiro tipo de metáfora é a orientacional. Nesse tipo de metáfora, as relações são formadas por meio de conceitos de direção, como para cima e para baixo. Um exemplo é a metáfora VIRTUDE É PARA CIMA; FALTA DE VIRTUDE É PARA BAIXO que possui realizações lingüísticas como levantar / baixar o nível, em que nível é entendido como educação, polidez, etc., que por sua vez são virtudes humanas: quando há falta de tais virtudes, o nível baixa, ao passo que quando tais virtudes estão presentes, o nível está alto. Os domínios que as metáforas orientacionais evocam são ainda menos estruturados do que aqueles envolvidos nas metáforas ontológicas, pois referem-se a conceitos relativos ao espaço físico. Para Kövecses (2002: 35), as metáforas orientacionais, na verdade, vão além da simples identificação da direção do movimento e da correspondente atitude ou reação ligadas a esse movimento. Segundo ele, tais metáforas possuem uma função de organizar coerentemente grupos de metáfora:

“their cognitive job is to make a set of target concepts coherent in our conceptual system. ... It would perhaps be more appropriate to call this type of conceptual metaphor “coherence metaphor”, which would be more in line with the cognitive function these metaphors perform.” (Kövecses, 2002: 36-37).

Assim, de um modo geral (no inglês), a orientação para cima parece indicar uma situação favorável, enquanto a direção contrária parece associada a situações negativas. Por isso, os domínios “para cima e para baixo” emprestam coerência a um grupo de metáforas.

Apesar de a orientação cognitiva trazer contribuições fundamentais sobre a metáfora e como ela se relaciona com o pensamento e a experiência, há várias questões importantes que essa linha não contempla. Entre elas está a questão da autenticidade dos dados. Gibbs (2002: 4) expressa esta crítica do seguinte modo:

“Analyses of metaphor in language and in thought are all based on individual analysts’ intuitions, which are often tied to theories/hypotheses they wish to support. ... Examples are made-up, and presumably reflect fundamental aspects of the “idealized” speaker-hearer, and not how people ordinarily speak or write in naturalistic discourse.”

Ou, segundo Cameron:

“many of the definitions [of metaphor] were based on, or validated through, decontextualised and constructed examples, often restricted in length to sentence-level or below.” (Cameron, 1999:106)

Isso nos traz de volta à questão do uso de corpora na pesquisa de metáforas. A disciplina que se ocupa do estudo de corpora eletrônicos é a Lingüística de Corpus, da qual trataremos na seção a seguir.

### 3. LINGÜÍSTICA DE CORPUS

A proposta de investigação de metáforas formulada neste trabalho baseia-se na Lingüística de Corpus (Berber Sardinha, 2004). A metodologia, mais especificamente, emprega a análise de padrões de associação (Biber, Conrad, & Reppen, 1998; Hunston, 2002).

A Lingüística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador (Berber Sardinha, 2004).

A Lingüística de Corpus adota uma visão probabilística da linguagem. A visão da linguagem enquanto sistema probabilístico prevê que embora muitos traços (features) lingüísticos sejam possíveis teoricamente, eles não

ocorrem com a mesma frequência. Por exemplo, a frequência de substantivos (no inglês e, com certeza, no português) é maior do que qualquer outra categoria; cerca de 25% das palavras são substantivos (Kennedy, 1998:103). Desse modo, a probabilidade de uma palavra ser um substantivo é maior do que outra classe gramatical.

Uma outra característica da visão probabilística de linguagem adotada pela Lingüística de Corpus é a de que as palavras atraem-se umas às outras com certa probabilidade. Essa atração mútua entre certas palavras cria padrões de co-ocorrência não aleatórios. Distingue-se normalmente três padrões na análise de corpus:

(1) Colocação: associação entre itens lexicais. Por exemplo, “café” e “manhã” possuem probabilidade alta de co-ocorrência no português, formando a colocação “café da manhã”.

(2) Coligação: associação entre itens lexicais e gramaticais ou entre categorias gramaticais. Por exemplo, substantivo seguido de adjetivo é um padrão comum em português.

(3) Prosódia semântica: associação entre itens lexicais e uma postura avaliativa (p.ex. negativa, positiva ou neutra). Por exemplo, no inglês, “cause” tem uma prosódia semântica negativa, pois associa-se a palavras desfavoráveis como problem(s), damage, death(s), disease, concern e cancer.

Porém, embora a Lingüística de Corpus possa oferecer meios para um estudo de metáforas de modo mais abrangente e com dados autênticos, ela não coloca à disposição do analista de metáforas métodos prontos para a extração de metáforas de corpora eletrônicos. Os métodos existentes ainda estão em desenvolvimento. O presente estudo pretende contribuir nesse sentido apresentando uma metodologia de identificação de metáforas em corpora eletrônicos.

