8

Adding Registers to a Previous Multi-Dimensional Analysis

Tony Berber Sardinha, Marcia Veirano Pinto, Cristina Mayer, Maria Carolina Zuppardi, and Carlos Henrique Kauffmann

Introduction

This chapter describes the process of conducting an additive Multi-Dimensional (MD) Analysis, which consists of incorporating new registers into an existing MD investigation. Unlike a "full" MD Analysis, an additive MD Analysis does not reveal any new dimensions of variation, but rather complements the existing dimensions by increasing the number of text varieties (e.g., registers) covered by the previous analysis. The potential to add text varieties is one of the major advantages of the MD framework in that the dimensions can continue to be expanded as they are applied to different varieties. An additive analysis precludes the need for running a factor analysis, which is required in a regular MD Analysis; however, it does not preclude the need to use the same linguistic variables that loaded on the dimensions of interest in the full MD study on which it is going to be based. Furthermore, the additive analysis does not need to include all the dimensions from the previous analysis—the researchers may choose which dimensions to use.

Additive MD analyses have been conducted with both synchronic and diachronic data on a variety of registers, generally by using the dimensions identified by Biber (1988), with a few rare exceptions (see Biber this volume). One important reason for the prevalence of the 1988 dimensions in additive MD analyses is that these dimensions have become stable reference points for register studies. Another reason is that both the Biber TagCount and the Multidimensional Analysis Tagger (MAT) programs provide automatic additive analyses on the 1988 dimensions.

Synchronic additive studies include Conrad (2001), who looked at variation in the language of biology and history texts published in textbooks and journal articles. Her main objective was to describe differences and similarities in the language patterns of these two disciplines in relation to three perspectives: the disciplines themselves, the kind of publication (journal articles and textbooks), and other nonacademic texts. Although she included the five dimensions of English variation (Biber 1988) in her study, she only reported the results obtained for Dimensions 1 (Involved vs.

Informational production) and 2 (Narrative vs. Non-narrative concerns), as she claims that they are sufficient "to illustrate the kind of new insights that multi-dimensional analysis can provide for understanding disciplinary texts" (Conrad 2001, 97). The results showed that, in relation to these two dimensions, history texts were more similar to conversation and popular non-fiction than ecology texts and that research articles were more informationally dense than textbooks (Dimension 1), regardless of the discipline.

Connor-Linton (2001) analyzed the political discourse surrounding the American nuclear arms policy in the United States and in the Soviet Union during the Cold War context of the 1980s. The texts were (portions of) essays and books by four specialists in the arms race—namely, Herman Kahn's *In Thermonuclear War* (1961), Helen Caldicott's *Missile Envy* (1984), Freeman Dyson's *Weapons and Hope* (1984), and Admiral Noel Gayler's essay "The Way Out: A General Nuclear Settlement" (1984). These texts were added to Biber's (1988) Dimensions 1 (Involved vs. Informational production) and 4 (Overt expression of persuasion, which Connor-Linton refers to as "Overt persuasion effort"), in order to assess the presence of involvement and persuasion in these texts as a way to understand how the authors communicate their worldview on the arms race. The results suggest that written nuclear discourse is particularly marked by persuasion (Dimension 4) and that texts like Kahn's, which make repeated use of the pronouns "we/us" and "they/them" (Dimension 1) "require the reader to share his 'particularistic' scope of identification"—that is, it tends to lead to a less critical reading of the arguments presented (Conor-Linton 2001, 93).

Forchini (2012) examined the degree of naturalness in movie dialogues by carrying out two different additive MD analyses using Biber's (1988) MD model as a reference study and comparing the mean scores obtained in them. The first additive MD Analysis used a corpus of eleven movie transcripts. The second used the Longman Spoken American Corpus (LSAC)—a five-million-word corpus containing at least four hours of daily conversations of American speakers chosen to represent gender, age, ethnicity, and education across the United States. The results show that the mean scores of movie transcripts are very close to those of natural conversation on all five dimensions.

Conrad (2014) argued for the usefulness of complementing an additive MD Analysis using qualitative analysis techniques, such as interviews. She investigated the mismatch between engineering students' writing skills and the workplace demands for engineers by adding a corpus of both practitioner registers (reports and site visit observations) and student registers (university class reports, bridge descriptions, and lab reports) to Biber's (1988) dimensions of variation. The results of the additive analysis show that the practitioners' and students' texts differ considerably on Dimensions 3 (Elaborated vs. Situation-Dependent reference) and 5 (which she refers to as Abstract vs. Non-abstract style). On Dimension 3, student's texts are more suggestive of the elaborated side, whereas practitioners' texts are more suggestive of the situation-dependent side; on Dimension 5, students' texts are more abstract than practitioners' texts. The author then interviewed the students, engineering faculty, and practitioners to better understand why the texts scored the way they did on the dimensions. The analysis of the interview data suggested some of the motivations for the scores, such as the idea on the part of the students that writing looks "better" if sentences are longer

and do not include personal pronouns and, on the part of the practitioners, the need to use language that is unambiguous in order to avoid unintentional liability. The study concluded that these different representations of engineering texts are underlying the differences in the dimensional scores and that it is necessary to change the perception of students in relation to what a "good text" is, as it differs from what is perceived as a "good text" by practitioners in the field.

