

# Metaphor and Corpus Linguistics<sup>1</sup>

## *Metáfora e linguística de corpus*

---

Tony Berber Sardinha\*  
Catholic University of São Paulo  
São Paulo / Brasil

**ABSTRACT:** In this paper, I look at four different aspects of metaphor research from a corpus linguistic perspective, namely: (1) the lexicogrammar of metaphors, which refers to the patterning of linguistic metaphor revealed by corpus analysis; (2) metaphor probabilities, which is a facet of metaphor that emerges from frequency-based studies of metaphor; (3) dimensions of metaphor variation, or the search for systematic parameters of variation in metaphor use across different registers; and (4) automated metaphor retrieval, which relates to the development of software to help identify metaphors in corpora. I argue that these four aspects are interrelated, and that advances in one of them can drive changes in the others.

**KEYWORDS:** corpora; metaphor; metaphor identification; lexicogrammar; probabilities; Multi-Dimensional Analysis; metaphor retrieval software.

**RESUMO:** Neste artigo discuto quatro aspectos da pesquisa sobre metáfora do ponto de vista da linguística de corpus: (1) a lexicogramática das metáforas, que se refere aos padrões da metáfora linguística revelados pela análise de corpus; (2) probabilidades metafóricas, que é uma faceta da metáfora que emerge a partir dos estudos relacionados à frequência de metáforas; (3) dimensões da variação de metáforas, ou a busca por parâmetros sistemáticos de variação de uso de metáfora em diferentes gêneros; e (4) captura automática de metáfora, que está relacionada ao desenvolvimento de softwares que auxiliam na identificação de metáforas em corpora. I defendo que esses quatro aspectos são interrelacionados, e que progressos em um deles podem acarretar mudanças nos outros.

**PALAVRAS-CHAVE:** corpora; metáfora; identificação de metáfora; lexicogramática; probabilidades; Análise Multidimensional; software de captura de metáforas.

---

\* [tony@corpuslg.org](mailto:tony@corpuslg.org)

<sup>1</sup> I want to thank CPNq (Brasília, Brazil) for supporting my research.

## 1. Introduction

The field of metaphor studies is vast and has, a long tradition that dates back to ancient Greece. Over time, numerous theories of metaphor and a range of different methods for metaphor identification have been proposed (see GIBBS, 2008a). Corpus Linguistics is a newcomer to the field, but its influence is already being felt:

A related emerging concern for empirical studies of metaphor focuses on the true frequency of metaphors in language and other media. Claims about the importance or ubiquity of particular metaphorical patterns, in either language or thought, are often made without adequate empirical support, such as reporting the frequencies with which different metaphors are found in particular texts, or comparing the findings from one's own textual analysis of metaphor with those seen in large corpora. (GIBBS, 2008b, p. 12)

Notwithstanding the recognition of its role, Corpus Linguistics has only begun to make itself noticed in the vast field of metaphor scholarship. One reason is the fact that it is a relatively recent approach to metaphor analysis, with the first studies dating back to 1999 (DEIGNAN, 1999a; 1999b; 1999c). Another reason is that metaphor traditionally requires hand analysis, which is too time consuming to carry out in large corpora. A number of metaphor retrieval computer tools have been developed but they have not made an impact in the field, partly because they are not widely available and partly because their performance is still not particularly high (see section 5 below).

There is a growing body of research at the interface between metaphor and Corpus Linguistics. Deignan (2005) offers a detailed treatment of bottom-up approaches to metaphor analysis, with an emphasis on concordancing and how linguistic metaphor is signaled by recurring patterns of use. In Stefanowitsch and Gries (2006) several different approaches to metaphor identification in corpora are presented, which Stefanowitsch (2006, p. 2-6) classifies into seven distinct groups based on the kind of searching performed: manual, source domain vocabulary, target domain vocabulary, both source and target domains, metaphor markers, and extraction from corpora annotated for semantic fields or for conceptual mappings. Wikberg (2007) discusses central issues in using corpora for metaphor research, and concludes that close reading of text passages is necessary for determining metaphoricity. Berber Sardinha (2009) provides an overview of corpus-based and corpus-driven approaches to metaphor identification in corpora, showing how they can be retrieved by programs such

as WordSmith Tools Keywords (SCOTT, 2004) and the Metaphor Candidate Identifier (see section 5 below).

At the same time, there is crucial work being done at setting criteria for metaphor identification by hand analysis. This includes MIP (Metaphor Identification Procedure) and MIV (Metaphor Identification Through Vehicle Terms). MIP (PRAGGLEJAZ GROUP, 2007) and its more recent version MIPVU (STEEN, DORST, HERRMANN *et al.*, 2010) both lay down guidelines for metaphor identification. MIP/MIPVU details steps for coding metaphors at the word level, showing how to determine metaphoricity by taking into account the basic and contextual meanings of each word. MIV (CREET, 2006) also presents detailed procedures for metaphor identification, but singles out Vehicle terms (metaphorically used language), which may or may not be single words. Other equally important work in this area includes Steen (2007) and Cameron and Maslen (2010). The former gives a thorough account of issues in metaphor identification and interpretation, as well as how these relate to language and thought. The latter focuses on discourse dynamics and takes a comprehensive look at systematicity, that is, how recurrent connections between Topic (what the metaphor refers to) and Vehicle can reveal aspects of discourse. Work on methods for metaphor identification and interpretation, even though not strictly from a corpus perspective, can provide valuable insights into issues that affect corpus research, such as the lexical patterning of metaphorical language, and criteria for determining metaphor use.

I would argue the following are particularly important findings from previous research:

- 1) Metaphor use seems to correlate with lexicogrammatical patterns. Patterns used to express metaphor are typically different from patterns employed to denote literal language. In other words, metaphorically used language selects particular patterns.
- 2) In particular genres (articles, reports, speeches, etc.) or registers (academic, fiction, business, etc.)<sup>2</sup> (see e.g. BIBER, CONRAD, 2009), metaphorically

---

<sup>2</sup> There are numerous definitions of genre and register in the literature. Here, genres are understood as 'recognizable communicative events, characterized by a set of communicative purposes identified and mutually understood by members [...] of [a] community where they regularly occurs.' (BHATIA, 2004, p. 23). And register is defined 'as a cover term for any language variety defined by its situational characteristics, including the speaker's purpose, the relationship between speaker and hearer, and the production circumstances.' (BIBER, 2009, p. 823)

used language has probabilities of use that are different from those in literal language. Also, probabilities of metaphor use for particular words or expressions in specialized varieties differ from those in general language.

- 3) Metaphor use varies systematically across different genres and registers and this may give rise to dimensions of metaphor variation.
- 4) Specialized systems for metaphor retrieval by machine have been developed to automate metaphor identification from corpora.

Based on these, the major areas that I think should mature in CL metaphor research are the following:

- 1) The lexicogrammar of metaphor;
- 2) The probabilistic nature of metaphor use;
- 3) Variation in metaphor use;
- 4) Automating metaphor identification.

In the following sections, I will focus on each of these points in turn. In order to make these points, I will report findings from previous research.

