METAPHOR USE AND REGISTER VARIATION

Register variation and metaphor use: A multi-dimensional perspective

Tony Berber Sardinha*

São Paulo Catholic University
ABSTRACT

The main goal of the research reported here is to assess the relationship between metaphor use and register variation in English. The corpus used was the VU Amsterdam Metaphor Corpus (VUAMC), comprising 84 texts taken from the BNC Baby, all hand-coded for metaphor. The method follows the Multi-Dimensional framework for register variation introduced and developed by Biber (1988 *et seq*.). One of the outcomes of MD research are dimensions of variation, the parameters that underlie variation across texts; by deriving such dimensions, MD analysts are able to identify the systematic patterns of co-occurrence across linguistic features and determine the degree of relationship between such patterns and situational characteristics. To date, no MD study had looked at register variation from the perspective of metaphor use in English. The corpus was fully re-tagged for part of speech with the Biber tagger, annotated for semantic features with the Eng-CG tagger, and post-processed by the Biber Counter and by scripts especially designed for this investigation. Guided by five research questions, results showed that: (1) the existing dimensions of variation for English (Biber, 1988) cannot account for metaphor use; (2) there are four dimensions of variation in the corpus, two of which are related to metaphor, while the others are strictly grammatical; (3) metaphor accounts for 41% of register variation; (4) metaphor use is more strongly predicted by text groupings that are independent of register; (5) there are two text types, or linguistically defined text groups, that combine both grammatical and metaphor features.
1. **Introduction**

Corpus-based research into metaphor has provided valuable insights into a range of aspects of metaphor use, including frequency, patterning and distribution within and across texts. One of the most consistent findings coming out of this research is that metaphor use varies across different specialized text forms, such as genres, registers, and text types.

Evidence of a relationship between metaphor use and specialized texts is offered both directly and indirectly in the literature. Indirect evidence is offered by different studies that look each into a distinct text variety. If we concentrate on frequency alone, which is just one of the features considered in these studies, we will realize that it varies considerably across different varieties. For instance, for classroom discourse, the average rate of metaphor use is one every 37 words in classroom discourse (Cameron, 2003), and one every 22 in business conference calls (Berber Sardinha, 2008a). Taken together, these two studies suggest that register variation may be influenced by metaphor frequency, *in other words, that metaphor may be an intrinsic element in register variation*. A problem with approaching metaphor across different studies like this is that many variables are not controlled for, and the differences noted may be due to unknown influences rather than to true situational or linguistic characteristics relevant to these registers. This problem can be ameliorated by looking at metaphor and register variation directly in register-diversified corpora. Studies focusing on metaphor variation directly control for identification method, among other issues, and therefore offer a more straightforward perspective on variation.
Few studies have taken a direct perspective on metaphor and register variation, not least because of the large amount of effort involved in coding metaphor in multiple subcorpora. There are two main sources for register variation studies in metaphor use. One is a set of studies comprising Steen et. al (2010), Krennmayr (2011), Dorst (2011), Kaal (2012), and Herrmann (forthcoming) all of which employed the VU Amsterdam Metaphor Corpus (VUAMC), a sample of the BNC Baby entirely hand-coded for metaphor using MIPVU as the identification method; and Steen et al. (2010) and Pasma (2011), who analyzed a similar Dutch corpus. Analyses of the VUAMC show that metaphor relates to register mainly with respect to frequency and word class. As far as frequency, metaphor occurs more often in academic writing (where 17.5% of the words are metaphorically used), followed by news (15.3%), fiction (10.9%), and conversation (6.7%). And as far as word class, most metaphorically used words are verbs and prepositions in all registers except academic writing. Results also reveal that the share of different parts of speech varies significantly across registers, suggesting intricate patterns of metaphor variation. The second literature source has only one study to date, namely Berber Sardinha (2011b), which takes a multi-dimensional perspective on metaphor and register variation. Based on a small corpus of Brazilian Portuguese comprising newspaper, academic, and conversation texts, this study has identified two tentative dimensions of metaphor variation, namely ‘involved narrative production versus metaphor use’, and ‘non-specific metaphor use.’ The first one reflects an interaction between frequency of metaphor use and narrative concerns, and shows a split between conversation, on the ‘low metaphor frequency, narrative’ end of the dimension, and the other two registers, which are ‘high metaphor frequency, non-narrative’. In turn, the second dimension marks a distinction between conversation,
being more ‘specific’ with respect to metaphor use, as it selects mostly topics related to
‘people’, and the other registers, which are more diverse with respect to topic.

Altogether this body of research presents a complex picture of the relationship
between metaphor and register variation. As Kaal (2012, p.56) argues, ‘the study of
metaphor would greatly benefit from a register-variation approach that is able to
separate register-defining metaphor use from general and shared patterns’. To
accomplish this, this chapter resorts to the (Multi-Feature) Multi-Dimensional
framework (Biber, 1988 et seq.) and multivariate statistical techniques in order to
examine the interplay between register variation and metaphor use in English.

In keeping with the theme of this volume, metaphor and discourse specialization, this
chapter takes a comparative perspective on specialization, by looking at a set of
different registers. The registers analyzed here include academic writing, news, fiction
and conversation, and they range from highly specialized, such as academic writing, to
less specialized, i.e. conversation.

1.1. The Multi-Feature Multi-Dimensional approach

The Multi-Feature Multi-Dimensional, or simply Multi-Dimensional (MD), approach
was developed by Biber (1988 et seq.) with the aim of finding the communicative
parameters that underlie register variation. The MD approach has been applied to a
number of different contexts, ranging from specific genres, like TV episodes (Rey,
2001), to broad registers such as conversation (Biber, 2004), and even to whole
languages, such as English (Biber, 1988), Spanish (Biber, Davies, Jones, & Tracy-
Ventura, 2006; Parodi, 2007) and Portuguese (Berber Sardinha, 2011a). With the
exception of Berber Sardinha (2011b) and the project reported here, no MD study has
incorporated metaphor in the set of features analyzed. The steps in a full MD analysis are the following:

1. A corpus is collected that represents the registers needed to attain the research objectives.
2. A review of the literature indicates the features necessary for the analysis of the corpus.
3. A corpus is tagged automatically, interactively and/or by hand for those features.
4. The frequencies of each feature are counted, and normalized to control for text size.
5. The data are screened statistically, and features that do not meet the necessary requirements for factor analysis are discarded.
6. Frequencies for the remaining features are further standardized to a mean of 0 and a standard deviation of 1 to control for differences in feature frequency.
7. An initial unrotated Factor Analysis is run to determine the number of factors in the data as well as the variables that will remain in the analysis.
8. Further factor extractions are performed, and rotation of these factors indicate which variables are present on each factor, with their respective weights.
9. Each text receives a factor score, based on the standardized frequencies of the variables on each factor.
10. A mean score is calculated for each register, and registers are placed on a scale representing the range of scores on each factor.
11. Factors are interpreted and given labels to express their underlying communicative or discourse parameters. These are the dimensions of register variation.
METAPHOR USE AND REGISTER VARIATION

12. Statistics are computed to assess the amount of variation captured by each dimension.

2. Method

2.1. Corpus

The corpus used for this study is the VUAMC (VU Amsterdam Metaphor Corpus), which in turn is a sample of BNC Baby that was hand coded for metaphor by the members of the Metaphor in Discourse project (Steen, et al., 2010). The corpus was downloaded off the Oxford Text Archive and after processing with the Biber tagger, it had the size shown in Table 1.

INSERT TABLE 1 ABOUT HERE

2.2. Metaphor identification

The VUAMC corpus had been coded for metaphor using MIPVU, a procedure that consists of the following steps:

1. Find metaphor-related words (MRWs) by examining the text on a word-by-word basis.

2. When a word is used indirectly and that use may potentially be explained by some form of cross-domain mapping from a more basic meaning of that word, mark the word as metaphorically used (MRW).

3. When a word is used directly and its use may potentially be explained by some form of cross-domain mapping to a more basic referent or topic in the text, mark the word as direct metaphor (MRW, direct).
4. When words are used for the purpose of lexico-grammatical substitution, such as third person personal pronouns, or when ellipsis occurs where words may be seen as missing, as in some forms of co-ordination, and when a direct or indirect meaning is conveyed by those substitutions or ellipses that may potentially be explained by some form of cross-domain mapping from a more basic meaning, referent, or topic, insert a code for implicit metaphor (MRW, implicit).