Zuppardo (2013) described variation in the language of aircraft manuals by adding a large corpus of commercial aircraft maintenance manuals from five different manufacturers to the five dimensions of variation in English (Biber 1988). Her results showed that aircraft manuals were marked by abstract/impersonal (Dimension 5) and informational language (Dimensions 1), but not by situation-dependent reference (Dimension 3), narrative (Dimension 2), or persuasive discourse (Dimensions 4). Like Conrad (2014), Zuppardo argued for the possibility of using the results of additive MD analyses in language teaching, as the additive MD profiles offer a linguistic basis for teachers and course developers to prepare materials for English for Specific Purposes classes.

Jonsson (2016) investigated the linguistic characteristics of "conversational writing" by adding a corpus of computer chat and conversation to the 1988 dimensions. The chats were both synchronous (one turn at a time) and super-synchronous (simultaneous turns), and the conversations were from the Santa Barbara Corpus of Spoken American English; although Biber (1988) included conversations, the author argued that the Santa Barbara Corpus provided an "updated reference as regards face-to-face conversations" (Jonsson 2016, 205). Unlike most studies, Jonsson included Dimension 6 (On-line informational elaboration)[1] because real-time elaboration appears to be an important aspect of both computer-mediated conversation and face-to-face conversations. The results show that the MD profile of the Santa Barbara Corpus conversations is very close to the London-Lund corpus conversations in Biber (1988) and that the super-synchronous chats are more "conversation-like" than both the synchronous chats and face-to-face conversations.

Fewer diachronic studies exist in the additive MD literature, but they also generally rely on the 1988 dimensions. An early diachronic study is Biber and Finegan (1988), who investigated a corpus of fiction, essays, and letters covering a period of four centuries, which were added to Dimensions 1 (Involved vs. Informational production), 3 (Explicit vs. Situation-Dependent reference), and 5 (Abstract vs. Non-abstract concerns). This study referred to the original dimensions by letters, so that Dimension 1 was called Dimension A; Dimension 3, Dimension B; and Dimension 5, Dimension C—probably because the numbers of the dimensions were not well-known at the time. Furthermore, the polarity of Dimension 1 (A) was inverted, meaning the original positive pole (Involved production) was turned negative and the original negative pole (Informational production) was turned positive. These inversions were meant "to facilitate register comparisons across dimensions" (Biber et al. 1998, 229) by placing the poles of interest where they make more sense for a particular study. Inverting the order of the poles does not affect the correlations identified by the factor analysis, only the order of the labels. The results suggest that fiction, essays, and letters developed a more oral (Dimension 1), situation-dependent (Dimension 3), and Non-abstract (Dimension 5) style over time.

Atkinson (1992) looked at the evolution of medical research writing by both performing a rhetorical analysis focusing on the broad genre characteristics of the genre and adding a corpus of medical research texts from 1735 to 1985, consisting of case reports, disease reviews, treatment-focused reports, experimental reports, and prepared speeches, to Biber's (1988) five dimensions of variation. The MD profile showed that the register became gradually more informationally dense and abstract/impersonal (Dimensions 1 and 5), but less narrative (Dimension 2) and not marked by explicit or situation-dependent references or the overt expression of persuasion (Dimensions 3 and 4). According to Atkinson (1992, 363), such results suggest that medical research writing was not influenced by paradigm shifts in the field, but evolved gradually over the years.

Biber and Hared (1994) studied the transition to literacy in Somali using the press registers reportage, institutional editorials, letters to the editor, analytical articles, announcements, sports reviews, and stories over a period of sixteen years. They used the three dimensions of variation in Somali—namely, Structural elaboration: Involvement versus Exposition, Lexical elaboration: On-line versus Planned/Integrated production and Argumentative versus Reported presentation of information (Biber and Hared 1992a, b). The results showed that the evolution of language patterns in Somali was greatly influenced by the introduction of written varieties of texts in 1973.

Atkinson (1996) studied the variation in scientific writing from 1675 to 1975 by analyzing a corpus of 202 texts from the Philosophical Transactions of the Royal Society of London. He added these texts to the first five dimensions of variation from Biber (1988). He discovered a major shift from moderately involved language to highly informational language (Dimension 1) accompanied by a distancing of the author from the text (Dimension 5), which in turn led to highly "abstract" texts that displayed a gradual loss of narrative markers (Dimension 2), a tendency toward explicit reference (Dimension 3), and no persuasive tones (Dimension 4).

Biber and Finegan (2001) looked at diachronic relations among personal written registers (journals/diaries, personal letters, fiction prose), specialist written registers (legal opinions, medical prose, news reportage, and scientific prose), and speech-based registers (drama, fiction dialogue, and sermons) taken from the ARCHER corpus over a 340-year period (1650–1990). They added the texts to the dimensions from Biber (1988) and found that personal written registers became more informational (Dimension 1) and impersonal (Dimension 5). The speech-based registers became generally more involved and situation-dependent. The specialist expository registers (legal opinions and medical and scientific prose) followed a trend toward a more literate style—that is, they became more informationally dense (Dimension 1), less narrative (Dimension 2), and more marked by elaborate reference[2] features (Dimension 3) and impersonality (Dimension 5).