However, we must first distinguish between two basic types of CL metaphor research: whole corpus and concordance-based. In the former, researchers code all the metaphors in the whole corpus, usually by hand, and then retrieve the metaphors based on the hand analysis done ahead of time; in the latter, they run concordances for particular items and then analyze only those occurrences. Whole corpus analyses are affected by the amount of data that need to be coded. Concordance-based is not, because analysis is typically carried out on a sample (e.g. one thousand lines) of concordance lines extracted from the corpus. Concordance-based analyses are influenced by the choice of search terms, since these will define what will and will not be found in the corpus.

These points overlap and draw on each other; for instance, the more we know about the linguistic patterning underlying metaphor use, the better we can establish both the probabilities of use and the dimensions of variation of metaphor across registers, and vice-versa. And the more we know about the patterning of metaphor, its probabilities and variation, the better positioned we are to determine which aspects of metaphor the computer can be taught to recognize with reasonable degrees of accuracy.

One further point that has not deserved much attention in CL metaphor research is extending the scope of inquiry beyond English. The vast

majority of the literature focuses on metaphors in English, and few other languages are reported at all. There are exceptions, notably Steen *et al.* (2010) analysis of 130,000 words of Dutch. A basic ingredient, the corpus, can be easily compiled for a large number of languages, given the wide availability of electronic texts on the Web. Other resources may be harder to find, which may hinder progress of this kind of research in other languages. I will present findings of analyses of Portuguese corpora below.

## 2. The lexicogrammar of metaphor

One of the ways in which a metaphor reveals itself in corpora is by its patterns of usage, which typically contrasts with the patterns of non-metaphorical language. This has proved valuable as a criterion for both metaphor manifestation and identification.

To illustrate, I will use data from my own analysis of autobiographical narratives in Brazilian Portuguese (BERBER SARDINHA, 2007B). These were recorded by the Museu da Pessoa (Museum of the Person), an organization that aims at preserving history by recording people telling a personal narrative about their lives. These recordings are then transcribed and many of them are made public on the institution's website. I collected a corpus of such narratives and used both hand and machine analyses to identify the metaphors in them.

One set of metaphorical patterns that emerged in the analysis was the following:

TABLE 1  
Metaphorical patterns of PEGAR

Pattern	Direct English translation
PEGAR <sub>past tense</sub> + , + <i>Vb past</i> <sup>3</sup>	GRAB/CATCH/PICK UP <sub>past tense</sub> + , + <i>Vb past</i>
PEGAR <sub>past tense</sub> + <i>Determiner</i> + [Disease]	GRAB/CATCH/PICK UP <sub>past tense</sub> + <i>Determiner</i> + [Disease]
PEGAR <sub>past tense</sub> + e + <i>Vb past</i>	GRAB/CATCH/PICK UP <sub>past tense</sub> + e + <i>Vb past</i>

<sup>3</sup> I use the following convention to represent patterns: CAPITALS = lemmas; subscript italics = grammatical constraint on lemma; *Italics* = part of speech; [Square brackets] = Semantic fields; other formats: actual strings/words; the plus sign = followed by, up to five words away.

These are exemplified in the concordance below.

1	,	o	pai	pegou	,	disse	que
2	nesse	trabalho	eu	pegava	,	eu	fiz
3	taquara	e	eu	peguei	,	fiz	por
4	é	que	me	pegou	a	depressão	estou
5	E	ai	ela	pegou	a	malária	.
6	e	a	mãe	pegou	e	disse	:
7	frente	dele	,	pegou	e	disse	:
8	que	a	gente	pegava	e	fazia	,
9	com	dó	de	pegar	e	levar	pra
10	,	ai	ela	pegou	e	falou	:
11	.	Aí	ele	pegou	e	danou	pra
12	Aí	o	velho	pegou	e	veio	mais
13	.	Aí	ele	pegou	e	casou	com
14	pitimbado	.	Ele	pegou	e	falou	:
15	.	Aí	ela	pegou	e	falou	:
16	,	ai	eu	peguei	e	falei	com
17	no	pé	,	pegou	e	falou	comigo
18	tá	bom	,	peguei	e	fui	lá
19	.	Aí	eu	peguei	e	disse	:
20	fiz	?	Eu	peguei	e	fui	na
21	?	Então	ela	pegava	e	mandava	eu
22	.	Aí	eu	peguei	e	falei	.
23	Então	,	eu	peguei	e	comprei	esse
24	“	Aí	eu	pegava	e	fazia	a
25	de	carro	,	peguei	e	vendi	o
26	!	Aí	eu	peguei	e	falei	para
27	,	ai	eu	peguei	e	liguei	para
28	Maria	do	Carmo	pegou	e	falou	“
29	Cultura	,	então	peguei	e	pedi	a
30	um	dia	ele	pegou	e	começou	falando
31	tempo	que	ele	pegou	essa	doença	,
32	.	Lá	ela	pegou	essa	febre	.
33	panqueca	.	Denise	pegou	uma	anemia	profunda
34	ficou	doente	,	pegou	uma	doença	grave
35	ele	pegou	,	pegou	uma	hepatite	o
36	doente	.	Ela	pegou	uma	febre	que
37	logo	.	Ele	pegou	uma	infecção	intestinal

FIGURE 1: Concordances for metaphorical patterns of PEGAR

The metaphors realized by these patterns appear in the table below.

TABLE 2  
Metaphorical patterns of PEGAR

Pattern	Metaphor
PEGAR <sub>past tense</sub> + , + Vb past	AN ACTION IS AN OBJECT
PEGAR <sub>past tense</sub> + Determiner + [Disease]	A DISEASE IS AN OBJECT
PEGAR <sub>past tense</sub> + e + Vb past	AN ACTION IS AN OBJECT

Examples of each pattern are shown below:

TABLE 3  
Examples of metaphorical patterns of PEGAR

Pattern and Metaphor	Concordance lines in Figure 1	Example
PEGAR <sub>past tense</sub> + , + Vb past AN ACTION IS AN OBJECT	1 – 3	O pai <u>pegou</u> , <u>disse</u> que uma coisa que a gente tinha que ter era estudo, entende? My father turned and said that one thing we needed to have was schooling, see what I mean?
PEGAR <sub>past tense</sub> + Det + [Disease] A DISEASE IS AN OBJECT	4 – 5 31 – 32 33 – 37	Ela <u>pegou a malária</u> . E foi indo, foi a causa da morte dela. She caught malaria. She carried on, this was the cause of her death. E ela estava em Marajó. Lá ela <u>pegou essa febre</u> . She was in Marajó. She caught this fever over there. <u>Pegou uma hepatite</u> , o cabelo foi tudo para o beleléu, tudo num mês. <i>He caught hepatitis, his hair fell out, all in a month's time.</i>
PEGAR <sub>past tense</sub> + e + Vb past AN ACTION IS AN OBJECT	6 – 30	Subindo um viaduto, assim, e a mãe <u>pegou e disse</u> : “Cardoso, caiu uma, caiu um papelão.” <i>Going up an overpass, like that, and my mom turned around and said: “Cardoso, a piece of cardboard dropped out”.</i>

As can be seen in the translated examples, in the metaphor AN ACTION IS AN OBJECT PEGAR means something equivalent to the English ‘turn (a)round and’. In the other metaphor, A DISEASE IS AN OBJECT, it means its direct equivalent, ‘to catch’.