5. When a word functions as a signal that a cross-domain mapping may be at play, mark it as a metaphor flag (MFlag).

6. When a word is a new-formation coined, examine the distinct words that are its independent parts according to steps 2 through 5. (Steen, et al., 2010, p.25)

2.3. Tagging

The corpus had its original part of speech tags removed and was then tagged over with the Biber tagger, a software program regularly employed in MD research that automatically identifies over 400 different grammatical features. It was later post-processed by the Biber Counter program, which calculated the frequencies of 127 selected features. To identify semantic features, the Eng-CG tagger was used (beta.visl.sdu.dk/visl/en).

2.4. Variables

A review of the metaphor literature indicated the following features as potentially relevant:

1. Metaphor frequency: density of metaphor use differentiates registers, as noted in the introduction.
METAPHOR USE AND REGISTER VARIATION

2. Metaphor signaling: words that act “as a signal that a cross-domain mapping may be at play” are considered metaphor signals, or flags (Steen, et al., 2010, p.26).

3. Metaphor manifestation: metaphors can be expressed:
   a. Directly: Words whose metaphorical status is signaled directly (Steen, et al., 2010, p.39), by words such as resembling, as and like. E.g.: He’s like a favorite old coat. (Steen, et al., p.93).
   b. Indirectly: Words whose metaphoricity is not explicitly signaled (Steen, et al., 2010, p.33). This is how metaphors manifest themselves by default. Eg: high wages (Krennmayr, 2011, p.31).
   c. Implicitly: Words whose metaphorical status is realized by substitution (e.g. it in to embark on such a step is not necessarily to succeed immediately in realizing it, where it refers back to the metaphorically used word step) or ellipsis (but he is [an ignorant pig], where is receives the code for implicit, elliptical metaphor, in place of the omitted fragment in brackets) (Steen, et al., 2010, p.40).

4. Metaphor clustering: Metaphor cases are distributed unevenly in text, forming clusters of neighboring metaphors (Cameron & Stelma, 2004).

5. Metaphor conventionalization: Conventionalized metaphors, or those that “go unnoticed in everyday life” (Deignan, 2005, p.5, 40-47), represent a large share of metaphor use.

6. Metaphor semantics: Assigning metaphor cases to semantic groupings is useful in categorizing metaphor use (Cameron & Maslen, 2010). In addition, semantic
fields have been used as a starting point for metaphor detection (Berber Sardinha, in press; Hardie et al., 2007; Kaal, 2012; Krennmayr, 2011).

7. Metaphor word class: Previous studies found that frequency of different parts of speech distinguishes registers (Dorst, 2011; Kaal, 2012; Krennmayr, 2011; Pasma, 2011; Herrmann, forthcoming).

2.5. Research Questions

The following questions were formulated and will be addressed in turn in the chapter:

1. Is there a relationship between Biber’s (1988) dimensions of variation for English and metaphor use? In other words, by using the main dimensions of variation (Biber, 1988), and given that we know the scores for each text on them, are there significant associations between these scores and metaphor use? If so, which parameters of variation are associated with metaphor use?

2. If we run a new Factor Analysis, what dimensions of variation are specific to the VUAMC corpus? Do they differ from those obtained by Biber (1988)? What is the role of metaphor on the dimensions?

3. Are differences among mean register scores on each dimension significant? That is, to what extent can these dimensions predict register?

4. What are the text groupings that cut across register categories for the individual dimensions?

5. What are the text types (i.e., text varieties defined solely according to linguistic criteria; Biber, 1988) present in the corpus?
METAPHOR USE AND REGISTER VARIATION

2.6. Variables

A review of the literature on metaphor was operationalized as a variable set, which is specified as follows.

Three different kinds of variables were extracted from the corpus: structural, semantic, and metaphor-related. The source for structural variables were the tags assigned to each word by the Biber tagger and then post-processed by the Biber counter. For semantic variables, the source was the tagging done by the Eng-CG tagger. And for metaphor variables, the source was the manual annotation added to the corpus by the MIP-VU team.

The search strings used for extracting metaphor-related words and flags from the VUAMC appear in Table 2.

| INSERT TABLE 2 ABOUT HERE |

| INSERT TABLE 2 ABOUT HERE |

| All borderline cases were incorporated in the actual counts for each metaphor type. |

| Metaphor variables were combined with both structural and semantic variables and processed through scripts developed by the author using algorithms such as: |

| For each metaphor category (indirect, direct, implicit, flag): |

| a. Find each metaphor-related word in each text; |
| b. Detect the major part of speech assigned to it; |
| c. Count how many times this metaphor-related word was used in each text as that part of speech; |
| d. Repeat for each semantic field. |

The major part of speech categories were the following 20:
adjective, adverb, article, coordinating conjunction, determiner, existential *there*,
foreign word, infinitive marker, interjection, *not*, noun, number, pre-quantifier,
preposition, pronoun, qualifier, subordinating conjunction, *that* clause, verb, wh-word.

This algorithm produced a breakdown of the frequency of each metaphor category
by part of speech and by semantic class. For instance, indirect metaphors were counted
as:

Indirect metaphors, nouns;
Indirect metaphors, adjectives;
Indirect metaphors, adverbs;
Indirect metaphors, action semantic field;
Indirect metaphors, building semantic field;
Indirect metaphors, container semantic field;

And so on.

In addition to structural, semantic and metaphor-related variables, two other
variables were computed.

The first was metaphor bundles, which captures metaphor conventionalization. It
was operationalized through a count of high frequency lexical bundles (Biber, Conrad,
& Cortes, 2004) that incorporates a metaphor-related word. Bundles were processed as
follows:

1. For each file, create 3-grams, that is, sequences of three neighboring words;
2. For each of these 3-grams, determine if it incorporates metaphorically used
   words by matching it to the annotated corpus;
METAPHOR USE AND REGISTER VARIATION

3. For each 3-gram that has a metaphorically used word, match it to the list of COCA (Corpus of Contemporary American English; corpus.byu.edu/coca) 3-grams;


For example, in the excerpt in Table 3 (file FEF, academic), there are two metaphorically used words and four matching COCA 3-grams.

INSERT TABLE 3 ABOUT HERE

The second extra variable was metaphor clustering, which operationalizes metaphor distribution. A metaphor cluster was defined as a sequence of at least three metaphor-related words (direct, indirect or implicit metaphors) at a distance not greater than $D$ from each other. $D$ was calculated as the average distance between metaphors, which was eight ($208,915$ tokens / $25,429$ MRWs = 8.21, rounded off to 8).

Table 4 gives details of the number of features present in the corpus by variable type.

INSERT TABLE 4 ABOUT HERE

The total number of features present in the corpus excludes semantic variables because these were only counted for metaphor-related words, and therefore their count is already incorporated in the metaphor tagset. Hence, all variables fell in two groups: structural and metaphor-related.
The figure for structural variables refers to the features selected by the Biber counter, and it is smaller because the counter considers only the most important tags in the tagset, collapsing tags into groups.

All counts were normed per 1,000 words. For example, the frequency of indirect metaphors in text A1E was 86. Since this text had 598 tokens, its normed count for indirect metaphors was 143.8, that is, $86/598 \times 1000$. Norming controls for text size, thus enabling comparisons across texts of different sizes.

3. Results

3.1. Question 1: Relationship between metaphor and previous dimensions of variation

Answering this question did not require that a new factor analysis were run, but that the factors already extracted by Biber (1988) be applied to the VUAMC corpus. Hence, for each text in the corpus, text scores were computed using the same variables, means and standard deviations as used in Biber (1988). After that, correlations were run between the scores for each text on Biber’s (1988) dimensions and the main metaphor variables, namely, counts for indirect, direct, and implicit metaphor, and for metaphor flags. Results appear in Table 5.

In this table...
METAPHOR USE AND REGISTER VARIATION

the only type that had had any significant scores at all, meaning that any relationship between these dimensions and metaphor use is restricted to indirect metaphors.

Dimension 1 (information vs involvement) had a negative coefficient, meaning that a high score on the ‘information’ side of dimension would predict a large number of metaphors in the text, while a high score on ‘involvement’ would indicate few metaphors. In turn, Dimension 3 (explicit vs situation-dependent reference) had a positive score, indicating that a high scoring text on ‘explicit reference’ would have many metaphors, while a ‘situation-dependent’ text would have few metaphors. It must be stressed that because no high coefficients were found, no direct relationship must be assumed between these dimensions and metaphor use.