Souza (2014) investigated the diachronic variation in a corpus comprising all of the cover stories of *Time* Magazine published between 1923 and 2011. She added the texts to the dimensions from Biber (1988) and found that *Time* cover stories were marked by increasing density of information (Dimension 1), explicit references (Dimension 3), and abstract/impersonal language (Dimension 4).

Veirano Pinto (2014) looked at the verbal language of American movies by adding a large diachronic multi-genre corpus[3] to Biber's (1988) dimensions of variation in English.

The corpus comprised the major movie titles released in the United States from the 1930s to the first decade of the 2000s. The genres included action/adventure, comedy, drama, and horror/suspense/mystery. The results showed that the movie dialogues were highly marked by involvement (Dimension 1), moderately marked by situation-dependent references (Dimension 3) and persuasive language (Dimension 4), and not marked by narrative (Dimension 2) or abstract/impersonal language (Dimension 5). They also revealed that socio-economic aspects, such as stricter/laxer censorship and the greater/lesser financial power of major studios across the years, left linguistic fingerprints that allowed for the identification of transition periods in moviemaking.

Conducting an additive MD Analysis

An MD additive analysis is technically less demanding and usually less time consuming than a full MD Analysis, since it does not require conducting multivariate statistical procedures. Yet it provides a wealth of detail about the linguistic profile of the texts under consideration and affords multiple comparisons with the registers from the previous analyses. The methodological steps needed to conduct an additive MD Analysis are provided below. These are illustrated with reference to an additive analysis of English Web registers onto Biber's (1988) dimensions. For an overview of how to conduct a full MD Analysis, see Egbert and Staples (this volume) and Friginal and Hardy (2014). The major steps needed for the addition of new registers to an existing MD Analysis are as follows:

1. Selecting a study whose MD Analysis will serve as the basis for the additive analysis.
2. Choosing the existing dimension(s) of interest.
3. Identifying the linguistic features included in the existing dimensions.
4. Tagging the corpus for these linguistic features.
5. Computing the normed frequency counts for the linguistic features.
6. Standardizing the normed counts.
7. Computing the dimension scores.
8. Computing the mean dimension scores for the new registers.
9. Comparing the new registers to the ones in the original study.

There are two ways to perform steps 5 through 9 using (1) a spreadsheet program, such as Excel, or (2) a specialized software program, such as the Biber TagCount or the MAT (see Nini this volume). These options are discussed below.

Selecting a study whose MD Analysis will serve as the basis for the additive analysis

The MD Analysis onto which new registers will be added should be selected in accordance with the goals of the study conducted. Researchers should consult the previous literature on full MD analyses to select a study that fits their needs. If specialized

MD Analysis computer software is used to compute the dimension scores, the choice of a reference study will be constrained by the particular reference study that the software is programmed to use. At the time of writing, both the Biber TagCount and the MAT use Biber's (1988) dimensions as a reference; therefore, if either of these programs is used, the reference study will necessarily be Biber (1988). The Biber TagCount computes scores for the first five dimensions whereas the MAT, for the first six. However, if analysts decide to calculate the dimension scores themselves, they should select a study that makes available the mean and standard deviations for the linguistic features that loaded on the dimensions pertinent to the study (see step 6), because these are needed for the calculation of the dimension scores. Fortunately, such data are available in the MD literature for several different studies. For example, the means and standard deviations for English registers are available in Biber (1988, pp. 77–78, 247–69); for Somali, in Biber (1995, 110–11); for Korean, in Biber (1995, 108–9); and for Brazilian Portuguese, in Berber Sardinha, Kauffmann, and Mayer Acunzo (2014, 69–74). If the mean and standard deviations for the reference study of interest were not published, the researcher may try to obtain the data by contacting the authors directly.

Choosing the existing dimension(s) of interest

As previously mentioned, not all dimensions of a full MD Analysis need to be used when performing an additive MD Analysis. Analysts may choose to utilize only the dimensions pertinent to the goals of their study. Sometimes a single dimension may be sufficient to meet a particular research goal. For example, Quaglio (2009) used a single dimension—namely, Dimension 1 (Involved versus Informational production)—to measure the naturalness of the dialogues in the *Friends* sitcom.

Identifying the linguistic features included in the existing dimensions

It is a good idea to make a list of the linguistic features loading on each of the poles of the reference dimensions to start an additive analysis. For each dimension pole, you should not include the features that loaded with a higher loading on a different dimension; these are normally shown in brackets in the literature. For instance, in Dimension 2 from Biber (1988), all of the features on the negative pole are in brackets (see Table 8.1) and, therefore, should not be included in the computation of the dimension scores for that dimension. Sorting out the features will enable you to focus only on those features that will actually be used to compute the additive dimension scores. For example, if the reference analysis for the additive study is Biber (1988) and the registers will be added to the first five dimensions found in that study, the additive MD Analysis will use fifty-four linguistic features out of the sixty-seven features entered in the factor analysis (see Appendix 1 for a list of these features).