I labeled the instances of ‘turn (a)round and’ as AN ACTION IS AN OBJECT, but these might as well have been named in other ways, for instance as AN IDEA IS AN OBJECT, since they might imply ‘grabbing an idea’ and expressing it in words or actions. Labeling metaphors, especially conceptual ones, is tricky, and there are no specific guidelines. This is certainly an area where more clarity is needed; this will become more pressing as research that resorts to metaphor categorization intensifies.

Semantic categories are very useful in formulating patterns. In this particular case, they were applied after the fact, by looking at and grouping citations in a concordance. They can be more useful, though, if applied as search terms to query a corpus, because researchers need only to specify the semantic grouping and not each individual word in it. The problem is of course that it requires a semantically annotated corpus. Increasing the availability of semantically annotated corpora (in several languages) is another front that needs development both in Corpus Linguistics in general and in CL metaphor research in particular.

By contrast, the basic patterns for literal uses of PEGAR are:

TABLE 4  
Non-metaphorical patterns of PEGAR

Pattern Example	Concordance lines in Figure 2	Direct English (translation) Examples
PEGAR + <i>Det</i> + [Concrete]		GRAB + <i>Det</i> + [Concrete]
Pegou a bagagem	1	Picked up the baggage
Pegou a bola	7	Grab that ball
Pegava o saco	20	Picked up a bag
Pegou uns pincéis	30	Picked up a few brushes
PEGAR + <i>Prep</i> + <i>Det</i> + [Concrete]		SIT/HOLD/PICK UP + <i>Prep</i> + <i>Det</i> + [Concrete]
Me pegava no colo	12	Sit on someone’s (lap)
Pegar no lixo	13	Pick something up from the trash

These are illustrated in the following concordance:

1	?	Conclusão	?	<b>Pegou</b>	a	bagagem	e
2	levantava	cedo	e	<b>pegava</b>	a	enchadinha	.
3	falava	assim	,	<b>pegava</b>	a	identidade	da
4	menina	saiu	,	<b>pegou</b>	a	lata	.
5	,	a	gente	<b>pegando</b>	a	meia	com
6	lá	fora	,	<b>peguei</b>	a	receita	joguei
7	que	tinha	que	<b>pegar</b>	aquela	bola	.
8	eu	gostava	de	<b>pegar</b>	aquela	coisinha	de
9	aqui	!	Vou	<b>pegar</b>	as	cobertas	“
10	segunda	época	,	<b>pegou</b>	as	provas	e
11	no	coisa	e	<b>pegava</b>	as	terras	,
12	,	que	me	<b>pegava</b>	no	colo	e
13	de	pegar	.	<b>Pegar</b>	no	lixo	.
14	bisavó	,	aí	<b>pegou</b>	no	mato	,
15	ele	atravessava	,	<b>pegava</b>	o	bonde	,
16	a	mais	e	<b>pegava</b>	o	dinheiro	pra
17	pagar	pra	ele	<b>pegar</b>	o	diploma	.
18	que	a	gente	<b>pegava</b>	o	gibi	,
19	ele	foi	lá	<b>pegar</b>	o	prêmio	,
20	,	né	?	<b>Pegava</b>	o	saco	,
21	Aí	um	dia	<b>peguei</b>	um	anúncio	e
22	notável	,	ele	<b>pegava</b>	um	livro	.
23	em	quando	ele	<b>pegava</b>	um	passarinho	esse
24	,	pra	ela	<b>pegar</b>	um	pedaço	a
25	,	mas	nunca	<b>peguei</b>	um	peixe	o
26	favor	do	cara	<b>pegar</b>	um	pente	para
27	às	vezes	vocês	<b>pegavam</b>	uma	quantidade	de
28	empregad	não	podia	<b>pegar</b>	uma	revista	,
29	Eu	gosto	,	<b>pegar</b>	uma	tesoura	.
30	de	cor	,	<b>pegou</b>	uns	pincéis	,
31	teve	interesse	em	<b>pegar</b>	uns	quadros	meus

FIGURE 2: Concordance with non-metaphorical uses of PEGAR (GRAB)

This illustrated the existence of a lexicogrammar of individual metaphors, which patterns the way metaphor choices are made in texts. Patterns may signal metaphor (or non-metaphor) with a certain likelihood. The next point looks at the cumulative effects of the presence of metaphor in corpora, from a probabilistic point of view.

### 3. Metaphor probabilities

Patterns of metaphor use occur in language with particular probabilities of occurrence attached to them. There is little research in this aspect of metaphor use, even though this is an important characteristic of metaphor, because it may reveal how likely it is that we encounter metaphors in written and spoken texts. Theory emphasizes that a metaphor is a frequent linguistic feature, and that all language users are likely to come across or employ metaphors to express various meanings. Empirical research also makes similar claims. For instance, according to Deignan and Potter (2004, p. 1236) 'non-literal language is extremely common, often accounting for a substantial proportion of the corpus citations of a word.' Gibbs and Franks (2002, p. 151) likewise note that their data 'show just how prominent metaphor was.' And Moules *et al.* (2004) observe that they were 'struck with how often metaphors arise in the language of grief'. Such claims imply that the probability of metaphor use is high in language, and so in order to verify whether they are true, we must look at the probability of use of metaphor in corpora.

I did research by looking at metaphor probability in 2007, and this involved determining the metaphor status (metaphor versus non-metaphor) of each individual word in an 85,000 word corpus of teleconferences held at investment banks in Portuguese in Brazil; these meetings were attended by bank staff, investors and journalists, and were broadcast over the phone. I then searched a large general corpus of Portuguese (Banco de Português, +220 million words) for the same words found to be used metaphorically in the teleconference corpus. Finally, I compared the frequency of metaphor versus non-metaphor across the two corpora.

In that study, probabilities were calculated in three different ways.

First, all metaphorically used words (MUW) tokens as a proportion of all word tokens in the specialized corpus. This can answer the question of how likely it is that any one word token is a MUW:

$$4311 \text{ MUW tokens} / 85438 \text{ word tokens in the corpus} = .05 \text{ (5\%)}$$

This indicates that a small share of the words in the corpus are MUWs. The likelihood of word tokens being an MUW is therefore approximately 1 in 20. Literal is the default status for words in the corpus.

Second, all MUW tokens as a proportion of their joint frequency (including both metaphors and non-metaphors) in the specialized corpus. This

can provide an answer to the question of how likely it is that an MUW selects a metaphorical meaning:

4311 MUW tokens / 5021 sum of frequency of all MUW types = .86 (86%)

This suggests that MUW types tend to be re-used metaphorically in the same corpus. That is, of all the words in the corpus, those that had taken on a metaphorical meaning tend to so more often than otherwise (that is, be used literally). Metaphor is the default status for MUWs.