The answer to question 1 is therefore that there is no strong statistical association between the existing dimensions for English and metaphor use. In other words, it would not be possible to predict metaphor use in quantitative terms simply by knowing the text scores for those dimensions. This prompts an analysis that incorporates metaphor use, so as to better take account of the role of metaphor in register variation.

3.2. Question 2: Dimensions of variation

A fresh multi-dimensional analysis must include the extraction of factors, rather than relying on previously determined ones. To do that, variables were selected that met two main criteria:

a. The subjects to variables (STV) ratio should be at least 2, that is, there had to be twice as many texts (subjects) as there were variables. There are many different STV ratios in the literature, from just above 1 to 20 (Zhao, 2009). In the MD literature, STV’s range from 2.6 for Korean (150 texts, 58 features), 5.3 for
Tony Berber Sardinha

Tuvalan (222 texts, 42 features), 7.2 for English (481 texts, 67 features), and 7.4 for Somali (483 texts, 65 features) (Biber, 1995, p.89-100). In this study, hand analysis of metaphor constrains the size of the corpus, even though the VUAMC corpus has more tokens than either the Somali or the Korean corpus (Biber, 1995), but fewer texts. At the same time, metaphor needs to be explored in as much detail as possible, which contributes a large number of features. These two considerations led to the decision to adopt an SVT that was compatible with both prescriptions in the statistics literature and the lower range of previous MD analyses. Consequently, as there are 84 texts in the corpus, the maximum number of variables was 42.

b. These 42 variables should have the highest communalities in the data set, and no communality should be less than .2 (Biber, 1995, p.138).

To obtain these 42 highest communalities, an initial factor analysis was run, using a subset of the total 250 variables. This subset comprised all 47 variables used to compute the five main dimensions of variation for English in Biber (1988), in addition to the 52 metaphor-related variables that had a mean normed frequency of .1 (i.e., they occurred at least 10 times per 1,000 words). This factor analysis solution was unrotated, using Principal Axis Factoring as the extraction method. The 42 variables retained appear in Table 6.

INSERT TABLE 6 ABOUT HERE

It turns out that there were 26 structural variables and 16 metaphor-related variables. Indirect metaphor was by far the most common type on the list, and it was also the most
frequent in the corpus. This suggests that the other kinds of metaphor (direct and implicit) and metaphor signaling devices (flags) may not be as relevant to mark register variation in English.

A further unrotated factor analysis was run with these 42 variables. The scree plot indicated four factors as the likely number of factors in the data (each dot along the line is a factor). After the fourth dot, there is a break on the line, forming a 'scree' (Figure 1).

INSERT FIGURE 1 ABOUT HERE

A third factor analysis was then conducted, this time requiring that four factors be extracted. Principal Axis Factoring was used for extraction, with Promax as the rotation method, which are standard choices for MD analyses (Biber, 1988). To test for sampling adequacy, the Kaiser-Meyer-Olkin (KMO) test was carried out, achieving .811 out of a maximum of 1, thus signaling that the data set used in this analysis (84 texts and these 42 variables) is adequate. Bartlett’s test of sphericity was also performed, to rule out the possibility that the data were an identity matrix (where all diagonal cells are 1, and off-diagonal ones are 0), which would mean that the variables were not correlated except with themselves. The result was significant ($\chi^2 = 3055.1, p=.000$), suggesting that the data were not an identity matrix, thus being adequate.

Factor 1 had two poles (negative and positive), shown in Table 7; features in parentheses had a higher loading on a different factor, and therefore were not entered in the computation of text scores for this factor. Hence, they did not play a part in the distribution of registers along the dimensions, but were taken into account during the interpretation of the data. The table shows the features that loaded on this factor; as with
other factors, and consistent with Biber (1988), only those features with a loading of at least .35 were considered.

**INSERT TABLE 7 ABOUT HERE**

Factor 1 did not include any metaphor-related variables, and was very similar to Biber’s (1988) Factor 1. In his research, it was interpreted as marking a distinction between Involvement and Information in language production, with features in the positive pole indicating the former, and those in the negative pole signalling the latter. A comparison of the features shared by both factors shows the overlap is not perfect, as the following features that are on our Factor 1 were on different factors in Biber (1988): modals of prediction, in Biber’s dimension 4, Overt expression of persuasion; past tense in his dimension 2, Narrative concerns; and both agentless passives and passive nominal modifiers in his dimension 5, Abstract information. With the exception of past tense verbs, all others were entered in the computation of our factor 1 text scores. In addition, Biber’s factor 1 had more features than ours, with 23 exclusive variables versus 17.

To determine the real extent of overlap, text scores were computed for each text on both this factor and on Biber’s (1988) factor 1, for all texts in the VUAMC corpus. The variables chosen to compute scores for Biber’s factor 1 were those used in his study, whereas the variables to calculate scores for our factor 1 were those shown in Table 7. Results showed an almost perfect correlation (r=.978, p=.000), thus suggesting that these two factors are tapping into the same dimension, even if with slightly different
METAPHOR USE AND REGISTER VARIATION

features. Hence, this factor received the same label as Biber's (1988) factor 1 (Involved versus Informational production).

Mean text scores for each register were calculated and plotted on the graph (Figure 2).

Again, this distribution resembles Biber's (1988) dimension 1, with conversation and fiction being more 'involved' and news and academic writing being more ‘informational’.

The absence of metaphor-related features is revealing, because it indicates that metaphor does not play a role in this dimension of English. In other words, metaphor does not seem to be a key device for expressing either involvement or information in English.

Factor 2 concentrated most of the metaphor-related variables, and had only one pole (positive). Table 8 gives details of its features.

Indirect metaphor was the only metaphor type present on the factor. As with the previous factor, correlations were computed between the scores for each text on this factor 2 and their scores on Biber's (1988) factors. As Table 9 shows, correlations were moderate with factors 1 and 3, and low or non-existent with the others. Therefore, this factor seems to be unique, in that it does not share any of the previous underlying
parameters of register variation for English. This is to be expected, since no previous study of register variation in the MD framework included metaphor variables. At the same time, it does suggest that metaphor has a role to play in register variation in English, and should be considered for inclusion in MD studies.

The salient metaphor features comprised indirect metaphors and its subtypes, that is, indirect metaphor nouns and adjectives, in addition to specific semantic fields conveyed by these metaphors, namely abstraction (view, assessment, scale, etc.), action (collapse, come, charge, etc.), perception (skilled, vocational, clear, etc.), process (flow, growth, drift, etc.), and features (power, interests, technique, etc.). They also encompassed clusters, which reflect metaphor distribution within texts, and bundles, which represents conventionalization.

The fact that most metaphor variables were present on this factor, coupled with the observation that all of them are indirect metaphor, the most frequent kind, led to the hunch that this factor might be correlated with the total count of metaphors in the texts. To test this hypothesis, correlations were run between individual text scores on this factor and the total number of metaphors in each text (the sum of counts for indirect, direct and implicit metaphors) (Table 10). It must be remembered that this total was not previously included in the analysis, and therefore had not been directly accounted for.
Results show a very high correlation with factor 2, confirming the importance of metaphor density on this factor. The correlation was high because this factor includes the most frequent type of metaphor (indirect), and when its various manifestations were summed up, this count approached the total number of metaphors in each text. Consequently, a suitable interpretive label for this factor is simply Metaphor Density.

The plot in Figure 3 shows the mean scores for each register on Dimension 2.

The plot shows a basic opposition between metaphor dense registers (academic prose and news) and metaphor sparse ones (fiction and conversation).

The sample below illustrates a very dense presence of metaphor in academic text FEF (score = 24.91):

Since it is the equation [formula] that needs to be satisfied, we should choose the potential as a vector (called, not without logic, the vector potential), defined by the equation [formula]. Substituting the above equation into eqn (3.1) we get [formula].

The ten metaphors in this 41-word excerpt represent a density of one metaphor every 4.1 words, or a normed count of 244 metaphors per thousand words. For the whole text, the count is 241 metaphors per thousand words, which is higher than the average, at 146 metaphors per thousand words.

The text is also dense with metaphors clusters: the normed count for clusters is 170, higher than the corpus average of 74. In the sample, these clusters are present:

satisfied...potential...called

potential...called...without
TONY BERBER SARDINHA

called...without...potential
without...potential...defined
Substituting...above...into
above...into...get

There are three bundles in the sample (figures in brackets are COCA frequencies): to be satisfied (323), potential as a (177), defined by the (915). The overall normed count for bundles in the text is 165, higher than the corpus average of 123.

The salient semantic fields for this dimension are frequent in the text; below are some examples (bold identifies semantic features).