Tagging the corpus for these linguistic features

The texts in the corpus should be annotated with features that reflect those identified in the previous step. Preferably, automatic taggers should be used to annotate the texts;

Table 8.1 Linguistic features in Biber (1988) Dimension 2

Narrative concerns features	**Loadings**
Past tense verbs	.90
Third-person pronouns	.73
Perfect aspect verbs	.48
Public verbs	.43
Synthetic negation	.40
Present participial clauses	.39
Non-narrative concerns features	
(present tense verbs	−.47)
(attributive adjectives	−.41)
(past participial WHIZ deletions	−.34)
(word length	−.31)

if possible, the same tagger should be used as in the source MD Analysis. However, if using the same tagger is not possible, then a different tagger may be used, as long as the necessary features are adequately annotated (see Nini this volume). The tagger generally used in MD analyses of English is the Biber tagger, a program first developed by Douglas Biber for his 1988 study (see Biber this volume). Studies of other languages use different taggers; for example, for Brazilian Portuguese, the PALAVRAS parser (Bick 2014) was used. For the study reported in this chapter, the texts were tagged with the current version of the Biber tagger. See Biber (1988, 211–45) for an overview of the linguistic features covered by the tagger and how the tagging algorithm works. Bear in mind that the tagger does not need to identify the features used in the reference MD Analysis directly. In fact, the features used in most reference MD analyses were not identified directly by the tagger, but were instead computed by a post-processing program that reads the tagged texts and combines, modifies, and renames the tags. For most MD analyses, the post-processing program is the TagCount developed by Biber. In its current version, the TagCount program computes the counts of the features used in Biber (1988) and in more recent analyses (e.g., Biber 2006). The MAT (see Nini this volume) also calculates the features in Biber (1988), and the PALAVRAS TagCount provides the features used in the MD Analysis of Brazilian Portuguese (Berber Sardinha et al. 2014).

Computing the normed frequency counts for the linguistic features

After the texts are tagged, it is necessary to count the linguistic features and norm these counts, generally to a rate per 1,000 words. The process of normalizing the frequency counts can be done manually, using a spreadsheet program such as Excel, or by using a computer program such as the Biber TagCount or MAT. Considering the large number of features, the computation of the frequency counts is usually done automatically by customized software developed for the study or by specialized programs like the

TagCount program developed by Biber and the MAT software developed by Nini (this volume).

The texts used in an MD Analysis generally vary in length (i.e., in number of words); therefore, the counts of linguistic features are not directly comparable, as a longer text will have more occurrences of particular features simply because it has more words. Therefore, after the texts are tagged, their raw frequency counts need to be normed (or normalized) so that they can be compared accurately. The most common procedure is to normalize the raw frequency counts to a text length of 1,000 words. In this process, the raw frequency of a given feature is divided by the total number of words in the text and then multiplied by 1,000:

$$\frac{\text{raw frequency of linguistic feature}}{\text{number of tokens in the text}} \times 1{,}000$$

For example, consider a comparison of three texts: A, with 1,000 words; B, with 2,500 words; and C, with 1,500 words. The raw frequency of adverbs in these three texts is: A, 20; B, 60; and C, 20. In this case, the computation of the frequency counts normed to a rate per 1,000 words would be as follows:

Text A: (20 adverbs ÷ 1,000 words) × 1,000 = 20 adverbs per 1,000 words
Text B: (60 adverbs ÷ 2,500 words) × 1,000 = 24 adverbs per 1,000 words
Text C: (20 adverbs ÷ 1,500 words) × 1,000 = 13.3 adverbs per 1,000 words

When performing the calculations using a spreadsheet, the formula above should be used for each linguistic feature in every text. An example of how this can be done is shown in Table 8.2. Column A includes each text; column B, the raw frequency counts of the tag; column C, the number of tokens in the text; and column D, the formula for the normalization.

Biber's TagCount program computes the normed frequencies of the linguistic features present in texts by itself. The output of the program is a text file, shown in Table 8.3. In this file, each text is represented by a record comprising twelve lines. Therefore, the first twelve lines represent the first file, the next twelve lines, the second file, and so on. The normed counts for each particular feature appear in a particular

Table 8.2 Spreadsheet with the formula for the calculation of normed frequency counts

	A	B	C	D
	Filename	**tag1—raw frequency count**	**Number of tokens in the text**	**normed frequency formula: (B/C) × 1000**
1	text_A.txt	20	1,000	=(B2/C2)*1000
2	text_B.txt	60	2,500	=(B3/C3)*1000
3	text_C.txt	20	1,500	=(B4/C4)*1000