Third, the frequency in the reference corpus of all MUW types found in the specialized corpus as a proportion of their joint frequency (including both metaphors and non-metaphors) in the reference corpus. This can help answer how likely it is that MUWs in the specialized corpus are metaphors in language in general:

15220 MUW tokens / 21854 sum of frequency of all MUW types = .7 (70%)

This shows that MUWs in the specialized corpus tend to be MUWs in general language as well, albeit to a lesser degree.

However, when I looked at each word individually and compared their probability of metaphor use in the specialized corpus against the general corpus, I noticed that the vast majority showed ‘upward resetting’ (HALLIDAY, 1991), that is, their probability of metaphor use was higher in the specialized corpus:

TABLE 5  
Resetting of probability of metaphor use

Upward resetting (general corpus < specialized corpus)	323
No resetting (general corpus = specialized corpus)	63
Downward resetting (general corpus > specialized corpus)	37
Total	423

Examples of upward resetting MUWs are shown in Table 6.

TABLE 6  
Examples of upward resetting

Word (literal translation)	MUW Prob. in General Corpus	MUW Prob. in Specialized Corpus	Example (adapted translation)
Aliança ( <i>alliance</i> )	.01	1.00	<i>o (banco) firmou uma <u>aliança</u> estratégica com (companhia x)</i> (the bank formed a strategic <u>alliance</u> with (company x))
Parada ( <i>stopped</i> )	.03	1.00	<i>a economia <u>parada</u></i> ( <u>slow</u> economy)
Bola ( <i>ball</i> )	.04	1.00	<i>acho que precisa de uma <u>bola</u> de cristal para saber o que vai acontecer</i> (I think we need a crystal ball to know what's going to happen)
Jogamos ( <i>throw</i> )	.05	1.00	<i>é isso que nós <u>jogamos</u> na projeção</i> (that's what we <u>build into</u> our projection)
Atingidos ( <i>hit</i> )	.05	1.00	<i>bases essenciais para que estes objetivos sejam <u>atingidos</u></i> (essential basis for us to <u>meet</u> our objectives)
Fotografia ( <i>snapshot</i> )	.05	1.00	<i>ai tem uma <u>fotografia</u> do que é a transação</i> (there's a <u>snapshot</u> of what a transaction is)
Depositado ( <i>deposited</i> )	.05	1.00	<i>que mostra a confiança que o mercado tem <u>depositado</u> no (banco)</i> (that shows the trust that the market has <u>placed</u> in the bank)
Bala ( <i>bullet</i> )	.06	1.00	<i>eu entenderia que eles estariam guardando <u>bala</u></i> (I would understand that they were holding <u>fire</u> )
Loteria ( <i>lottery</i> )	.07	1.00	<i>mas isso é <u>loteria</u>, não temos idéia do que vai ocorrer</i> (but this is a <u>lottery</u> , we have no idea what's going to happen)
Travada ( <i>locked</i> )	.07	1.00	<i>nós temos a moeda dólar <u>travada</u> para a aquisição</i> (we have the dollar <u>locked</u> in for the acquisition)
Empatar ( <i>tie</i> )	.08	1.00	<i>como isso poderia <u>empatar</u> o balanço em reais</i> (how that might <u>balance</u> the books in Brazilian reais)
Canal ( <i>channel</i> )	.09	1.00	<i>entre outras estratégias que o banco pode adotar, um <u>canal</u> alternativo</i> (among other strategies that the bank may adopt, an alternative <u>channel</u> )
Chute ( <i>kick</i> )	.10	1.00	<i>isso não foi um <u>chute</u>, foi sim uma análise metodológica</i> (that was not a <u>guess</u> , it was a methodological analysis)

Characteristically, these are words of the financial domain. Their metaphoricality is strengthened in the specialized corpus.

Taken together, these findings seem to suggest that metaphors are not evenly distributed across texts; rather, they are typical of certain words/patterns and not others. On the basis of this evidence, metaphors might be seen as a matter of degree (more/less probable) rather than of category (yes/no). In addition, certain metaphors seem to be typical of particular genres or registers rather than of 'language as a whole'. The next section will explore the consequences of that from the point of view of variation.

#### **4. Dimensions of metaphor variation**

In the previous section, I presented evidence to suggest that the frequency of use of metaphor varies between specialized and general language. The question that arises is whether there is variation across different genres and registers as well. If the answer is affirmative, then this may suggest that metaphor use is patterned at the level of both lexicogrammar and register.

One way in which language use at the level of register may be seen to be systematically patterned is through dimensions of variation. This concept was introduced by Biber (1985; 1988) to refer to underlying parameters of variation, where 'each dimension represents a different set of co-occurring linguistic features' (BIBER, 2009, p. 829). He has developed a method for identifying these dimensions which was termed Multi-Dimensional Analysis of Variation (MDA), which can be defined as:

a corpus-based methodological approach to, (i) identify the salient linguistic co-occurrence patterns in a language, in empirical/quantitative terms, and (ii) compare registers in the linguistic space defined by those co-occurrence patterns. (BIBER; DAVIES; JONES *et al.*, 2006, p. 5).

To carry out an MDA, the following steps need to be taken:

- (a) "An appropriate corpus is designed based on previous research and analysis. Texts are collected, transcribed (in the case of spoken texts), and input into the computer. (In many cases, pre-existing corpora can be used.)
- (b) Research is conducted to identify the linguistic features to be included in the analysis, together with functional associations of the linguistic features.
- (c) Computer programs are developed for automated grammatical analysis, to identify or 'tag' all relevant linguistic features in texts.

- (d) The entire corpus of texts is tagged automatically by computer, and all texts are edited interactively to ensure that the linguistic features are accurately identified.
- (e) Additional computer programs are developed and run to compute normed frequency counts of each linguistic feature in each text of the corpus.
- (f) The co-occurrence patterns among linguistic features are identified through a factor analysis of the frequency counts.
- (g) The 'factors' from the factor analysis are interpreted functionally as underlying dimensions of variation.
- (h) Dimension scores for each text are computed; the mean dimension scores for each register are then compared to analyze the salient linguistic similarities and differences among registers." (BIBER, 2009, p. 825-826).

MDA research has identified dimensions of variation for a number of different languages and varieties. The first MDA description is that of English, which consists of six dimensions, namely: (1) Involved vs. informational production; (2) Narrative vs. Non-narrative concerns; (3) Explicit vs. Situation-dependent reference; (4) Overt expression of persuasion; (5) Abstract vs. Non-abstract information; and (6) On-line informational elaboration.

There were no previous studies that focused explicitly on metaphor variation. Nor were there MDA studies that included variables relating to metaphor use. Nevertheless, there is mounting evidence that metaphor use varies across registers. For instance, Cameron's (2003) study of metaphor in classroom discourse found a rate of metaphor use of 1 out of every 37 words. My own study of conference calls (referred to in the previous section, BERBER SARDINHA, 2008) showed that metaphor was used at a rate of 1 out of every 20 words. My research into metaphor use in autobiographical narratives (BERBER SARDINHA, 2010b) indicated the rate of metaphor use to be at 1 out of every 115 words. And Krennmayr's (personal communication) study of several registers indicated that metaphor use varied from 18.4% of word tokens in academic discourse, to 16.6% in news, to 11.8% in fiction, to 7.8% in conversation. Different identification methods were used in these studies, as well as different definitions of what is counted as a metaphorical unit, therefore these figures are not directly comparable. This is confirmed by my own analyses of the same autobiographical narrative corpus; an early analysis showed a rate of 1 metaphor every 364 words, but more recently this changed to 1 in every 115, due to changes in the procedures for metaphor identification.