Abstraction: All our variables can be functions of space…

Action: Since charged particles rarely travel close to the velocity of light…

Feature: if the current density is specified,…

Perception: Using a technical term borrowed from electron optics,…

Process: When current flows through two materials,…

News has the second highest score on this dimension, indicating that is also metaphor dense. Below is a text sample (file A9J, score 18.51) that illustrates the salient features on the factor:

All these have in fact only constituted the outer layer of an inner mechanism of civilian rebellion. The second, and complementary, part of the Intifada strategy concentrated on the construction of an indigenous national authority. As Israel’s authority was being lifted, the population began creating an alternative authority of their own.

The 16 metaphors in this 52-word sample represent a high metaphor density of one metaphor every 3.25 words. In the whole text, the normed total metaphor count is 278,
considerably higher than the average of 146. All of the metaphors in the sample are indirect.

Ten clusters are formed by these metaphors:

- these...in...outer
- in...outer...layer
- outer...layer...of
- layer...of...inner
- of...inner...mechanism
- inner...mechanism...of
- part...of...concentrated
- of...concentrated...on
- concentrated...on...construction
- on...construction...of

For the whole text, the normed frequency for clusters is 216, much higher than the corpus average of 74.

The sample is also teeming with 13 bundles: *have in fact* (160), *in fact only* (46), *the outer layer* (70), *outer layer of* (88), *of an inner* (86), *part of the* (60849), *concentrated on the* (493), *on the construction* (176), *the construction of* (3249), *construction of an* (159), *of an indigenous* (80), *authority of their* (30), *of their own* (8854). For the whole text, the normed bundled frequency is 197, again higher than the corpus average of 123.

The sample has one metaphor in the semantic field of action (*construction*), but the text has several occurrences of the salient semantic characteristics, such as those shown below.

Abstraction: *The struggle reflects a struggle in mass consciousness.*
TONY BERBER SARDINHA

Action: a rotation Unified National acts in the role of field-general

Feature: we have the capacity to bring coherence and organization to…

Perception: it is a conscious effort.

Process: its pains and sufferings, are all reminiscent of the process of birth.

Fiction is on the negative pole of the dimension, which means it exemplifies metaphor sparseness. The sample below is from text FAJiii (score of -10.74):

She has in front of her a typewriter, her recorder and her notes. A servant is preparing lunch on a fire. Claudia hears something. She looks up into the sky in the direction of Ol Doinyo Lengai.

The only metaphor-related word is in, and as a result there are no clusters in the sample. Likewise, there are no frequent COCA bundles formed with in, and no semantic features. The whole text has only 73 metaphors per 1,000 words, which is only half of the 146 found on average in the corpus. This has the effect of reducing both clusters and bundles counts as well; the text has only 84 bundles and 13 clusters per 1,000 words, as opposed to 123 and 74 on average, respectively, for the corpus.

Finally, the lowest scoring register on this dimension is conversation. The following excerpt from text kcc (score of -14.24) illustrates its metaphor sparseness:

PS0F5: He bought himself er a padded shirt, you know?
PS0F7: A padded shirt like Alfred had for Christmas?
PS0F6: Mum.
PS0F7: Robert had for Christmas.
PS0F5: He had one for Christmas, did he?
PS0F7: Well I, I, I asked him what he wanted for Christmas.
PS0F5: Yeah?
METAPHOR USE AND REGISTER VARIATION

PS0F7: And he said get me a padded shirt, he said.

This sample has no metaphors, and consequently no clusters, bundles or metaphor-related semantic features. The full text has a normed frequency of only 45 metaphors, 47 bundles and 5 clusters, as opposed to corpus averages of 146, 123, and 74, respectively. The difference between the total of metaphors (45) and clusters (5) indicates that those few metaphors are a long way from each other along the text, which highlights the absence of metaphors in the text.

Factor 3 had no metaphor-related variables, just like factor 1. Its pattern matrix appears in Table 11.

INSERT TABLE 11 ABOUT HERE

Factor 3 has five features on the positive pole, but only two of them have higher loadings on it, namely coordinating conjunction as a phrasal connector and singular noun nominalization; the other features have higher loadings on the negative pole of factor 1, Informational production. The negative pole has four features, three of which are more salient on this factor (perfect aspect verbs, third person pronouns, and past tense verbs); the other variable, private verbs, has a higher score on factor 1, for Involved Production.

The salient features on the positive pole are both part of Biber’s (1988) Dimension 3 positive pole, Explicit Reference. And salient features on the negative pole all appear in Biber’s (1988) Dimension 2 positive pole, Narrative concerns. Correlations between this factor and Biber’s (1988) factors are shown in Table 12.
These correlations indicate a high degree of inverted overlap between this factor and Biber’s (1988) Dimension 2, narrative concerns, that is, texts that score high on this factor tended to score low on Biber’s Dimension 2, and vice-versa. In other words, the positive pole of this factor is associated with non-narrative concerns, and the negative, with narrative concerns. There is also a moderate correlation with dimension 3, Explicit reference. This factor seems to be a blend of Biber’s (1988) Dimension 3 and Dimension 2, and was therefore labeled ‘Explicit reference versus Narrative Concerns’.

The mean scores for each register were computed and plotted on the graph in Figure 4.

The plot shows academic writing to be the register that depends on explicit reference the most. News is mostly neutral with respect to either parameter; conversation is mostly narrative and situation-dependent for reference, and fiction is the most marked for both narrativity and situation-dependent reference.

However, the absence of metaphor is relevant, and suggests that metaphor is not a key element in marking either parameter of variation, namely narrativity and explicit reference. Just like with factor 1, which also saw an absence of metaphor, this factor does not rule out the possibility of using metaphor to narrate events or to make references explicit, but it indicates that metaphor is not a central device for these functions in English, at least from a quantitative, register variation perspective.

Factor 4 is the last factor extracted, and its pattern matrix is shown in Table 13.
Factor 4 is a mixture of both metaphor and structural features. The metaphor features are all related to implicit metaphor, namely the total count for this type of variable, and the use of pronouns as implicit metaphorical devices. The structural features, in turn, are possibility modals, infinitive verbs, and (negatively) phrasal coordinating conjunctions (which has a higher weight on factor 3). An examination of the occurrences of these features revealed that the salient metaphor and grammatical features seldom co-occur in the same clauses (as in *which allows them to be legally sold in the UK*, which combines an implicit metaphor and a to-infinitive clause), and therefore metaphor and structure mostly work in separate syntactic units within the texts. Implicit metaphor is a rare feature (mean frequency of .3 per 1,000), and metaphorical pronouns realizing metaphor are even more infrequent (.2 per 1,000). Structural features are much more frequent: infinitive verbs occur 13.02 times per 1,000, possibility modals 6.2, and phrasal coordination 1.9. Hence, the salient grammatical features are much more pervasive than the metaphor ones.

Infinitive verbs serve a variety of functions, such as reporting “intentions, desires, efforts, perceptual states, and various other general actions” (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p.693). Coordinating conjunctions “link elements with the same syntactic role” (Biber et al. 1999: 79), in this case, within phrases. And modals of possibility express the likelihood of actions or states.

Correlations with Biber’s (1988) dimensions are all low, as Table 14 shows. This suggests that this factor represents a dimension not yet described for English.
Correlations between text scores for factor 4 and variables not on the factor were computed. Correlations are moderate or weak, which is to be expected since these variables did not load on the factor, and the highest coefficient is with to-complement clauses related expressing stance (.635, \( p < .01 \)), which means the to-infinitive feature on the factor is in fact related to stance. As Biber explains:

“Stance expressions can convey many different kinds of personal feelings and assessments, including attitudes that a speaker has about information, how certain they are about its veracity, how they obtained access to the information, and what perspective they are taking.” (Biber, 2006, p.87)

Below are examples of to-clause stance expressions found in the corpus (marked in bold).

Verb-controlled (.528, \( p < .01 \)): coherent enough to allow for the existence of a unified resistance

Adjective-controlled (.334, \( p < .01 \)): Leinster look certain to face an uphill struggle

Noun-controlled (.325, \( p < .01 \)): The East German Government's decision to let the refugees leave …

Suasive verbs were also common (.412, \( p < .01 \)), some of which occurred in to-clauses, as in (marked in bold):

it could no harm to concede a point he had already taken.