Table 8.3 Output from the Biber TagCount program

en_blogs_01.txt.txt 42.3 5.0 782														
5.1	0.0	7.7	49.9	2.6	0.0	0.0	2.6	2.6	3.8	2.6	0.0	0.0	0.0	0.0
0.0	2.6	2.6	3.8	0.0	3.84	79.5	98.5	57.5	20.5	7.7	6.4	1.3	0.0	1.3
0.0	0.0	44.8	5.1	9.0	35.8	9.0	0.0	1.3	0.0	1.3	0.0	3.8	0.0	3.8
0.0	5.1	0.0	0.0	3.8	3.8	9.0	3.8	2.6	1.3	83.1	12.8	1.3	89.5	7.7
1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	1.3	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	1.3	0.0	0.0	0.0	0.0	0.0	11.5	1.3	1.3	14.1	12.8	0.0	3.8	5.1
5.1	2.6	3.8	0.0	1.3	1.3	2.6	12.8	6.4	2.6	0.0	0.0	0.0	0.0	0.0
-20.34		-3.25		-1.41		-5.58		1.72		0.0	0.0	0.0	0.0	0.0
en_blogs_02.txt.txt 38.8 4.6 853														
5.9	1.2	5.9	75.0	8.2	2.3	3.5	4.7	11.7	12.9	1.2	0.0	0.0	2.3	0.0
1.2	2.3	9.4	10.6	0.0	2.33	52.9	107.9	41.0	23.4	12.9	5.9	1.2	1.2	3.5
0.0	3.5	46.9	3.5	3.5	28.1	5.9	3.5	0.0	0.0	1.2	0.0	4.7	5.9	2.3
0.0	4.7	1.2	0.0	2.3	14.1	18.8	8.2	2.3	4.7	72.7	32.8	2.3	110.2	5.9
7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0
0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	1.2	0.0
1.2	0.0	1.2	2.3	1.2	0.0	1.2	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.1	2.3	2.3	8.2	22.3	2.3	7.0	8.2
1.2	2.3	5.9	0.0	0.0	2.3	0.0	30.5	1.2	7.0	0.0	3.5	5.9	0.0	0.0
-6.36		-3.04		4.34		-5.21		2.00		0.0	0.0	0.0	0.0	0.0

section of a line. For example, the normed count for private verbs appears on the second line of each record, between columns 2 and 5 (i.e., between the second and fifth spaces). The format of this file is suitable for processing in SAS, using code similar to the one shown below. In the example, the code refers to a study named *internet*, whose data are in a file called *internet.data*, in a folder called folders/myfolders/internet. You should substitute the details of your project on these lines. The remainder of the code refers to the linguistic features; the contents of each line are identified by a pound sign followed by the number of the line. For instance, line 2 is indicated by "#2," line 3 by "#3," and so on. These lines should not be modified if you are using the output of the (current version of the) Biber TagCount program.

```
DATA internet;
INFILE "/folders/myfolders/internet/internet.data";
input reg $ 1-5 filename $ 6-60 ttr $ 61-65 wrlengh $
66-70 wcount $ 71-75
#2 prv_vb 1-5 that_del 6-10 contrac 11-15 pres 16-20
pro2 21-25 pro_do 26-30 pdem 31-35 gen_emph 36-40 pro1
41-45 it 46-50 be_state 51-55 sub_cos 56-60 prtcle 61-
65 pany 66-70 gen_hdg 71-75
#3 amplifr 1-5 wh_ques 6-10 pos_mod 11-15 o_and 16-20
wh_cl 21-25 finlprep 26-30 n 31-35 prep 36-40 adj_attr
41-45 pasttnse 46-50 pro3 51-55 perfects 56-60 pub_vb
61-65 rel_obj 66-70 rel_subj 71-75

#4 rel_pipe 1-5 p_and 6-10 n_nom 11-15 tm_adv 16-20 pl_
adv 21-25 advs 26-30 inf 31-35 prd_mod 36-40 sua_vb 41-
45 sub_cnd 46-50 nec_mod 51-55 spl_aux 56-60 conjncts
61-65 agls_psv 66-70 by_pasv 71-75

#5 whiz_vbn 1-5 sub_othr 6-10 vcmp 11-15 downtone 16-20
pred_adj 21-25 allmodal 26-30 allconj 31-35 allpasv
36-40 allwh 41-45 allwhrel 46-50 alladj 51-55 allpro
56-60 have 61-65 allverb 66-70 vprogrsv 71-75

#6 that_rel 1-5 jcmp 6-10

#7 nonf_vth 16-20 att_vth 21-25 fact_vth 26-30 lkly_vth
31-35 att_jth 36-40 fact_jth 41-45 lkly_jth 46-50 nfct_
nth 51-55 att_nth 56-60 fct_nth 61-65 lkly_nth 66-70
spch_vto 71-75

#8 mntl_vto 1-5 dsre_vto 6-10 efrt_vto 11-15 prob_vto
16-20 x1_jto 21-25 x2_jto 26-30 x3_jto 31-35 x4_jto
36-40 x5_jto 41-45 all_nto 46-50 nonfadvl 51-55 atadvl
56-60 fctadvl 61-65 lklydvl 66-70

#9 all_vth 1-5 all_jth 6-10 all_nth 11-15 all_th 16-20
all_vto 21-25 all_jto 26-30 all_to 31-35 all_advl 36-40
```

```
#10 act_ipv 1-5 act_tpv 6-10 mentalpv 11-15 commpv 16-
20 occurpv 21-25 copulapv 26-30 aspectpv 31-35 humann
36-40 prcessn 41-45 cognitn 46-50 abstrcn 51-55 concrtn
56-60 tccncrt 61-65 quann 66-70 placen 71-75

#11 groupn 1-5 sizej 6-10 timej 11-15 colorj 16-20
evalj 21-25 relatnj 26-30 topicj 31-35 actv 36-40 commv
41-45 mentalv 46-50 causev 51-55 occurv 56-60 existv
61-65 aspectv 66-70

#12 dim1 1-10 dim2 11-20 dim3 21-30 dim4 31-40 dim5
41-50;
RUN;
```

To use the output of the Biber TagCount program in SPSS, the output file has to be converted to a spreadsheet, with each text included on a single line. The conversion should be carried out by a third-party program, as the Biber TagCount program does not perform the conversion itself. Table 8.3 partially shows the data output of the Biber TagCount program, and Table 8.4 displays these results converted into a spreadsheet format.