Despite these problems, this combined evidence may suggest that different registers use metaphors at different rates, and that perhaps casual spoken non-scripted registers such as conversation and personal narratives employ fewer metaphors than information-laden written registers such as academic or news.

To verify that, I decided to conduct an MDA of three major registers of Brazilian Portuguese, and include in the variable set a number of metaphor-related variables.

The corpus used for this study consisted of a small subset of the Brazilian MDA Corpus, which in turn is taken from the much larger Brazilian Corpus (1 billion words; <http://corpusbrasileiro.pucsp.br>):

TABLE 7  
MDA of metaphor variation corpus

Register	Tokens	Texts
Conversation	17,042	8
Academic	16,915	8
Newspaper	18,165	67
Total	52,122	83

The corpus was compiled to meet a target of around 50 thousand words, distributed roughly equally among its registers. The target was chosen because it did not seem too large for manual analysis. Previous studies that involved close reading of entire corpora have used less data, such as Cameron (2003), who took a corpus of 27,000 words of classroom talk, Cameron (2010), with a 27,000 word corpus of reconciliation discourse, and Charteris-Black (2004), whose corpus of American political speeches was 33,000 words long. There is no consensus on corpus size for such research projects, and other studies used larger data sets, such as Steen *et al.* (2010), which is based on a corpus of 190,000 of English data and 130,000 of Dutch.

Initially, the variable pool included 57 variables. The corpus was tagged for part of speech by the Tree-Tagger (trained for Portuguese), and for metaphor features by hand. After that, variable frequencies were taken and examined, and a number of low frequency variables were dropped. An initial factor analysis was run (in SPSS) and communalities were examined. Some variables were dropped based on their communalities, either because they were too low (<.4 according to STEVENS, 2002, p. 410) or too high (1 or higher). The result was a final set of 25 variables, shown below.

To code metaphors, I drew on the concepts of metaphor Topic and Vehicle. A metaphor Topic is that which is being referred to metaphorically. A Vehicle, in turn, is that which is used metaphorically. For instance, in the metaphor ‘waste of time’, ‘time’ is the Topic, and ‘waste’ the Vehicle. Time is being metaphorized in terms of a precious resource that should not be wasted.

#### Metaphor variables

Variable	Example <i>Translation</i>
1) metaphor density	Metaphorically used types / total word tokens in the text
2) metaphor <u>topic</u> : people	Fiz <u>amigos</u> <i>I made friends</i>
3) metaphor <u>topic</u> : social	Pressão de <u>diversos países</u> <i>Pressure by different countries</i>
4) metaphor <u>topic</u> : abstract	Queda da <u>participação</u> <i>Drop in participation</i>
5) metaphor <u>vehicle</u> : movement/position	<u>Elevadas</u> taxas <i>High taxes</i>
6) metaphor <u>vehicle</u> : object/buildings	Relacionamento <u>construtivo</u> <i>Constructive relationship</i>
7) metaphor <u>vehicle</u> : other	<u>Campo</u> de tensões <i>Field of tension</i>
8) vehicle word POS: verb	<u>Aumentar</u> o tempo <i>Increase the time</i>

#### Linguistic variables

- 1) adjectives
- 2) adverbs
- 3) demonstratives
- 4) future tense
- 5) nouns
- 6) past participles
- 7) past tense verbs
- 8) possessives
- 9) prepositions
- 10) pronouns: 1st person
- 11) pronouns: 2nd person

- 12) pronouns: 3rd person
- 13) proper nouns
- 14) public verbs
- 15) be as main verb (ser, estar)
- 16) subordinate clauses
- 17) verbs

In order to determine how many factors are present in the data, a graph known as ‘scree plot’ is normally used in MDA. It plots the eigenvalues, or variances of the factors. Researchers look at the line searching for points where it breaks, indicating major differences in factor variances. The scree plot for the initial factor solution seemed to indicate a three-factor solution, as shown in the figure below.

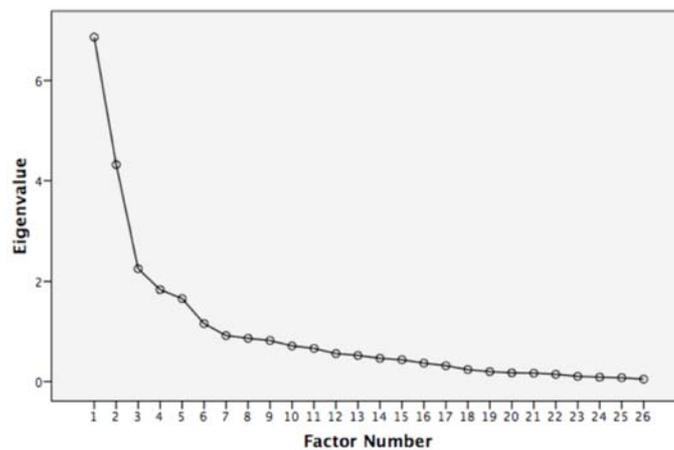


FIGURE 3 - Scree plot for metaphor MDA data

A three-factor Promax rotated analysis was then run on the data. The total variance captured was 47%, which is close to Biber’s (1988) final 6-factor solution (at 49%). This suggested that the factor analysis seemed to have tapped into a good portion of the variation present in the data. Factor intercorrelations were small, at  $-.088$  (between factors 1 and 2),  $-.405$  (between 1 and 3), and  $.29$  (2 and 3). This is again similar to Biber (1988), where they ranged from  $-.49$  to  $.3$ .

The structure of the first factor is shown below.

TABLE 8  
Factor 1 structure

Variable	Loadings
adverbs	.84
subordinate clause	.76
pron 3rd person	.75
tokens	.69
demonstratives	.65
past tense	.61
adjectives	.58
pron 1st person	.57
pron 2nd person	.53
possessives	.42
verbs (other)	.41
be as main verb (ser, estar)	.39
public verbs	.36
Metaphor density	-.58
proper nouns	-.56

This factor encompassed a large number of linguistic features and only one metaphor variable (density). Adverbs, subordination, be as main verb, first and second person pronouns are all features occurring on Biber's Dimension 1 (BIBER, 1988, p. 105-107), signaling involved production. Public verbs and past tense are present on Biber's 1988 Dimension 2, indicating narratives. Adjectives appear on his Dimensions 1, 2, and 5, associated with informational, non-narrative and abstract discourse. And demonstratives occur on his Dimension 6, linked to online informational elaboration. In all, these features seem to indicate verbal, narrative, involved discourse produced under real-time conditions. The proper nouns at the bottom of the factor suggest an informational focus.

The distribution of registers along this dimension is shown below.

TABLE 9  
Dimension 1 'Involved narrative production vs metaphor use'.