#### 4. OBJETIVOS

Os objetivos do trabalho são:

1. Desenvolver e aplicar uma metodologia para identificação de metáforas com corpora eletrônicos baseada na presença de colocações.

2. Identificar metáforas no corpus de estudo.
3. Mapear as metáforas lingüísticas em metáforas conceptuais.
4. Compreender como a presença das metáforas contribui para a caracterização do gênero teleconferências de apresentação de resultados.

## 5. METODOLOGIA

O corpus de estudo é composto de 17 teleconferências conduzidas por um grande banco brasileiro, cujas transcrições estavam disponíveis na WWW. As teleconferências foram realizadas em português e aconteceram durante vários anos, desde o final da década de 1990 até o início dos anos da década de 2000. O corpus possui 98.515 tokens e 6.160 types. O objetivo das teleconferências era comunicar aos investidores a situação de seus investimentos e a do banco em geral. Os participantes eram diretores do banco, de um lado, e investidores ou representantes de fundos de investimentos do outro. Também participaram a equipe técnica, que operava os telefones e a ordem dos falantes e também organizava a sessão de perguntas e respostas. As pessoas estavam situadas em vários lugares do Brasil e, às vezes, do mundo. Os participantes tinham acesso a arquivos PowerPoint com gráficos e tabelas para acompanharem os comentários.

A metodologia de análise dessas teleconferências tem como ponto de partida o tipo de evidência apresentada por Lakoff e Johnson (1980) para embasar suas metáforas conceptuais. Segundo os autores, por exemplo, no inglês americano, ARGUMENT IS WAR (DISCUSSÃO É GUERRA) é uma metáfora porque as pessoas discutem usando vocabulário semelhante ao que empregam para falar da guerra e vice-versa (Lakoff e Johnson, 1980: 7). Ou seja, há léxico compartilhado no uso dos conceitos de war e to argue. Traduzindo essa abordagem para os termos da Lingüística de Corpus, diríamos que os padrões de associação (Biber, 1996) de ARGUMENT são em boa parte semelhantes aos colocados de WAR. Daí o mapeamento entre os dois campos conceptuais. Uma das coisas que sinaliza esse mapeamento metafórico é a distância semântica considerável entre ARGUMENT e WAR (vide abaixo). Ou seja, os sentidos de ARGUMENT e WAR são distintos, a não ser quando considerados metaforicamente: a definição de ARGUMENT encontrada em um dicionário não contempla o sentido de WAR e vice-versa. Esse mapeamento

é, ainda, cultural, relativo, pois outras culturas podem entender conceitos como argumentar de outro modo e não como guerra.

Os procedimentos específicos são descritos abaixo e se constituem numa abordagem bottom-up/corpus-driven de identificação de metáforas:

1. Identificação das palavras-chave do corpus. Palavras-chave (key words) são aquelas cujas frequências são estatisticamente diferentes das encontradas em um corpus de referência, calculadas por meio da ferramenta KeyWords, do programa WordSmith Tools (Scott, 1997).

2. Seleção de uma amostra de palavras-chave. A amostra foi escolhida de tal modo que as palavras que fizeram parte da amostra pareciam indicar assuntos, conceitos e participantes centrais para o gênero.

3. Identificação dos conjuntos de colocados de cada palavra-chave. Os colocados (collocates) são as palavras que ocorrem nas imediações de uma outra palavra (o nóculo, ou node word, que, em nosso caso, são as palavras-chave). Os colocados mostram a padronização do nóculo, que por sua vez revela como essa palavra é usada, em termos de suas colocações, coligações e prosódia semântica. A padronização, em resumo, indica os sentidos associados a ela (Sinclair, 2003). A identificação dos conjuntos de colocados foi feita com o programa Concord, do WordSmith Tools. Foram feitas concordâncias dessas palavras e a computação dos colocados de cada um, com a função show collocates. O horizonte estabelecido para inclusão dos colocados foi de 5 palavras para cada lado do nóculo. A frequência mínima exigida dos colocados foi 2. A tabela de colocados oferecida pelo comando show collocates do WordSmith Tools foi então transferida para o Analisador de tabela de colocados do WordSmith Tools, disponibilizado pelo CEPRIIL / LAEL, no endereço <http://lael.pucsp.br/corpora>. Foram calculadas as estatísticas de Mutual Information e de T-Score para cada colocado, em relação ao nóculo (palavra-chave). O ponto de corte para a seleção dos colocados foram os valores de  $MI > 3$  ou  $T > 2$ . Esses valores foram sugeridos por Stubbs (2001). Isso significa que um colocado que possuísse o valor mínimo exigido em pelo menos uma das estatísticas foi aceito. A razão da duplicidade de estatísticas é que elas se completam. Enquanto Mutual Information indica a força da colocação, T-Score exprime certeza da colocação (Hunston, 2002: 73). Isso significa que Mutual Information tende a privilegiar palavras de baixa frequência mas que tenham tendência de