Standardizing the normed counts

Standardizing the normed frequencies is a process whereby the frequency counts are scaled as standard deviation units (Biber 1988, 94–95). This process consists of standardizing all frequencies to a mean of 0 and a standard deviation of +1 or −1. In a full (non-additive) MD Analysis, this can be accomplished in most statistical software packages by converting the normed frequencies into z-scores. In SPSS, click *Analyze*, then *Descriptive statistics*, select the variables, check the *Save standardized values as variables*, and click *OK*. This will create new variables in your dataset, whose names begin with a Z, followed by the original name of the variable. In an additive analysis, standardizing the counts in SPSS or SAS into z-scores is not possible because

Table 8.4 Partial spreadsheet with normed frequency counts of Web registers

register	filename	ttr	wrlengh	wcount	prv_vb	that_del	contrac
ency	en_ency_01.txt.txt	30.0	5.0	823	7.3	2.4	2.4
ency	en_ency_02.txt.txt	29.3	4.4	727	8.3	.0	.0
ency	en_ency_03.txt.txt	32.5	5.0	690	2.9	.0	.0
ency	en_ency_04.txt.txt	30.8	5.2	506	4.0	.0	.0
ency	en_ency_05.txt.txt	32.5	4.9	537	9.3	1.9	.0
ency	en_ency_06.txt.txt	30.3	4.9	555	3.6	.0	14.4
ency	en_ency_07.txt.txt	30.5	5.6	765	5.2	1.3	.0
ency	en_ency_08.txt.txt	31.8	5.4	600	.0	.0	.0
ency	en_ency_09.txt.txt	28.8	5.5	508	2.0	.0	.0

the mean and standard deviations come from a separate study. The Biber TagCount and the MAT programs will standardize the normed counts based on the mean and standard deviations from the reference study automatically, but they will not report the standardized counts.

To calculate the standardized counts, you will need the normed frequency counts from your corpus as well as the mean and standard deviations from the previous reference MD Analysis. These data will then be entered into the following formula for each linguistic feature:

$$\frac{\text{normed frequency in the corpus} - \text{mean frequency in the previous study}}{\text{standard deviation of the previous study}}$$

For example, take the feature word length (wrlengh) in the first text of the encyclopedia register (Table 8.4). The normed frequency of this feature is 5. The mean frequency for the same variable in Biber's 1988 study is 4.5, and the standard deviation is .4. Entering these data in the formula, we obtain a standardized count of 1.25:

$$\frac{5 - 4.5}{.4} = 1.25$$

This indicates that the word length for this particular text is 1.25 standard deviations higher than the mean. Table 8.5 shows how to calculate the standardized counts for word length (wrlengh) in a spreadsheet.

Computing the dimension scores

In this part of the analysis, a dimension (or factor) score is computed for each text in your corpus. Only the features with the highest loadings on a given factor should be used for the calculations. Those that load on a particular factor but have a higher loading on a different factor are generally mentioned in parentheses in the MD

Table 8.5 Spreadsheet exemplifying the calculation of standardized counts

	A	B	C	D	E
	Filename	**wrlengh—normed frequency count (x)**	**wrlengh—mean score on '88 (y)**	**wrlengh—standard deviation on '88 (SD)**	**Standardized frequency formula: (x–y)/SD**
1	en_ency_01.txt.txt	5	4.5	.4	=(B2–C2)/D2
2	en_ency_02.txt.txt	4.4	4.5	.4	=(B3–C3)/D3
3	en_ency_03.txt.txt	5	4.5	.4	=(B4–C4)/D4
4	en_ency_04.txt.txt	5.2	4.5	.4	=(B5–C5)/D5
5	en_ency_05.txt.txt	4.9	4.5	.4	=(B6–C6)/D6

literature. These are considered for the interpretation of the dimensions, but are not entered in the calculation of the dimension scores.

To compute the text scores, sum up the standardized frequencies (z-scores) of all the linguistic characteristics loading on the positive pole of the dimensions for the reference study. Then add the linguistic features loading on the negative pole (if it exists) and subtract the summation of the negative pole from the summation of the positive pole (cf. Biber 1988, 95). This procedure applies to both a "full" and an additive MD Analysis, but in a "full" MD Analysis, the factor scores can be calculated directly in the same statistical package where the factor analysis was conducted because the z-scores are available in the dataset. In a full analysis, the factor scores are computed in SPSS, as follows. Click *Transform*, *Compute Variable*, type the name of the factor score variable in the *Target Variable* box, for instance F1 (for the factor 1 scores). Select the z-score for the first variable loading in the positive pole of the factor (remember to select only those variables that had their highest loading on this factor) and drop it in the *Numeric Expression* box. Type a space and a plus sign after it in the *Numeric Expression* box and choose the next z-score, until all variables loading on the positive pole of the factor have been entered in the *Numeric Expression* box. Enclose this expression in parentheses, and type a minus sign after the closing parenthesis. Now repeat these steps for the variables loading on the negative pole. Suppose two variables loaded on the positive pole (var1 and var2), and two loaded on the negative pole (var3 and var4), the expression entered in the *Numeric Expression* box would look like this (without the period at the end): (Zvar1 + Zvar2) – (Zvar3 + Zvar4). Click *OK* to run this procedure; SPSS will create a new variable in your dataset, called F1, with the factor score for each text on factor 1. To illustrate this with a real example, consider English Dimension 3, Explicit versus Situation-Dependent Reference[4] (Table 8.6; Biber 1988, 89).