Mean factor scores for each register

DIMENSION 1

+	
22	
21	conversation
20	
19	
...	
0	
-1	academic
-2	newspaper
-3	
...	
-20	
-	

F=53.051, p=.000, R<sup>2</sup>=.617

The first factor is generally the one that captures most of the variation. This is reflected in the distance between conversation, at the top, and the other two registers at the bottom. The register with the highest score on this dimension was conversation, which means conversations have high quantities of the positive features (mostly verbal features, as indicated above), and low quantities of negative ones (proper nouns and metaphors). I labeled this dimension 'Involved narrative production versus metaphor use', because the positive features seem to highlight the involved nature of conversation, while at the same time revealing that involvement seems to be achieved with very little need for metaphors. Proper nouns are missing in conversation because they are generally replaced by pronouns. This appears to confirm the earlier hunch that in casual spoken registers, metaphor is not a frequent feature.

The structure of the second factor is shown below.

TABLE 10  
Factor 2 structure

FACTOR 2	
Variable	Loadings
verbs as vehicles words	.92
social topics	.77
object/building vehicles	.73
abstract topics	.69
other vehicles	.60
movement/position vehicles	.40
<hr/>	
(adjectives	-.32)
(Metaphor density	-.42)

In this factor, a large number of metaphor features are clustered together, and there are only positive features, since the variables at the negative end of the scale have higher loadings in other factors and are therefore disregarded for the computation of factor scores (but they are considered during factor interpretation). This paints a non-specific picture of metaphor use, one that does not seem to differentiate between different kinds of topics and vehicles. It seems to suggest that those three registers appear to have no preferences for particular kinds of metaphorically used words. Abstract and social topics are linked to particular kinds of vehicles, but not to metaphor density (it is in brackets because it had a higher loading on factor 1). This suggests that there is some association between abstraction and metaphorical language, but not between abstraction and metaphor frequency.

The distribution of registers along dimension 2 is shown below.

TABLE 11  
Dimension 2 'non-specific metaphor use'

Mean factor scores for each register

DIMENSION 2

+	
11	
...	
2	
1	newspaper
0	
-1	academic
-2	
-3	
-4	
-5	conversation
-6	
...	
-9	
-	

F=5.775, p=.005, R2=.145

Unlike in the previous factor, in this one registers are not distributed far apart, suggesting there is not much difference between them. I called this dimension 'non-specific metaphor use' because of the lack of correlation between particular kinds of metaphor and registers. Newspaper is the less metaphor specific register, which suggests that it will employ just about any kind of metaphorically used word or refer to about any topic metaphorically. Conversation, on the other hand, seems to be a little less non-selective, but not enough to have any noticeable preference (otherwise the variables in the factor would have broken differently across the positive and negative ends). The fact that conversation is also metaphorically sparse (as suggested in the previous factor) may also influence these results, since there may not be enough metaphors to go around in conversation to constitute some sort of solid preference for any particular topic or vehicle.

Finally, the structure of the third factor is shown below.

TABLE 12  
Factor 3 structure

FACTOR 3	
Variable	Loadings
nouns	.89
prepositions	.84
past participle	.50
(proper nouns	.49)
(movement/position vehicles	.34)
(abstract topics	.30)
<hr/>	
(pron 1st person	-.41)
(pron 2nd person	-.40)

None of the variables that entered in the calculation of factor scores for dimension 3 is metaphor-related, namely nouns, prepositions and past participles. The remaining variables (in brackets) have higher loadings in other factors. These three variables seem to suggest an information focus, since nouns and prepositions are used in nominal groups which can package information densely. And past participles can form part of passive voice, which is a common feature of elaborate informative and/or argumentative registers. Pronouns, which cluster together on the negative pole of the dimension, are indicative of an interactive focus. This distribution of variables resembles in part that of Biber's (1988) first dimension, 'Involved vs. informational production', and so our dimension was named after that.

Metaphors are often thought of as devices that can help express abstract ideas as more concrete ones. Thus, it is interesting that characteristics normally associated with abstraction and information, such as the ones in this factor, are not linked to higher metaphor use (metaphor density). There is some association to abstract topics and to movement and position vehicles (metaphors of things going up and down, in and out, etc.), though, but these have higher loadings on factor 2, shown previously.

The distribution of registers on dimension 3 appears below.

TABLE 13  
Dimension 3 'informational versus involved production'

Mean factor scores for each register

DIMENSION 3

+	
4	
...	
2	academic
1	
0	newspaper
-1	
-2	
-3	
-4	conversation
-5	
...	
-10	
-	

F=17.582, p=.000, R2=.349

Academic is the most informational register; newspaper is at the center, suggesting that on average it is both informational and involved. Conversation is at the bottom end of the scale, representing involved production. Once again, metaphors do not seem to come into play in defining conversation. This again reflects the scarcity of metaphor in this register.

On this factor, the ordering of registers is different from that on the other factors. In the previous factors, it was conversation – academic – news (regardless of polarity), and here it is academic – news – conversation. The ordering in and of itself is not particularly revealing, since registers are aligned on the scale according to their scores. What has remained consistent across the factors is the larger difference between the scores for conversation, on the one hand, and for the remaining registers, on the other. This suggests that conversation is a more distinctive register, which in turn perhaps reflects the basic distinction between spoken and written language, with the written registers (academic and newspaper) sharing more characteristics between themselves than with the spoken register.

In this section, I looked at the extent to which variation in metaphor is systematic and whether it can give rise to dimensions. Statistically significant

results suggest that there is systematic variation in metaphor use across registers, with conversation standing in contrast with both academic and newspaper as a more metaphor-scarce language variety. The type of metaphor used in registers was not a good predictor of variation, though. There was some evidence to suggest that abstract topics are often metaphorized in informational registers. Metaphor density, on the other hand, was a strong component in the factors, forming a pole in factor 1. Registers seemed to be distinguished in terms of the quantity of metaphors present in them, with written registers sharing most of the metaphors, and conversation the fewest.

It must be stressed that these dimensions are not final. Larger corpora must be analyzed before a definitive set of dimensions is agreed on. Biber himself carried out preliminary analyses (BIBER, 1985) before arriving at the six dimensions that are currently referred to. Problems such as the subjective nature of metaphor identification and the labor intensive nature of such work on large quantities of text surely impose limits on both the range of registers that can be investigated and the number of texts that are included to represent each register. Work on dimensions of variation has been made possible in large part by automatic taggers (especially the 'Biber tagger', which is a reference in MDA). Similarly, if research in metaphor dimensions of variation is to continue and expand, then software for metaphor identification must be developed. This is the topic of the next section.

## **5. Automated metaphor retrieval**

I have been engaged in developing software for metaphor extraction for several years. This has led to several prototypes of the Metaphor Candidate Identifier (BERBER SARDINHA, 2006; 2007a; 2010b; 2010b), a program that is intended to find possible metaphors (i.e., candidates) in corpora. It has been made available online for several years under different names (Metaphor Tagger, Metaphor Identifier). Support for the online versions has stopped while development of a desktop version is underway. The version I will report on here is number 4 (desktop), and it works as follows:

- (a) For each word token in a corpus (of Portuguese or English), grab its collocates from 5 words to the left to 5 words to the right.
- (b) For each of these collocates, determine its part of speech and lemma.
- (c) Build list of node and collocate pairs, including lemma and part of speech.