ocorrer juntas muitas vezes; já T-Score tende a identificar pares de palavras que ocorram muitas vezes devido à alta frequência de um dos (ou os dois) elementos. Os conjuntos de colocados foram, então, formados somente por aqueles colocados que passaram pelo ponto de corte de MI ou de T.

4. Identificação do grau de *similaridade semântica* (ou de distância semântica) entre as palavras-chave e seus colocados. A intenção é identificar quais pares de palavras são potencialmente indicativos de mapeamentos metafóricos: expressões metafóricas tendem a possuir palavras com similaridades semânticas menores do que palavras de outras expressões (Berber Sardinha, 2002). Por exemplo, *ideas* e *plants*, que fazem parte da metáfora *IDEAS ARE PLANTS* apresentada em Lakoff e Johnson (1980: 47), possuem uma similaridade de 1,67, enquanto *car* e *automobile*, considerados praticamente sinônimos, possuem um escore de similaridade igual a 3,47. Já palavras não relacionadas, seja por metáfora seja por outra ligação aparente, demonstram uma similaridade semântica mais baixa ainda. Por exemplo, *elephant* e *pencil* possuem um escore de apenas 0,76. Os pontos de corte sugeridos por Berber Sardinha (2002) são 1 e 2, ou seja, os potenciais candidatos a metáfora possuem escore de similaridade que se situa na faixa até 2 de grau de distância semântica segundo aferida pelo teste Leacock-Chodorow. A similaridade semântica é uma medida de proximidade (ou de distanciamento) de sentido entre duas palavras (ou expressões) (Leacock & Chodorow, 1998). No cálculo da distância semântica são levados em conta várias relações de sentido, tais como sinonímia, meronímia, antonímia, etc. Para o cálculo, serão consideradas as relações documentadas em um dicionário eletrônico, mais especificamente, o WordNet, que é um thesaurus que contém cerca de 129 mil palavras, do inglês. Atualmente, ele está disponível para a língua inglesa, o que significa que as palavras-chave serão traduzidas antes de terem sua distância semântica aferida. Infelizmente o WordNet para o português ainda está em preparação. O cálculo da similaridade foi feito com a rotina em Perl chamada *distance* (Pedersen & Patwardhan, 2002), empregando uma das medidas de distância estatística disponíveis na rotina, como *lch* (Leacock e Chodorow), utilizando a instalação online do programa (versão 0.05) disponível em <http://lael.pucsp.br/corpora/similarity>.

5. Neste ponto, chega-se a uma lista das metáforas do corpus, que foi interpretada e classificada conceptualmente.

## 6. ANÁLISE DO CORPUS EM BUSCA DOS CANDIDATOS À METÁFORA

Conforme apresentado acima, o primeiro passo da análise consiste na identificação de palavras-chave. Assim, foi preparada uma lista de frequência de palavras do corpus de estudo com o programa WordSmith Tools (Scott, 1997). A seguir, com o programa KeyWords do WordSmith Tools essa lista foi contrastada com a lista de frequência de palavras do Banco de Português (Berber Sardinha, 2003), que possui as dimensões a seguir:

Tokens	230.460.560
Types	607.392

**Quadro 1. Corpus de referência**

Foram fornecidas pelo KeyWords as primeiras 500 palavras, ordenadas por significância estatística, segundo aferida pelo teste Log-Likelihood.

Dessas 500 palavras, foi feita então a seleção daquelas que possuíam centralidade temática em relação ao gênero teleconferências. A seleção foi feita intuitivamente, baseando-se no conhecimento do pesquisador sobre o contexto, os participantes e a temática envolvida nesses gêneros. Desse modo, das 500 palavras-chave iniciais, chegou-se a uma lista com 14 palavras-chave, que são:

PALAVRA	TELEC*	%	BP**	%	KEYNESS	P
BANCO_NOME***	707	0,72	5437	<0.01	6602,0	0,00000
TRIMESTRE	588	0,60	9895	<0.01	4612,8	0,00000
BANCO	919	0,93	137496	0,06	3346,6	0,00000
CRÉDITO	394	0,40	30582	0,01	1923,6	0,00000
CARTEIRA	272	0,28	12102	<0.01	1620,2	0,00000
RESULTADO	256	0,26	56073	0,02	751,2	0,00000
CLIENTES	168	0,17	20643	<0.01	672,8	0,00000
AQUISIÇÃO	112	0,11	5241	<0.01	656,0	0,00000
TÍTULOS	155	0,16	3818	0,02	421,9	0,00000
ÍNDICE	131	0,13	40254	0,02	305,5	0,00000
MERCADO	245	0,25	149939	0,06	297,7	0,00000
PATRIMÔNIO	85	0,09	15454	<0.01	278,2	0,00000
DESPESA	24	0,02	3127	<0.01	93,4	0,00000
RECEITA	66	0,07	37183	0,02	88,3	0,00000

\* Freqüências no corpus de teleconferências.

\*\* Freqüências no Banco de Português, o corpus de referência.

\*\*\* Código criado para se referir ao nome do Banco, a fim de proteger sua identidade.

**Quadro 2. Palavras-chave estudadas**

Essas 14 palavras indicam os principais conceitos, temas e participantes envolvidos nas teleconferências.

Entre os participantes, figuram aí o próprio Banco (representado pela palavra banco e pelo nome da empresa, codificado como banco\_nome), o cliente e o mercado. Também aparecem os principais elementos presentes nas transações financeiras mencionadas nas teleconferências: carteira (de crédito), patrimônio, índice (de algum dado econômico ou da bolsa de valores) e título, além da designação das próprias atividades financeiras, como despesa, receita e resultado e da referência temporal padrão na contabilidade, o trimestre.

O próximo passo referiu-se à identificação das colocações e seus respectivos conjuntos de colocados referentes às 14 palavras-chave.

Os colocados selecionados apareceram em 78 colocações, que são listadas no quadro a seguir.

nódulo	Colocado	fc	fnc	MI	T	distância	Campo fonte
BANCO	ACESSAR	8	4	3,422	1,815	1,856	PONTE
CRÉDITO	ACESSAR	8	3	4,228	1,642	1,856	PONTE
BANCO_NOME*	ACESSAR	8	4	3,800	1,857	1,856	PONTE
BANCO	ALAVANCAR	4	2	3,422	1,283	1,856	MÁQUINA
BANCO	ATINGIRAM	5	2	3,098	1,251	2,079	COMPETIÇÃO
CARTEIRA	ATINGINDO	8	3	4,762	1,668	1,163	COMPETIÇÃO
CARTEIRA	ATINGIR	11	2	3,717	1,308	1,163	COMPETIÇÃO
BANCO	AUMENTAMOS	8	3	3,006	1,518	1,856	PARA CIMA É BOM
PATRIMÔNIO	AUMENTOU	8	2	5,855	1,393	1,520	PARA CIMA É BOM
RECEITA	AUMENTA	3	3	8,220	1,726	1,163	PARA CIMA É BOM
RECEITA	AUMENTO	46	2	3,697	1,308	1,163	PARA CIMA É BOM
BANCO	CONSERVADORA	11	6	3,547	2,241	1,163	POLÍTICA
CRÉDITO	CONSERVADORA	11	2	3,184	1,262	1,163	POLÍTICA
BANCO	CRESCER	29	7	2,370	2,136	2,079	ORGANISMO
BANCO	CRESCENDO	67	10	1,677	2,174	2,079	ORGANISMO
BANCO	CRESCIMENTO	281	31	1,240	3,214	2,079	ORGANISMO
CARTEIRA	CRESCEMOS	11	3	4,303	1,645	1,856	ORGANISMO
CARTEIRA	CRESCER	29	7	4,127	2,495	0,000	ORGANISMO
CARTEIRA	CRESCENDO	67	6	2,696	2,074	0,000	ORGANISMO
CARTEIRA	CRESCIMENTO	281	13	1,744	2,531	0,000	ORGANISMO
CRÉDITO	CRESCEMOS	11	4	4,183	1,892	0,000	ORGANISMO
CRÉDITO	CRESCER	29	6	3,370	2,215	0,000	ORGANISMO
CRÉDITO	CRESCER	31	5	3,010	1,961	0,000	ORGANISMO
CRÉDITO	CRESCENDO	67	9	2,746	2,555	2,079	ORGANISMO