There were two instances of numbers (.402, \( p < .01 \)) used as implicit metaphor (in bold):
METAPHOR USE AND REGISTER VARIATION

adjustments you need to make are mental ones

the slip was a telling one

Finally, causative verbs (.351, p<.01) are frequent, as in (marked in bold):

its blankness will slowly dissolve and let life in

These features seem to be connected to a modalized, stance-marked, opinion-based style, which prompted the decision to label this dimension as elaborated opinion.

The plot in Figure 5 shows that the distribution of registers on this dimension is compact, as there is not much difference between register scores.

INSERT FIGURE 5 ABOUT HERE

All 12 top-scoring texts are news, and the sample below is from the highest scoring one (file AHD, 13.6 score):

Small children can sometimes be persuaded to eat cow cake in the same way that old people will eat cat food, but the chief result of all this farming was to produce huge food mountains which we could then refuse to give to the Third World, or to the socialists of Eastern Europe, in order to teach them the error of their ways. (…) The masts may seem unsightly to those unused to them, but I am sure it is simply a matter of getting accustomed.

The sample shows that the salient features are exploited by the writer to elaborate an opinion on farming.

By contrast, texts with low scores on this dimension have fewer of these characteristics. The sample below is from a piece of news about auctions (text a39, -4.3 score), which only includes phrasal coordination as a salient feature.
They are mixing Islamic art from the Middle East with Indian and South-East Asian. (...) Homely Persian pottery of the twelfth to thirteenth century combines animal and bird images with sinuous plant motifs and marks the beginning of lustre wares; they are mostly priced in the 500 to 1,000 range and rate high on charm.

The text provides a description of the items for auction, but the author does not elaborate an opinion on the event.

In view of these results, the answer to question 2 is that four dimensions of variation were identified, two of which replicate Biber's (1988) dimensions. The other two are novel metaphor-related dimensions that included metaphor in different degrees. Dimension 2 comprises most metaphor features, while dimension 4 includes only two metaphor characteristics. Thus, dimension 2 is more representative of the role of metaphor use in register variation than dimension 4.

As far as the role of metaphor in register variation, these dimensions suggest that registers vary with respect to metaphor density and elaborated opinion. Higher density registers tend to be more literate, such as academic writing and news, while lower density ones tend to be more oral, such as conversation and fiction. This corroborates previous research that saw metaphor sparseness in conversation (Berber Sardinha, 2008b; Kaal, 2012).

3.3. Question 3: Significance of register distinctions

Table 15 gives the statistics for indicating whether there are statistical differences among the mean register scores for each dimension. The $F$ and $p$ columns indicate whether the mean register scores were statistically different, and the $R^2$ value shows how much of the variation in the data is captured by the register differences.
METAPHOR USE AND REGISTER VARIATION

The differences across registers on the first three dimensions are significant, which indicates that register scores are statistically different from each other. At the same time, the amount of variation (R\textsuperscript{2}) captured by each factor varies considerably. Factor 1 is the most powerful, accounting for about 84% of the variation across the registers. This is normal, since earlier factors tend to concentrate more variation than later ones. Factor 3 is much less powerful, with just about 55% of variation accounted for, and factor 2 is the least discriminating, accounting for 41% of variation. There are no statistical differences among the registers on dimension 4, though, as could be anticipated given the fact that the four registers have very similar scores on this dimension. Hence, the only metaphor dimension that relates significantly to register is metaphor density. Most of the variation across the registers is not explained by metaphor density alone, as R\textsuperscript{2} is below .5. It must be noted, though, that accounting for 41% of the variation is acceptable in MD analyses; in Biber (1988, p.127), for example, dimension 4 (overt expression of persuasion) had an R\textsuperscript{2} of only .169, and it was considered a valid dimension all the same. The R\textsuperscript{2} statistic is therefore an index of the power to predict register distinctions (or whatever classifications are considered), not strictly of the validity of a dimension. It is noteworthy that metaphor by itself, and specifically its presence in text, can account for so much variation, since it is less frequent than most grammatical features typically used in MD analyses, one might have anticipated a weaker association between register and metaphor.
3.4. Question 4: Cutting across register differences

The scores for each text on a dimension can be compared to each other in order to generate clusters of texts that share similar scores. These text clusters thus bypass the classification based on registers. A cluster analysis was then run on the text scores for each individual dimension, using the Two-Step Cluster Analysis option in SPSS (having log-likelihood as the distance measure and BIC as the clustering criterion). This procedure determines “which solution provides the best ‘fit’ to the data”, answering the question “in which solutions are the texts within in each cluster maximally similar while the clusters themselves are maximally distinct?” (Biber, 1995, p.321). This use of cluster analysis is similar to text type identification in MD research (see next section), where text types are defined as groups of texts sharing similar linguistic characteristics. The clusters extracted here are also linguistically- rather than situationally-based, but the main difference is that in text type analysis, all dimensions are considered at once, whereas here each dimension was considered in turn. Hence, the resulting clusters are not text types, as they do not cut across all dimensions; rather, they are dimension-specific text groupings.

Table 16 gives the results of the cluster analysis.

The cluster analysis suggested either two- or three-cluster solutions for each dimension. Significance tests were conducted for the clusters on each dimension and a comparison was made for the $R^2$ obtained for the cluster solutions and for the register-based analyses reported in question 3. As can be seen, the cluster solutions account for
METAPHOR USE AND REGISTER VARIATION

more variation than registers on all dimensions except the first. The most dramatic increases were for the metaphor dimensions (2 and 4). This implies that metaphor variation is better modeled by these text clusters than by the original register classification. For the 'non-metaphor' dimensions, the differences in $R^2$ are small in comparison to register distinctions, which suggests that for grammar, register categories are as good predictors of variation as text clusters. For metaphor dimensions, the reverse is true: non-register distinctions are more powerful discriminators of metaphor use.

The actual clusters for the metaphor dimensions are shown in Table 17.

INSERT TABLE 17 ABOUT HERE

For dimension 2, the three clusters represent three distinct bands of metaphor density:

- low band: texts have low metaphor density, applies to all registers, but especially conversations (all of them are in this band) and fiction (92%);
- mid band: texts have average metaphor density, applies to most registers, especially academic;
- high band: texts have high metaphor density, restricted to some academic and news texts.

The low and mid-density bands combined account for 91% of all texts. Thus, the unmarked status for English texts seems to be 'low' or 'mid-density'. Conversations are particularly marked for anything other than low density, which again corroborates previous research (Berber Sardinha, 2008b; Kaal, 2012) on metaphor in conversation.
'High density' is typical of very few texts, being therefore a marked characteristic of English discourse.

For dimension 4, the three clusters indicate different levels of opinion elaboration:

- Less elaborated, applies to all registers, typical of conversation (73%);
- Somewhat elaborated, relates to all registers, typical of academic and fiction (67% each);
- Highly elaborated, restricted to a small number of news texts.

In summary, these are the profiles for each register:

- Academic: mid-range metaphor density, somewhat elaborated.
- News: mid to low metaphor density, all levels of opinion elaboration.
- Fiction: low metaphor density, somewhat elaborated.
- Conversation: low metaphor density, low on opinion elaboration.

3.5. Question 5: Text types

Text types are "text categories that are linguistically well defined" (Biber, 1995, p.320). They are derived from the data through cluster analysis, which groups together texts with similar text scores across dimensions. Unlike in the previous section, this time all four dimensions were entered in the analysis, not each one individually. The result is therefore a categorization that represents all dimensions at once. Text types are linguistically cohesive because the dimensions are based on co-occurring linguistic features, not on situational categories defined ahead of time (the registers).

A cluster analysis was run through the Two Step Clustering procedure in SPSS, using BIC as the clustering criterion and Log-Likelihood as the distance measure, as in the previous section. As said before, this procedure determines the quantity of clusters in the data by itself, and in this case, it yielded two clusters, that is, two text types.
Table 18 shows the mean scores on each dimension for each text type, and Figure 6 displays these scores graphically.

Each text type represents about half of the texts in the corpus, and they have diametrically opposing means. The text types can be summarized as follows:

Text type 1: Involved, lower metaphor density, narrative, less elaborated opinion

Text type 2: Informative, higher metaphor density, explicit reference, more elaborated opinion

4. **Concluding remarks**

This study revealed a number of different findings about the relationship between metaphor and register in English, some of which are highlighted below.

The first main finding is the identification of two metaphor dimensions of variation in English, one related to metaphor density, where metaphor is the chief element, and the other to opinion elaboration, where metaphor plays a marginal role, in quantitative terms.