- (d) Search for each node-collocate pair in a database of metaphor patterns (built during training).
- (e) If match is found, consider that word token a potential metaphor; if not, consider it as not being a potential metaphor.

The basis of the program is a large metaphor pattern database, consisting of over 541 thousand patterns. An example of a pattern found in the database is:

NL\_CW2R varrer\_mapa (translation: sweep map)  
 NL: Node is a lemma  
 CW: Collocate is a word (not lemma)  
 2R: Collocate is at two words to the right of the node

This pattern will capture the expression ‘varrer do mapa’ (sweep off the map).

Not all patterns have positional constraints such as this; others will capture occurrences within the whole width of the collocational span. Others will be formed by semantic fields (represented in square brackets), such as:

abaixo [not concrete] (translation: under/below)

This pattern will match expressions such as ‘abaixo das expectativas’ (below expectations).

Semantic fields are entered in a separate database, in the form of word lists. Currently, the program will not do word disambiguation, and will simply match words in the lists to those in the corpus; errors may occur because of that, for instance, by treating ‘meia’ (sock) to be ‘meia’ (half). There are no word disambiguation programs for Portuguese available.

The program (written as a script in Unix shell and Perl) works reasonably fast, being able to process a million words in under five minutes in a standard desktop computer with 4 GB RAM.

The MCI outputs segments of text where it has found a metaphor pattern. The screen below shows the output of the analysis of a text on economics:

000001 , a oferta da moeda aumenta no mercado e a tendência  
 000002 , o dólar voltou a cair e a taxa Ptax terminou  
 000003 , que vem puxando para cima o saldo da balança comercial  
 000004 . " O dólar está caindo porque houve muita emissão nos  
 000005 . Além disso , o fluxo de dólares para o Brasil  
 000006 . O dólar comercial à vista caiu 1 ,78% , cotado  
 000007 0 ,89% . O dólar comercial à vista caiu 1 ,78%  
 000008 2003 , quando o dólar caiu 14 ,35% no período .  
 000009 Brasil reverteu a tendência de queda a partir de março ,  
 000010 Com a entrada maior de dólares no País , a oferta  
 000011 Econômica , a cotação da moeda norte-americana caiu 15 ,7% entre  
 000012 O dólar comercial à vista caiu 1 ,78% , cotado por  
 000013 a cotação da moeda norte-americana caiu 15 ,7% entre abril e  
 000014 a tendência da taxa de câmbio é cair . As informações  
 000015 alto endividamento . " O dólar está caindo porque houve muita

In this particular case, all of the 15 lines were correctly picked up as they all have at least one metaphor:

1. offer of currency grows
2. the dollar has fallen again
3. balance of trade is being pulled upwards
4. the dollar is falling
5. the flow of dollars
6. the dollar has fallen
7. the dollar has fallen
8. when the dollar fell
9. downward trend
10. dollars went in
11. exchange rate fell
12. dollar fell
13. exchange rate fell
14. exchange rate trend is downward
15. high debt ... dollar is falling

I tested the MCI on a small corpus that was then hand coded for metaphors, made up by the following texts:

TABLE 14  
Test corpus for the MC

Register	Texts	Tokens	Types
News on the economy	15	5,510	1,562
News on science	15	10,321	2,947
Political speeches	20	24,712	3,865
Total	50	40,543	6,543

I computed the following metrics:

Precision: Metaphors correctly found divided by the total number of attempts (an attempt occurs when the program selects a metaphor candidate).

Recall: Metaphors correctly found divided by the total number of existing metaphors in the corpus according to manual analysis.

Results appear below.

TABLE 15  
MCI precision

Texts	Metaphors correctly found by MCI (True positives)	Attempts (True positives + False positives)	Precision %
Economy	427	478	89.3
Science	364	606	60
Politics	1136	1591	71.4
<i>Overall</i>	<i>1927</i>	<i>2675</i>	<i>72</i>

TABLE 16  
MCI recall

Texts	Metaphors correctly found by MCI (True positives)	Existing metaphors	Recall %
Economy	427	578	74
Science	364	535	68
Politics	1136	1563	73
<i>Overall</i>	<i>1927</i>	<i>2676</i>	<i>72</i>

Both precision and recall were 72% on average for the whole corpus. This means that 7 out of every 10 candidates MCI identified were really metaphors, and for every 10 metaphors in the corpus, 7 were correctly picked up. A performance at 70% is far from the ideal 100% that would be initially expected of a metaphor retriever, but this must be weighed against the difficulties involved in finding metaphors in texts by hand. This is demonstrated by several studies, such as Cameron (2003, p. 169), who reports an initial agreement of only 14% among analysts on a text in her corpus. Beigman Klebanov, Beigman and Diermeier (2008), in their study on newspaper metaphors, observe that agreement varied between 1.7% to 4%. And Steen *et al.* (2010) also show discrepancy between human analysts. At the same time, both Cameron and Steen *et al.* show that disagreement can be avoided by having very clear criteria for what counts as a metaphor, and it can also be resolved through discussion between the analysts. Such results underscore the difficulties involved in identifying metaphors, and imply that the gold standard must remain hand analysis, despite its shortcomings. I agree with that, but would further add that machine analysis must not be seen as substitute for manual analysis of metaphor. And that machine analysis should be considered as an extra rater in research teams.

This is because just as different people tend to find different but true metaphors, so does the computer when compared to people. In another study (BERBER SARDINHA, 2010a), I compared two independent analyses, by the MCI and by hand, and showed that the MCI correctly retrieved a large number of metaphors that were not noticed by hand and eye. Figure 4 shows the results of this study: the intersection between the two procedures (manual and MCI) is small. Inspection of the metaphors found revealed that the computer analysis was more consistent, never missing any one metaphor that it was taught to recognize, generally conventionalized ones. Human analysis, on the other hand, was better at finding metaphors that depended on context to be noticed, and also spotted innovative metaphors. The computer never gets 'distracted' or tired, while humans do, especially in activities that demand sustained attention such as metaphor identification in corpora.