CRÉDITO	CRESCIMENTO	281	21	1,901	3,357	2,079	ORGANISMO
MERCADO	CRESCIDO	17	4	4,241	1,895	2,079	ORGANISMO
PATRIMÔNIO	CRESCU	31	2	3,902	1,322	2,079	ORGANISMO
RECEITA	CRESCU	31	3	4,851	1,674	2,079	ORGANISMO
TRIMESTRE	CRESCU	31	6	2,696	2,072	1,856	ORGANISMO
TRIMESTRE	CRESCIMENTO	281	36	2,100	4,602	1,856	ORGANISMO
CARTEIRA	DETERIORAÇÃO	29	11	4,779	3,196	1,520	ORGANISMO
MERCADO	DETERIORAÇÃO	29	3	3,055	1,524	1,674	ORGANISMO
ÍNDICE	DETERIORAÇÃO	29	4	4,374	1,905	2,079	ORGANISMO
BANCO	ESTRATÉGIAS	5	2	3,098	1,251	0,981	GUERRA
BANCO	ESTRATÉGIA	53	11	2,152	2,572	0,981	GUERRA
DESPESA	REESTRUTURAÇÃO	23	3	6,741	1,717	0,000	CONSTRUÇÃO
BANCO_NOME*	ESTRUTURA	8	4	3,800	1,857	1,856	CONSTRUÇÃO
CARTEIRA	ESTÁVEL	27	3	3,008	1,518	1,269	CONDIÇÕES DO TEMPO
ÍNDICE	ESTÁVEL	27	3	4,062	1,631	2,079	CONDIÇÕES DO TEMPO
BANCO	EVOLUÇÃO	132	15	1,283	2,284	1,856	ORGANISMO
CARTEIRA	EVOLUÇÃO	132	11	2,592	2,768	0,000	ORGANISMO
CRÉDITO	EVOLUÇÕES	11	3	3,768	1,607	0,000	ORGANISMO
CRÉDITO	EVOLUÇÃO	132	17	2,686	3,484	0,000	ORGANISMO
PATRIMÔNIO	EVOLUINDO	16	2	4,856	1,368	0,000	ORGANISMO
RECEITA	EVOLUIU	4	2	7,220	1,407	1,856	ORGANISMO
RESULTADO	EVOLUINDO	16	2	3,264	1,269	1,856	ORGANISMO
RESULTADO	EVOLUÇÃO	132	8	2,220	2,222	2,079	ORGANISMO
TRIMESTRE	EVOLUÍRAM	2	2	5,065	1,375	2,079	ORGANISMO
TRIMESTRE	EVOLUINDO	16	8	4,065	2,660	1,856	ORGANISMO
ÍNDICE	EVOLUINDO	16	2	4,231	1,340	1,856	ORGANISMO
BANCO	FOCADO	5	3	3,684	1,599	1,856	INSTRUMENTO ÓTICO
BANCO	FOCAR	4	2	3,422	1,283	1,856	INSTRUMENTO ÓTICO
BANCO_NOME*	FOCADO	5	3	4,062	1,631	1,856	INSTRUMENTO ÓTICO
MERCADO	LÍDER	14	6	5,105	2,380	1,386	COMPETIÇÃO
MERCADO	LÍDERES	6	2	4,743	1,364	1,386	COMPETIÇÃO
CARTEIRA	MOSTRA	79	6	2,459	2,004	1,520	PERSONIFICAÇÃO
MERCADO	MOSTRADO	11	2	3,869	1,318	2,079	PERSONIFICAÇÃO
RESULTADO	MOSTRANDO	36	4	3,096	1,767	1,856	PERSONIFICAÇÃO
RESULTADO	MOSTRA	79	7	2,768	2,258	1,856	PERSONIFICAÇÃO
TRIMESTRE	MOSTRANDO	36	6	2,481	2,013	2,079	PERSONIFICAÇÃO
TRIMESTRE	MOSTRA	79	11	2,221	2,607	2,079	PERSONIFICAÇÃO
ÍNDICE	MOSTRA	79	9	4,098	2,825	1,674	PERSONIFICAÇÃO
ÍNDICE	MOSTRAMOS	35	3	3,687	1,599	1,674	PERSONIFICAÇÃO
ÍNDICE	MOSTRANDO	36	2	3,061	1,248	1,674	PERSONIFICAÇÃO
CARTEIRA	POSITIVAS	4	2	5,177	1,375	0,981	PERSONIFICAÇÃO
RESULTADO	POSITIVAMENTE	5	2	4,944	1,371	0,981	PERSONIFICAÇÃO
RESULTADO	POSITIVO	44	6	3,390	2,217	1,856	PERSONIFICAÇÃO
RESULTADO	POSITIVA	18	2	3,096	1,251	1,856	PERSONIFICAÇÃO