The second main finding is that metaphor accounts for 41% at most of register variation in English. Grammar, on the other hand, captures twice as much variation. Being a much less common feature of language use than grammatical structure, it is striking that metaphor can predict register by such a degree.
The third main finding is that metaphor cannot be accounted for by the existing dimensions for English. Simply knowing how a text scores on each of the main five dimensions of variation in Biber (1988) do not enable us to predict metaphor use. As a result, metaphor must be explicitly factored in in MD analyses, which was done here.

The fourth main finding is that structural variables tend to segregate metaphors, and vice-versa. This indicates that metaphor and grammar are two distinct levels of language that give rise to different frequencies and distributions in text, thus splitting apart onto different dimensions. In other words, although grammar and metaphor are associated, they provide a different perspective each on register variation.

The fifth main finding is the existence of text groupings that better account for metaphor variation than register categories. The chief grouping represents a scale of metaphor density that specifies three bands of density.

And the final main finding relates to the two text types detected, which embody the linguistic characteristics mutually shared across the texts with respect to the dimensions.

As regards further research, it is necessary to ‘unpack’ dimension 2 (metaphor density). This dimension is the most important for register variation but it is actually a composite of different metaphor features that provide a rather crude way of looking at register from a metaphor perspective. By virtue of aggregating many different metaphor characteristics, it simply distinguishes between registers with more metaphor from those with less metaphor. Although this is an important distinction, it is not nuanced enough to capture the various different roles that metaphor plays in language use. The reason why this dimension came about in the first place must have to do with the mutual segregation effect described above. In other words, if metaphor and grammar variables had not been added together to the variable pool, but instead metaphor had been...
considered on its own, the picture emerging might have been much sharper in terms of seeing fine-grained associations between registers and particular metaphor uses (see Herrmann, this volume). Further MD research should therefore consider metaphor features on their own, verify what kinds of dimensions can be detected, and determine to what extent these dimensions can capture more specialized modes of interaction between metaphor and textual categories such as register or text type.

The dimensions unveiled in this chapter must not be considered final, as more research is needed, for instance with a corpus containing a wider variety of registers. It is hoped, though, that a combination of metaphor research and MD analysis may be considered a promising avenue for further research.
AUTHOR NOTE/ACKNOWLEDGMENT

* I'm grateful to CNPq (Brasília, Brazil) for research grants that enabled me to conduct the research reported here. I'm also grateful to the following people: Tina Krennmayr for sharing her early findings and for answering my queries about the VUAMC, Berenike Herrmann for both her feedback on previous versions of this text and generous cooperation throughout, the VUAMC team for making available the corpus on which this research is based, and Doug Biber, for his inspiration and support over the years.
NOTES

1 This differs from the size of the corpus as reported in previous research (e.g. Steen, et al., 2010), primarily because of the different ways in which the CLAWS (the tagger that annotated the BNC) and the Biber tagger handle word segmentation.
References


METAPHOR USE AND REGISTER VARIATION


## Descriptive statistics

### Factor 1

<table>
<thead>
<tr>
<th>register</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>academic</td>
<td>-9.0504</td>
<td>4.17705</td>
<td>15</td>
</tr>
<tr>
<td>conversation</td>
<td>28.4308</td>
<td>6.76348</td>
<td>11</td>
</tr>
<tr>
<td>fiction</td>
<td>4.3029</td>
<td>5.18995</td>
<td>12</td>
</tr>
<tr>
<td>news</td>
<td>-4.9700</td>
<td>5.07909</td>
<td>46</td>
</tr>
</tbody>
</table>

### Factor 2

<table>
<thead>
<tr>
<th>register</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>academic</td>
<td>5.3750</td>
<td>8.67702</td>
<td>15</td>
</tr>
<tr>
<td>conversation</td>
<td>-10.6741</td>
<td>2.17070</td>
<td>11</td>
</tr>
<tr>
<td>fiction</td>
<td>-4.9289</td>
<td>2.63581</td>
<td>12</td>
</tr>
<tr>
<td>news</td>
<td>2.0856</td>
<td>6.10712</td>
<td>46</td>
</tr>
</tbody>
</table>

### Factor 3

<table>
<thead>
<tr>
<th>register</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>academic</td>
<td>4.4288</td>
<td>2.36938</td>
<td>15</td>
</tr>
<tr>
<td>conversation</td>
<td>-1.5376</td>
<td>1.91356</td>
<td>11</td>
</tr>
<tr>
<td>fiction</td>
<td>-4.4485</td>
<td>1.72228</td>
<td>12</td>
</tr>
<tr>
<td>news</td>
<td>0.0840</td>
<td>2.48833</td>
<td>46</td>
</tr>
</tbody>
</table>

### Factor 4
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>academic</td>
<td>-.1091</td>
<td>1.42066</td>
<td>15</td>
</tr>
<tr>
<td>conversation</td>
<td>-1.3696</td>
<td>1.08428</td>
<td>11</td>
</tr>
<tr>
<td>fiction</td>
<td>-.7402</td>
<td>.77542</td>
<td>12</td>
</tr>
<tr>
<td>news</td>
<td>.5562</td>
<td>3.54533</td>
<td>46</td>
</tr>
</tbody>
</table>
### Table 1 Composition of the VUAMC corpus after processing

<table>
<thead>
<tr>
<th>Register</th>
<th>Tokens</th>
<th>Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>68,276</td>
<td>15</td>
</tr>
<tr>
<td>Conversation</td>
<td>48,768</td>
<td>11</td>
</tr>
<tr>
<td>Fiction</td>
<td>45,663</td>
<td>12</td>
</tr>
<tr>
<td>News</td>
<td>46,208</td>
<td>46</td>
</tr>
<tr>
<td>Total</td>
<td>208,915</td>
<td>84</td>
</tr>
<tr>
<td>XML markup</td>
<td>Meaning</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>type=&quot;impl&quot;</td>
<td>Implicit metaphor</td>
<td></td>
</tr>
<tr>
<td>type=&quot;lit&quot;</td>
<td>Explicit metaphor</td>
<td></td>
</tr>
<tr>
<td>type=&quot;met&quot;</td>
<td>Indirect metaphor</td>
<td></td>
</tr>
<tr>
<td>seg function=&quot;mFlag&quot;</td>
<td>Metaphor flag</td>
<td></td>
</tr>
<tr>
<td>subtype=&quot;WIDLII&quot;</td>
<td>Borderline cases</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2 Search strings for metaphor cases*
<table>
<thead>
<tr>
<th>Tokenized text</th>
<th>Metaphor?</th>
<th>3-gram</th>
<th>Contains metaphorically used word?</th>
<th>Match COCA 3-gram list?</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td></td>
<td>The other equation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td>Other equation for</td>
<td></td>
<td></td>
</tr>
<tr>
<td>equation</td>
<td></td>
<td>Equation for the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>for</td>
<td></td>
<td>For the electric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
<td>The electric field</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>electric</td>
<td></td>
<td>Electric field strength*</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>field</td>
<td>Yes</td>
<td>field strength* formula*</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can</td>
<td></td>
<td>Can we do</td>
<td></td>
<td></td>
</tr>
<tr>
<td>we</td>
<td></td>
<td>We do the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>do</td>
<td></td>
<td>do the same</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
<td>the same thing</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>same</td>
<td></td>
<td>same thing for</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>thing</td>
<td>Yes</td>
<td>thing for the</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>for</td>
<td></td>
<td>for the magnetic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
<td>the magnetic quantities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>magnetic</td>
<td></td>
<td>magnetic quantities no*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantities?</td>
<td></td>
<td>magnetic quantities no* but*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* words not in this text sample.