The factor score for Dimension 3 would require the following calculation:

> 1988 Dim. 3 score = (standardized frequency of *wh* relative clauses on object positions + standardized frequency of pied-piping constructions + standardized frequency of *wh* relative clauses on subject positions + standardized frequency of phrasal coordination + standardized frequency of nominalizations) – (standardized frequency of time adverbials + standardized frequency of place adverbials + standardized frequency of adverbs)

Table 8.6 Dimension 3: Explicit versus Situation-Dependent Reference (Biber 1988)

Positive pole		**Negative pole**	
wh relative clauses on object positions	.63	time adverbials	−.60
pied-piping constructions	.61	place adverbials	−.49
wh relative clauses on subject positions	.45	adverbs	−.46
phrasal coordination	.36		
nominalizations	.36		

	A	B	C	D	E	F
1	Text	Standardized var1	Standardized var2	Standardized var3	Standardized var4	F1
2	1	1.5	1.1	1.8	0.2	=(B2+C2)-(D2+E2)

Figure 8.1 Calculation of a factor score in an Excel spreadsheet.

In an additive analysis, you should use a spreadsheet; preferably the same spreadsheet where the standardized scores were calculated. Again, suppose two variables loaded on the positive pole (var1 and var2) and two on the negative pole (var3 and var4) of factor 1 (where they had their highest scores across all factors), an Excel spreadsheet would calculate this factor score as shown in Figure 8.1.

Computing the mean dimension scores for the new registers

The mean dimension score for a register (or some other corpus section) being added to the reference dimensions is computed by calculating the arithmetic mean of factor scores calculated in the previous step. This is accomplished by summing up the scores for the texts and dividing this sum by the number of texts. These mean dimension scores can then be plotted on the mean dimension graphs reported in the reference study. Table 8.7 provides an example of the calculation of the mean dimension scores for the register encyclopedia. The scores for all the texts were summed and then divided by the number of texts. In an Excel spreadsheet, the formula = SUM(C2:C21) was entered into cell C22, and the formula = SUM(C2:C21)/20 into cell D22. The mean dimension score for this register is therefore 5.41.

Comparing the new registers to the ones in the original study

The final step of an additive MD Analysis is to compare the registers added to the original ones. This step consists of basically two sub-steps: calculating the mean dimension scores for the "new" registers and illustrating the linguistic characteristics of the "new" registers with examples. To illustrate both of these sub-steps, we present an additive analysis whereby selected Web registers were added to Dimension 3, Explicit versus Situation-Dependent Reference (Biber 1988).

The corpus used in this example comprises five English registers from the Web—namely, encyclopedia, frequently asked questions (FAQ), forums, questions–answers, and terms of use. Online encyclopedias are written collaboratively by users, such as Wikipedia. FAQs are lists of the most common questions asked by users, along with answers provided by the webmaster. Internet forums are sites where users may ask questions about and provide answers to a range of topics. Question-and-answer sites are less restrictive than forums as they are generally not organized into fixed topics or follow strict rules about what can be discussed. Terms of use are a set of rules that users must accept or a disclaimer that users must be aware of in order to use a particular website. Table 8.8 shows the design of the Web corpus.

In order to compare the registers on the Web corpus to each other and to the registers on the previous dimensions (e.g., those found by Biber 1988), it is necessary to compute the mean dimension scores for each of these registers (see Appendix 2). The mean dimension score is calculated by summing up the dimension scores for each

Table 8.7 Spreadsheet with the calculation of the mean dimension scores for the register encyclopedia for Dimension 3 (Biber 1988)

Register	Filename	Dimension 3	
ency	en_ency_01.txt.txt	3.74	
ency	en_ency_02.txt.txt	4.31	
ency	en_ency_03.txt.txt	7.68	
ency	en_ency_04.txt.txt	7.18	
ency	en_ency_05.txt.txt	1.60	
ency	en_ency_06.txt.txt	4.78	
ency	en_ency_07.txt.txt	4.75	
ency	en_ency_08.txt.txt	9.23	
ency	en_ency_09.txt.txt	8.31	
ency	en_ency_10.txt.txt	8.03	
ency	en_ency_11.txt.txt	1.14	
ency	en_ency_12.txt.txt	6.05	
ency	en_ency_13.txt.txt	2.88	
ency	en_ency_14.txt.txt	3.16	
ency	en_ency_15.txt.txt	6.52	
ency	en_ency_16.txt.txt	8.44	
ency	en_ency_17.txt.txt	7.80	
ency	en_ency_18.txt.txt	8.13	
ency	en_ency_19.txt.txt	−1.81	
ency	en_ency_20.txt.txt	6.22	
		108.14	5.41

Table 8.8 Design of the Web corpus

Register	Abbreviation	Count of texts	Count of tokens
Encyclopedia	ency	20	13,569
FAQs	faq	20	11,272
Forums	foru	20	12,014
Questions–answers	queans	20	9,059
Terms of use	tou	20	13,493
Total		100	59,407

text and dividing them by the number of texts in each register. A mean for the entire corpus can also be calculated by summing up all of the individual dimension scores and dividing the result by the number of texts in the corpus. The mean dimension scores are best represented in charts; vertical or horizontal charts (line, bar, or box plots) can be used to illustrate the differences and similarities among the registers graphically. Figure 8.2 shows a vertical plot commonly used in MD analyses, and Figure 8.3 illustrates a horizontal plot. In both, the Web registers are printed in uppercase.