TRIMESTRE	POSITIVAMENTE	5	3	4,328	1,648	1,068	PERSONIFICAÇÃO
CARTEIRA	REDUÇÃO	79	14	3,681	3,450	1,269	PARA BAIXO É RUIM
CRÉDITO	REDUZIMOS	7	2	3,836	1,318	1,856	PARA BAIXO É RUIM
CRÉDITO	REDUÇÃO	79	14	3,146	3,321	1,856	PARA BAIXO É RUIM
TRIMESTRE	REDUZINDO	3	2	4,481	1,354	1,269	PARA BAIXO É BOM
BANCO	SÓLIDOS	1	2	5,421	1,382	1,386	CONSTRUÇÃO
RESULTADO	SÓLIDO	14	4	4,457	1,910	1,674	CONSTRUÇÃO
MERCADO	VOLUMES	10	2	4,006	1,329	1,674	MASSA
TRIMESTRE	VOLUMES	10	3	3,328	1,561	1,520	MASSA
TRIMESTRE	VOLÁTIL	4	2	4,065	1,333	2,079	GÁS

### Quadro 3. Colocações e sua classificação em termo de metáforas conceptuais

O quadro contém as seguintes informações: (1) n: o nóculo, isto é, as palavras-chave do estudo; (2) c: o colocado do nóculo; (3) fc: frequência do nóculo; (4) fnc: frequência conjunta de nóculo e colocado; (4) MI: valor da estatística de Mutual Information; (5) T: valor da estatística de T-Score; (6) Distância: valor da distância semântica segundo calculado pela fórmula de Leacock-Chodorow; (7) Campo fonte: Classificação da colocação em termos de um campo fonte, indicativo de uma metáfora conceptual.

No quadro, aparecem as colocações que passaram nos testes de associação (MI ou T) e no ponto de corte do cálculo de distância semântica. Relembrando, o ponto de corte estabelecido para MI foi 3 e para T foi 2; valores iguais ou acima desses, em qualquer uma das duas estatísticas, sinalizava uma colocação. No caso do teste de distância semântica, Berber Sardinha (2002) sugere 2 como sendo o valor máximo desejado para estudos de metáfora, com o teste Leacock-Chodorow. Pares com valores iguais ou acima de 2,1 foram sinalizados como não metafóricos. Com isso, entre as colocações identificadas com MI/T, apenas duas não passaram no teste de distância semântica, que são: Resultado + influenciado (2 ocorrências) e Trimestre + Influenciaram (3 ocorrências). Em ambos os casos, a distância semântica foi de 2,773, portanto acima do ponto de corte escolhido.

Em seguida, foi feita a classificação das expressões metafóricas em metáforas conceptuais, tendo em vista os domínios sinalizados pelas colocações. Um dos problemas presentes na categorização de expressões metafóricas em metáforas conceptuais é a delimitação do campo fonte. Há várias questões relacionadas a esse problema: Como o pesquisador pode garantir que um

campo fonte escolhido para designar uma metáfora conceptual é o correto? Se há mais de uma opção, qual a mais acertada? Se não há nenhuma opção, de que mecanismos ele dispõe para intuir um campo fonte? Há fontes confiáveis que possam substituir a intuição?

Surgem divergências entre classificações feitas por autores diferentes. Por exemplo, a metáfora MARKET MOVEMENTS ARE NAUTICAL OR ARE WAYS OF MOVING IN THE WATER, proposta por (Charteris-Black & Ennis, 2001), é contestada por (Chung, Ahrens, & Sung, 2003:130), que sugerem duas outras metáforas: STOCK MARKET IS A BOAT e STOCK MARKET IS OCEAN WATER.

A classificação das metáforas conceptuais foi feita com base em dois critérios básicos: (1) existência de referência prévia na literatura a respeito da mesma metáfora ou de metáfora semelhante; (2) caso esse passo não fornecesse nenhuma informação a respeito da classificação da colocação em termos de uma metáfora conceptual, foi empregada a intuição do pesquisador e seu conhecimento lingüístico, de mundo e do universo pesquisado (economia, finanças, bancos e investimentos), conseguido como resultado de muitos anos como pesquisador do projeto DIRECT, voltado à linguagem dos negócios.

O conjunto de metáforas conceptuais identificadas aparece registrado no quadro a seguir, em ordem decrescente pela coluna Ressonância.