*Table 3 Identifying bundles*
<table>
<thead>
<tr>
<th>Variable type</th>
<th>Features in tag set</th>
<th>Features present in the corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural</td>
<td>401</td>
<td>127</td>
</tr>
<tr>
<td>Semantic</td>
<td>270</td>
<td>36</td>
</tr>
<tr>
<td>Metaphor</td>
<td>1162</td>
<td>123</td>
</tr>
<tr>
<td>Total</td>
<td>1833</td>
<td>250</td>
</tr>
</tbody>
</table>

*Table 4: Total features used in annotating the corpus*
<table>
<thead>
<tr>
<th>Count</th>
<th>Biber (1988) dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Indirect metaphor</td>
<td>- .574</td>
</tr>
<tr>
<td></td>
<td>p=.000</td>
</tr>
<tr>
<td>Direct metaphor</td>
<td>-.047</td>
</tr>
<tr>
<td></td>
<td>NS</td>
</tr>
<tr>
<td>Implicit metaphor</td>
<td>-.110</td>
</tr>
<tr>
<td></td>
<td>NS</td>
</tr>
<tr>
<td>Metaphor flag</td>
<td>-.074</td>
</tr>
<tr>
<td></td>
<td>NS</td>
</tr>
</tbody>
</table>

*Table 5 Correlation coefficients for main metaphor counts and text scores on each of Biber (1988) dimensions*
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>adj_attr</td>
<td>Attributive Adjective</td>
</tr>
<tr>
<td>2</td>
<td>advs</td>
<td>Adverb</td>
</tr>
<tr>
<td>3</td>
<td>agls_psv</td>
<td>Agentless passive verb</td>
</tr>
<tr>
<td>4</td>
<td>all_METIMPL</td>
<td>Implicit metaphors</td>
</tr>
<tr>
<td>5</td>
<td>all_METLIT</td>
<td>Explicit metaphors</td>
</tr>
<tr>
<td>6</td>
<td>all_METMET</td>
<td>Indirect metaphors</td>
</tr>
<tr>
<td>7</td>
<td>be_state</td>
<td>Verb 'Be' (uninflected present tense, verb and auxiliary)</td>
</tr>
<tr>
<td>8</td>
<td>bundles</td>
<td>Metaphor is in a high frequency COCA 3-gram</td>
</tr>
<tr>
<td>9</td>
<td>clusters</td>
<td>Metaphor is part of a cluster</td>
</tr>
<tr>
<td>10</td>
<td>contrac</td>
<td>Contraction</td>
</tr>
<tr>
<td>11</td>
<td>gram_METIMPL_pronoun</td>
<td>Implicit metaphor is a pronoun</td>
</tr>
<tr>
<td>12</td>
<td>gram_METLIT_verb</td>
<td>Direct metaphor is a verb</td>
</tr>
<tr>
<td>13</td>
<td>gram_METMET_adjective</td>
<td>Indirect metaphor is an adjective</td>
</tr>
<tr>
<td>14</td>
<td>gram_METMET_noun</td>
<td>Indirect metaphor is a noun</td>
</tr>
<tr>
<td>15</td>
<td>inf</td>
<td>Infinitive Verb</td>
</tr>
<tr>
<td>16</td>
<td>it</td>
<td>Pronoun 'it'</td>
</tr>
<tr>
<td>17</td>
<td>n</td>
<td>Noun</td>
</tr>
<tr>
<td>18</td>
<td>n_nom</td>
<td>Singular noun –nominalization</td>
</tr>
<tr>
<td>19</td>
<td>p_and</td>
<td>Coordinating conjunction – phrasal connector</td>
</tr>
<tr>
<td>20</td>
<td>pany</td>
<td>Nominal Pronoun (e.g. someone, everything)</td>
</tr>
<tr>
<td>21</td>
<td>pasttnse</td>
<td>Past Tense Verb</td>
</tr>
<tr>
<td>22</td>
<td>pdem</td>
<td>Demonstrative Pronoun</td>
</tr>
<tr>
<td>23</td>
<td>perfects</td>
<td>Verb – Perfect Aspect</td>
</tr>
<tr>
<td>24</td>
<td>pos_mod</td>
<td>Modals of possibility (can, may, might, could)</td>
</tr>
</tbody>
</table>
### Table 6 Variable set

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>prd_mod</td>
<td>Modal of prediction (will, would, shall)</td>
</tr>
<tr>
<td>prep</td>
<td>Preposition</td>
</tr>
<tr>
<td>pres</td>
<td>Verb (uninflected present, imperative &amp; third person)</td>
</tr>
<tr>
<td>pro1</td>
<td>First person pronoun / possessive</td>
</tr>
<tr>
<td>pro2</td>
<td>Second person pronoun / possessive</td>
</tr>
<tr>
<td>pro3</td>
<td>Third person pronoun (except 'it')</td>
</tr>
<tr>
<td>pro_do</td>
<td>Verb 'Do'</td>
</tr>
<tr>
<td>prtcle</td>
<td>Discourse Particle (e.g. now)</td>
</tr>
<tr>
<td>prv_vb</td>
<td>Private Verbs (e.g. believe, feel, think)</td>
</tr>
<tr>
<td>sem_abstractMET</td>
<td>Indirect metaphor is in the &quot;abstract&quot; semantic field</td>
</tr>
<tr>
<td>sem_actionMET</td>
<td>Indirect metaphor is in the &quot;action&quot; semantic field</td>
</tr>
<tr>
<td>sem_featureMET</td>
<td>Indirect metaphor is in the &quot;feature&quot; semantic field</td>
</tr>
<tr>
<td>sem_foodMET</td>
<td>Indirect metaphor is in the &quot;food&quot; semantic field</td>
</tr>
<tr>
<td>sem_perceptionMET</td>
<td>Indirect metaphor is in the &quot;perception&quot; semantic field</td>
</tr>
<tr>
<td>sem_processMET</td>
<td>Indirect metaphor is in the &quot;process&quot; semantic field</td>
</tr>
<tr>
<td>sem_thingMET</td>
<td>Indirect metaphor is in the &quot;thing&quot; semantic field</td>
</tr>
<tr>
<td>wh_ques</td>
<td>Wh-question</td>
</tr>
<tr>
<td>whiz_vbn</td>
<td>Passive postnominal modifier</td>
</tr>
</tbody>
</table>
Figure 1 Scree plot for unrotated solution
<table>
<thead>
<tr>
<th>Feature</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>pres</td>
<td>0.992</td>
</tr>
<tr>
<td>pro_do</td>
<td>0.872</td>
</tr>
<tr>
<td>prtcle</td>
<td>0.854</td>
</tr>
<tr>
<td>pro2</td>
<td>0.852</td>
</tr>
<tr>
<td>it</td>
<td>0.849</td>
</tr>
<tr>
<td>contrac</td>
<td>0.798</td>
</tr>
<tr>
<td>wh_ques</td>
<td>0.781</td>
</tr>
<tr>
<td>pro1</td>
<td>0.714</td>
</tr>
<tr>
<td>prv_vb</td>
<td>0.606</td>
</tr>
<tr>
<td>pdem</td>
<td>0.564</td>
</tr>
<tr>
<td>prd_mod</td>
<td>0.542</td>
</tr>
<tr>
<td>pany</td>
<td>0.488</td>
</tr>
<tr>
<td>pos_mod</td>
<td>0.354</td>
</tr>
<tr>
<td>pasttnse</td>
<td>-0.378</td>
</tr>
<tr>
<td>agls_psv</td>
<td>-0.422</td>
</tr>
<tr>
<td>whiz_vbn</td>
<td>-0.572</td>
</tr>
<tr>
<td>adj_attr</td>
<td>-0.617</td>
</tr>
<tr>
<td>n</td>
<td>-0.801</td>
</tr>
<tr>
<td>prep</td>
<td>-0.816</td>
</tr>
</tbody>
</table>

Table 7 Factor 1 Pattern Matrix
| Dimension 1: Involved versus Informational Production |

<table>
<thead>
<tr>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>-1</td>
</tr>
<tr>
<td>-2</td>
</tr>
<tr>
<td>-3</td>
</tr>
<tr>
<td>-4</td>
</tr>
<tr>
<td>-5</td>
</tr>
<tr>
<td>-6</td>
</tr>
<tr>
<td>-7</td>
</tr>
<tr>
<td>-8</td>
</tr>
<tr>
<td>-9</td>
</tr>
<tr>
<td>-10</td>
</tr>
</tbody>
</table>

*Figure 2 Mean scores for Dimension 1: Involved versus Informational Production*
### Table 8 Factor 2 Pattern Matrix

<table>
<thead>
<tr>
<th>Feature</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>all_METMET clusters</td>
<td>Indirect metaphors 0.936</td>
</tr>
<tr>
<td>gram_METMET_noun</td>
<td>Indirect metaphor is a noun 0.829</td>
</tr>
<tr>
<td>sem_abstract_METMET</td>
<td>Indirect metaphor is in the &quot;abstract&quot; semantic field 0.772</td>
</tr>
<tr>
<td>sem_action_METMET</td>
<td>Indirect metaphor is in the &quot;action&quot; semantic field 0.736</td>
</tr>
<tr>
<td>sem_perception_METMET</td>
<td>Indirect metaphor is in the &quot;perception&quot; semantic field 0.715</td>
</tr>
<tr>
<td>bundles</td>
<td>Metaphor is in a high frequency COCA 3-gram 0.684</td>
</tr>
<tr>
<td>sem_process_METMET</td>
<td>Indirect metaphor is in the &quot;process&quot; semantic field 0.648</td>
</tr>
<tr>
<td>sem_feature_METMET</td>
<td>Indirect metaphor is in the &quot;feature&quot; semantic field 0.569</td>
</tr>
<tr>
<td>gram_METMET_adjective</td>
<td>Indirect metaphor is an adjective 0.532</td>
</tr>
</tbody>
</table>
Table 9 Pearson correlations between Factor 2 and Biber’s (1988) factor scores