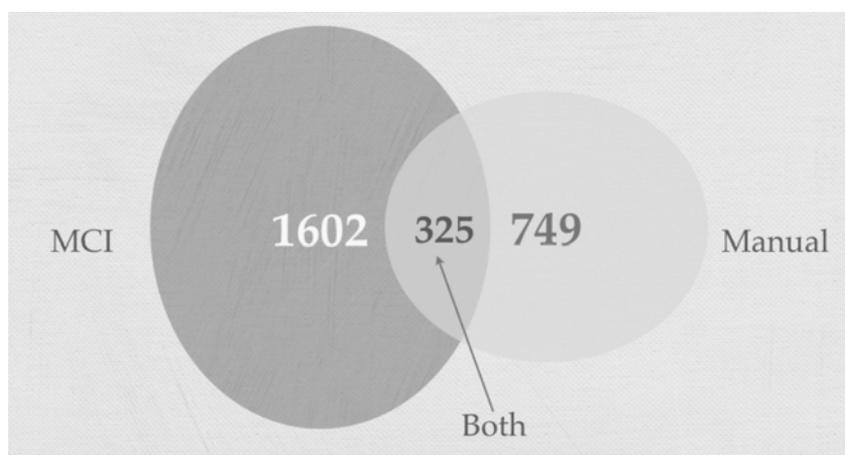


FIGURE 4: Comparison of true metaphors found by hand and by machine (MCI)

## 6. Final comments

In this paper, I argued that Corpus Linguistics has a great deal to contribute to metaphor studies, particularly with respect to research that shows:

1. The kinds of lexicogrammatical patterning that both arises from and signals metaphor in language use;
2. The extent to which metaphor use is patterned;
3. How metaphor varies across different genres and registers;
4. The extent to which such variation is systematic;
5. How research findings into linguistic patterning of metaphor can help develop tools to assist in automating at least in part the process of metaphor identification.

I also believe that these particular types of research can feed back on each other and support the development of resources to enable more CL metaphor research.

The development of resources such as metaphor identification assistance tools, semantically annotated corpora, and platforms for hand annotation of metaphors in corpora, among others, can all strengthen the important ties between metaphor and Corpus Linguistics. This way, the fields of metaphor and Corpus Linguistics can continue to mutually support and benefit from each other.

## References

- BEIGMAN KLEBANOV, B. *et al.* Analyzing disagreements. In: *Workshop on Human Judgements in Computational Linguistics*, Coling 2008, Manchester, UK. 2008.
- BERBER SARDINHA, T. A tagger for metaphors. Paper, RaAM - Researching and Applying Metaphor 6. Leeds, UK. 2006.
- BERBER SARDINHA, T. Análise de metáfora em corpora. *Ilha do Desterro*, v. 52, p. 167-201, 2007a.
- BERBER SARDINHA, T. Recontando a vida em narrativas pessoais: Um estudo de metáforas na perspectiva da Linguística de Corpus. *Organon*, v. 21, p. 143-160, 2007b.
- BERBER SARDINHA, T. Metaphor probabilities in corpora. In: ZANOTTO, M. S. *et al.* (Ed.). *Confronting Metaphor in Use: An Applied Linguistic Approach*. Amsterdam/Atlanta, GA: Benjamins, 2008.
- BERBER SARDINHA, T. Questões metodológicas de análise de corpora na perspectiva da Linguística de Corpus. *Gragoatá*, v. 26, p. 81-102, 2009.
- BERBER SARDINHA, T. Improving and evaluating the Metaphor Candidate Identifier. Paper, RaAM – Researching and Applying Metaphor Conference. Amsterdam, the Netherlands. 2010a.
- BERBER SARDINHA, T. MCI, um Identificador de Candidatos a Metáfora em corpora. In: SHEPHERD, T. *et al.* (Ed.). *Caminhos da Linguística de Corpus*. Campinas, SP: Mercado de Letras, 2010b. (Espaços da Linguística de Corpus).
- BERBER SARDINHA, T. A program for finding metaphor candidates in corpora. *The ESpecialist*, v. 31, n. 1, p. 49-68, 2010c.
- BHATIA, V. K. *Worlds of Written Discourse - A Genre-based View*. London, New York: Continuum, 2004. (Advances in Applied Linguistics).
- BIBER, D. Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics*, v. 23, p. 337-360, 1985.
- BIBER, D. *Variation across Speech and Writing*. Cambridge: Cambridge University Press, 1988.
- BIBER, D. Multi-dimensional approaches. In: LÜDELING, A.; KYTÖ, M. (Ed.). *Corpus Linguistics – An International Handbook*. Berlin / New York: Walter de Gruyter, 2009.
- BIBER, D.; CONRAD, S. *Register, genre, and style*. Cambridge; New York: Cambridge University Press, 2009. (Cambridge textbooks in linguistics).

- BIBER, D. *et al.* Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora*, v. 1, n. 1, p. 1-37, 2006.
- CAMERON, L. *Metaphor in Educational Discourse*. London: Continuum, 2003.
- CAMERON, L. What is metaphor and why does it matter? In: CAMERON, L.; MASLEN, R. (Ed.). *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities*. London: Equinox, 2010.
- CAMERON, L.; MASLEN, R. *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities*. London: Equinox, 2010. (Studies in applied linguistics).
- CHARTERIS-BLACK, J. *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan, 2004.
- CREET. *Metaphor Analysis Project*. 2006. Unpublished Work.
- DEIGNAN, A. *Metaphor and Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins, 2005.
- DEIGNAN, A.; POTTER, L. A corpus study of metaphors and metonyms in English and Italian. *Journal of Pragmatics*, v. 36, n. 7, p. 1231-1252, 2004.
- GIBBS, R. W. (Ed.) *The Cambridge Handbook of Metaphor and Thought*. New York: Cambridge University Press. 2008a.
- GIBBS, R. W. Metaphor and thought – The state of the art. In: GIBBS, R. W. (Ed.). *The Cambridge Handbook of Metaphor and Thought*. New York: Cambridge University Press, 2008b.
- GIBBS, R. W.; FRANKS, H. Embodied metaphor in women's narratives about their experiences with cancer. *Health Communication*, v. 14, n. 2, p. 139-165, 2002.
- HALLIDAY, M. A. K. Corpus studies and probabilistic grammar. In: AIJMER, K.; ALTENBERG, B. (Ed.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 1991.
- MOULES, N. J. *et al.* Making room for grief: walking backwards and living forward. *Nursing Inquiry*, v. 11, n. 2, p. 99-107, 2004.
- PRAGGLEJAZ GROUP. MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, v. 22, n. 1, p. 1-39, 2007.
- SCOTT, M. *WordSmith Tools, version 4*. Oxford: Oxford University Press, 2004.
- STEEN, G. *Finding Metaphor in Grammar and Usage*. Amsterdam / Philadelphia: John Benjamins, 2007.
- STEEN, G. *et al.* *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Amsterdam: John Benjamins, 2010.

STEFANOWITSCH, A. Corpus-based Approaches to Metaphor and Metonymy. In: STEFANOWITSCH, A.; GRIES, St. Th. (Ed.). *Corpus-based Approaches to Metaphor and Metonymy*. Berlin; New York: M. de Gruyter, 2006.

STEFANOWITSCH, A.; GRIES, St. Th. *Corpus-based Approaches to Metaphor and Metonymy*. Berlin; New York: M. de Gruyter, 2006. (Trends in linguistics. Studies and monographs, 171).

STEVENS, J. *Applied multivariate statistics for the social sciences*. 4. ed. Mahwah, N.J.: Lawrence Erlbaum Associates, 2002.

WIKBERG, K. The role of corpus linguistics in metaphor research. In: JOHANNESON, N. L.; MINUGH, D. (Ed.). *Selected Papers from the 2006 and 2007 Stockholm Metaphor Festivals*. Stockholm: Department of English, Stockholm University, 2007.

Recebido em 08/04/2011. Aprovado em 08/05/2011.