	Metáfora conceptual	Types	%	Tokens	%	Ressonância	%	Palavras-chave
1	NEGÓCIOS SÃO UM ORGANISMO	31	39.7%	263	58.1%	8153	86.3%	9
2	PERSONIFICAÇÃO	14	17.9%	65	14.3%	910	9.6%	5
3	PARA BAIXO É RUIM	3	3.8%	30	6.6%	90	1.0%	2
4	NEGÓCIOS SÃO UMA COMPETIÇÃO	5	6.4%	15	3.3%	75	0.8%	3
5	NEGÓCIOS SÃO UMA CONSTRUÇÃO	4	5.1%	13	2.9%	52	0.6%	4
6	PARA CIMA É BOM	4	5.1%	10	2.2%	40	0.4%	3
7	NEGÓCIOS SÃO UMA PONTE	3	3.8%	11	2.4%	33	0.3%	3
8	NEGÓCIOS SÃO GUERRA	2	2.6%	13	2.9%	26	0.3%	1
9	NEGÓCIOS SÃO POLÍTICA	2	2.6%	8	1.8%	16	0.2%	2
10	NEGÓCIOS SÃO UM INSTRUMENTO ÓTICO	3	3.8%	8	1.8%	24	0.3%	2
11	VARIAÇÕES DOS NEGÓCIOS SÃO VARIAÇÕES DO TEMPO	2	2.6%	6	1.3%	12	0.1%	2

12	NEGÓCIOS SÃO UMA MASSA	2	2.6%	5	1.1%	10	0.1%	2
13	PARA BAIXO É BOM	1	1.3%	2	0.4%	2	0.0%	1
14	COMPORTEAMENTO DOS NEGÓCIOS SÃO COMPORTEAMENTO DE UM GÁS	1	1.3%	2	0.4%	2	0.0%	1
15	NEGÓCIOS SÃO UMA MÁQUINA	1	1.3%	2	0.4%	2	0.0%	1
	Totais	78	100.00%	453	100.00%	9447	100.00%	41

#### Quadro 4. Metáforas conceptuais, suas colocações e ressonância

O quadro traz as seguintes informações: (1) metáfora conceptual: a metáfora subjacente às colocações; (2) colocações (types): número de colocações individuais que expressam aquela metáfora conceptual – a porcentagem incide sobre o valor total de 78; (3) ocorrências (tokens): número de ocorrências a que correspondem todas as colocações referentes àquela metáfora conceptual – a porcentagem incide sobre o valor total de 453; (4) Ressonância: medida de produtividade de uma metáfora conceptual (Charteris-Black, 2004:91), obtido por meio da multiplicação do número de colocações (types) pelo de ocorrências (tokens). A ressonância exprime uma medida de abrangência da metáfora conceptual; quanto maior seu valor, mais presente estará a metáfora no texto – a porcentagem incide sobre o valor total de 9447. A tabela está ordenada em ordem decrescente, segundo o valor de ressonância; (5) palavras-chave: número de palavras chave contidas entre as colocações daquela metáfora conceptual.

Os resultados indicam que os negócios são conceitualizados de muitas formas diferentes. Isso reflete, por um lado, as concepções de administração de empresa correntes na literatura. Conforme mostra Morgan {, 1996 #7274}, as empresas são comumente conceitualizadas no nível da administração como máquina, organismo e cérebro, por exemplo. Daí surgem metáforas sobre negócios como negócios são um ORGANISMO ou NEGÓCIOS SÃO GUERRA, encontradas nos textos do corpus.

Por outro lado, tais metáforas indicam modos comuns de conceitualizar muitos campos da atividade humana e não somente os negócios. Por exemplo, como visto acima, discussão é frequentemente conceitualizada como guerra e muitos elementos são personificados, corriqueiramente. Assim, tais metáforas das teleconferências parecem derivar da produtividade de

certos campos fonte, como guerras e organismos, que se aplicam a muitos campos alvo diferentes.

De qualquer modo, o fato de certos campos fonte serem mais típicos têm conseqüências para a caracterização do gênero em questão. No caso do campo alvo de organismos, uma delas é que negócios e investimentos passam a adquirir a capacidade de crescer por conta própria, devido a leis naturais da biologia. Assim, por extensão, “o banco cresce sozinho”; não é a atuação humana que faz o banco crescer (Charteris-Black, 2004).

## 7. COMENTÁRIOS FINAIS

O trabalho aqui apresentado visou à apresentação de uma metodologia de cunho corpus-driven/bottom-up de identificação de metáforas. A análise não partiu de inventários prévios de metáforas mas sim das próprias palavras do texto. Os exemplos foram obtidos a partir da análise de colocações de palavras-chave. A metodologia propiciou um conjunto rico de candidatos a metáfora, que foi então analisado qualitativamente, com apoio em concordâncias geradas por computador. Com isso, chegou-se a um conjunto variado de metáforas presentes nas teleconferências.