<table>
<thead>
<tr>
<th>Biber (1988) Factor</th>
<th>Correlation</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.536</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>-.383</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>.563</td>
<td>.000</td>
</tr>
<tr>
<td>4</td>
<td>.024</td>
<td>NS</td>
</tr>
<tr>
<td>5</td>
<td>.299</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 9 Pearson correlations between Factor 2 and Biber’s (1988) factor scores
### Table 10 Correlations between each factor and total metaphor count

<table>
<thead>
<tr>
<th>Factor</th>
<th>Correlation</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−.577</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>.949</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>.308</td>
<td>.004</td>
</tr>
<tr>
<td>4</td>
<td>.243</td>
<td>.026</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>academic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>news</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fiction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>conversation</td>
</tr>
</tbody>
</table>

*Figure 3 Mean scores for Dimension 2: Metaphor Density*
<table>
<thead>
<tr>
<th>Feature</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>(adj_attr)</td>
<td>Attributive Adjective 0.439</td>
</tr>
<tr>
<td>p_and</td>
<td>Coordinating conjunction – phrasal connector 0.416</td>
</tr>
<tr>
<td>(pdem)</td>
<td>Demonstrative Pronoun 0.405</td>
</tr>
<tr>
<td>n_nom</td>
<td>Singular noun –nominalization 0.391</td>
</tr>
<tr>
<td>(pres)</td>
<td>Verb (uninflected present, imperative &amp; third person) 0.355</td>
</tr>
<tr>
<td>(prv_vb)</td>
<td>Private Verbs (e.g. believe, feel, think) -0.405</td>
</tr>
<tr>
<td>perfects</td>
<td>Verb – Perfect Aspect -0.621</td>
</tr>
<tr>
<td>pro3</td>
<td>Third person pronoun (except ‘it’) -0.795</td>
</tr>
<tr>
<td>pasttnse</td>
<td>Past Tense Verb -0.922</td>
</tr>
</tbody>
</table>

*Table 11 Factor 3 Pattern Matrix*
<table>
<thead>
<tr>
<th>Biber (1988) Factor</th>
<th>Correlation</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.338</td>
<td>.002</td>
</tr>
<tr>
<td>2</td>
<td>-.874</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>.577</td>
<td>.000</td>
</tr>
<tr>
<td>4</td>
<td>.062</td>
<td>NS</td>
</tr>
<tr>
<td>5</td>
<td>.359</td>
<td>.001</td>
</tr>
</tbody>
</table>

*Table 12 Pearson correlations between Factor 3 and Biber’s (1988) factor scores*
Figure 4 Mean scores for Dimension 3: Explicit reference versus narrative concerns
<table>
<thead>
<tr>
<th>Feature</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>gram_METIMPL_pronoun</td>
<td>0.716</td>
</tr>
<tr>
<td>all_METIMPL</td>
<td>0.639</td>
</tr>
<tr>
<td>pos_mod</td>
<td>0.497</td>
</tr>
<tr>
<td>inf</td>
<td>0.491</td>
</tr>
<tr>
<td>(p_and</td>
<td>-0.388</td>
</tr>
</tbody>
</table>

*Table 13 Factor 4 Pattern Matrix*
### Table 14 Pearson correlations between Factor 4 and Biber’s (1988) factor scores

<table>
<thead>
<tr>
<th>Biber (1988) Factor</th>
<th>Correlation</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.043</td>
<td>NS</td>
</tr>
<tr>
<td>2</td>
<td>-.180</td>
<td>NS</td>
</tr>
<tr>
<td>3</td>
<td>.117</td>
<td>NS</td>
</tr>
<tr>
<td>4</td>
<td>.293</td>
<td>.007</td>
</tr>
<tr>
<td>5</td>
<td>.385</td>
<td>.000</td>
</tr>
</tbody>
</table>
Figure 5 Mean scores for Dimension 4: Elaborated opinion
Table 15 Significance tests for the differences among registers

<table>
<thead>
<tr>
<th>Dimension</th>
<th>F</th>
<th>p</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Involved versus informational production</td>
<td>141.8</td>
<td>.000</td>
<td>.836</td>
</tr>
<tr>
<td>2: Metaphor density</td>
<td>20.3</td>
<td>.000</td>
<td>.411</td>
</tr>
<tr>
<td>3: Explicit reference versus narrative concern</td>
<td>34.9</td>
<td>.000</td>
<td>.551</td>
</tr>
<tr>
<td>4: Elaborated opinion</td>
<td>1.8</td>
<td>NS</td>
<td>.029</td>
</tr>
<tr>
<td>Dimension</td>
<td>Qty. of Clusters</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>261.0</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>170.4</td>
<td>.000</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>143.8</td>
<td>.000</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>169.4</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Table 16 Cluster analysis results for individual dimensions*
### Dimension 2: Metaphor density

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>register</th>
<th>N</th>
<th>%</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>academic</td>
<td>2</td>
<td>13%</td>
<td>-8.8</td>
</tr>
<tr>
<td></td>
<td>conversation</td>
<td>11</td>
<td>100%</td>
<td>-10.7</td>
</tr>
<tr>
<td></td>
<td>fiction</td>
<td>11</td>
<td>92%</td>
<td>-5.3</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>16</td>
<td>35%</td>
<td>-3.6</td>
</tr>
<tr>
<td>2</td>
<td>academic</td>
<td>10</td>
<td>67%</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>fiction</td>
<td>1</td>
<td>8%</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>25</td>
<td>54%</td>
<td>3.1</td>
</tr>
<tr>
<td>3</td>
<td>academic</td>
<td>3</td>
<td>20%</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>5</td>
<td>11%</td>
<td>15.3</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>40</td>
<td>48%</td>
<td>-6.3</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>36</td>
<td>43%</td>
<td>3.4</td>
</tr>
</tbody>
</table>

### Dimension 4: Elaborated opinion

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>register</th>
<th>N</th>
<th>%</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>academic</td>
<td>5</td>
<td>33%</td>
<td>-1.8</td>
</tr>
<tr>
<td></td>
<td>conversation</td>
<td>8</td>
<td>73%</td>
<td>-1.9</td>
</tr>
<tr>
<td></td>
<td>fiction</td>
<td>4</td>
<td>33%</td>
<td>-1.7</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>19</td>
<td>41%</td>
<td>-2.2</td>
</tr>
<tr>
<td>2</td>
<td>academic</td>
<td>10</td>
<td>67%</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>conversation</td>
<td>3</td>
<td>27%</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>36</td>
<td>43%</td>
<td>-2.0</td>
</tr>
<tr>
<td></td>
<td>fiction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>21</td>
<td>46%</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>42</td>
<td>50%</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>6</td>
<td>13%</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6</td>
<td>7%</td>
<td>7.8</td>
</tr>
</tbody>
</table>

Table 17 Text clusters for dimensions 2 and 4
### Table 18 Cluster texts and centroids for each text type

<table>
<thead>
<tr>
<th>Text Type</th>
<th>Texts</th>
<th>Dim 1 (+) Involvement / (-) Information</th>
<th>Dim 2 metaphor density</th>
<th>Dim 3 (+) explicit reference / (-) narrative</th>
<th>Dim 4 elaborated opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40 (48%)</td>
<td>6.55</td>
<td>-4.92</td>
<td>-2.29</td>
<td>-1.19</td>
</tr>
<tr>
<td>2</td>
<td>44 (52%)</td>
<td>-5.96</td>
<td>4.48</td>
<td>2.08</td>
<td>1.09</td>
</tr>
</tbody>
</table>
This refers to all the semantic fields that can be assigned to a word.

Since this dimension does not incorporate metaphor, and due to space limitations, no text samples will be offered to illustrate typical texts of each register.

Text samples are chosen based on how well they illustrate a particular dimension. In this case, this sample was chosen because it has very few metaphors and therefore typifies ‘metaphor sparseness’. This particular sample does not claim to represent the register of fiction.

The Bayesian Information Criterion is a model selection criterion that determines the number of clusters that should be formed in the data.