Em geral, os resultados indicaram que muitas metáforas das teleconferências são típicas do campo da administração de empresas e da economia, como a do organismo, da competição (ambas em Clancy, 1989), a da guerra e a do gás (ambas em Charteris-Black, 2004), o que revela a afinidade desse gênero com outros da mesma área. Além disso, os campos fonte mais frequentes são gerais, aparecendo em muitos outros contextos.

Esperamos que a metodologia e os achados apresentados aqui sejam de valia para futuros estudos de metáfora com corpora, seja na área de negócios ou em outras.

Recebido em: setembro de 2007

Versão reformulada, recebida e aprovada em: setembro de 2010

E-mail: [tony@corpuslg.org](mailto:tony@corpuslg.org)

## REFERÊNCIAS BIBLIOGRÁFICAS

- BERBER SARDINHA, T. (2002). Metaphor in early applied linguistics writing: A corpus-based analysis of lexis in dissertations, *I Conference on Metaphor in Language and Thought*. Catholic University of São Paulo, Brazil.
- \_\_\_\_\_. (2003). The Bank of Portuguese, *DIRECT Papers* (Vol. 50). São Paulo / Liverpool: LAEL, PUCSP / AELSU, University of Liverpool.
- \_\_\_\_\_. (2004). *Linguística de Corpus*. São Paulo: Manole.
- BIBER, D., CONRAD, S., & REPPEN, R. (1998). *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- CAMERON, L. (1999). Identifying and describing metaphor in spoken discourse data. In L. Cameron & G. Low (Eds.), *Researching and Applying Metaphor* (pp. 105-132). Cambridge: Cambridge University Press.
- CHARTERIS-BLACK, J. (2004). *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan.
- CHARTERIS-BLACK, J., & ENNIS, T. (2001). A comparative study of metaphor in Spanish and English financial reporting. *The ESP Journal*, 20, 249-266.
- CHUNG, S. F., AHRENS, K., & SUNG, Y. H. (2003). Stock markets as ocean water: A Corpus-based, comparative study of Mandarin Chinese, English and Spanish. Proceedings of the 17 th Pacific Asia Conference on Language, Information and Computational (PACLIC). 2003. Singapore. pp. 124-133.
- CLANCY, J. J. (1989). *The Invisible Powers - The Language of Business*. Lexington, Mass: Lexington Books.
- DEIGNAN, A. (1999). Corpus-based research into metaphor. In L. Cameron & G. Low (Eds.), *Researching and Applying Metaphor* (pp. 177-199). Cambridge: Cambridge University Press.
- \_\_\_\_\_. (2005). *Metaphor and Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- HUNSTON, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- KENNEDY, G. (1998). *An Introduction to Corpus Linguistics*. New York: Longman.
- KÖVECSES, Z. (2002). *Metaphor - A Practical Introduction*. New York: Oxford University Press.
- LAKOFF, G., & JOHNSON, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

- LAKOFF, G., & JOHNSON, M. L. (2002). *As Metáforas da Vida Cotidiana*. Campinas: Mercado de Letras / EDUC (Tradução de Mara Sofia Zanotto e equipe).
- LEACOCK, C., & CHODOROW, M. (1998). Combining local context and WordNet similarity for word sense identification. In C. Fellbaum (Ed.), *WordNet: An Electronic Lexical Database* (pp. 265-294). Cambridge, Mass: MIT Press.
- MILNE, M., KEARINS, K., & WALTON, S. (2003). Business makes a “journey” out of “sustainability”: Creating adventures in wonderland? Accountancy and Business Law Working Paper Series. School of Business, University of Otago, New Zealand.
- MOON, J., CRANE, A., & MATTEN, D. (2003). *Can corporations be citizens? Corporate citizenship as a metaphor for business participation in society. (2nd edition)*. Nottingham: International Centre for Corporate Social Responsibility, Nottingham University.
- MORGAN, G. (1996). *Images of Organization*. Thousand Oaks, CA: Sage.
- NEUMAN, W. R. (1976). Patterns of recall among television news viewers. *Public Opinion Quarterly*, 40, 115-123.
- PEDERSEN, T., & PATWARDHAN, S. (2002). distance Perl package (Version 0.1): University of Minnesota, Duluth.
- SCOTT, M. (1997). WordSmith Tools. Version 3. Oxford: Oxford University Press.
- SINCLAIR, J. M. (2003). *Reading Concordances*. London: Pearson Longman.
- STEEN, G. (1999). Metaphor in discourse: Towards a linguistic checklist for metaphor analysis. In L. Cameron & G. Low (Eds.), *Researching and Applying Metaphor* (pp. 81-104). Cambridge: Cambridge University Press.