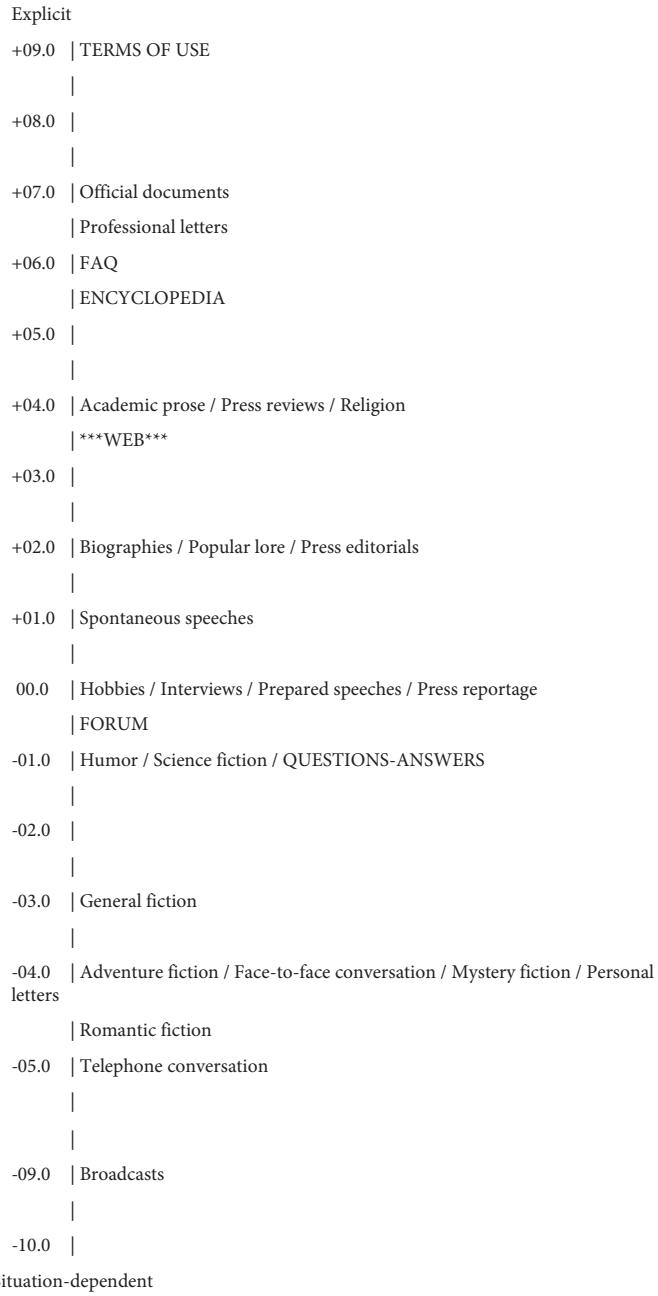

Figure 8.2 Vertical plot of Web registers added to Biber's (1988) Dimension 3, Explicit versus Situation-Dependent reference.

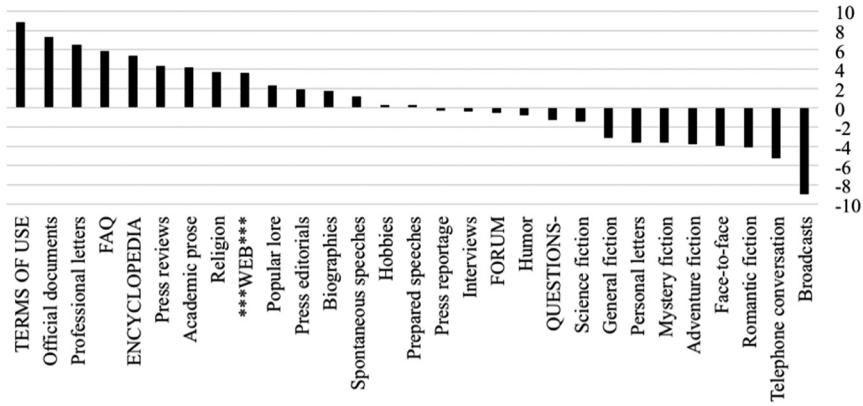

Figure 8.3 Horizontal bar plot of Web registers added to Biber's (1988) Dimension 3, Explicit versus Situation-Dependent reference.

The mean dimension score for the whole of the Web register corpus is on the positive pole of Dimension 3, which means that overall the Web is marked by explicit reference. However, the individual Web registers appear in different places on the dimension, suggesting that the mean for the corpus is hiding significant differences across the different registers in the corpus. This shows the risks involved in calculating a dimension mean that cuts across register categories.

Finally, the comparisons should include examples to illustrate the major linguistic characteristics marking the dimension. The sample in Example 1 (dimension score 8.9) depicts the following characteristics of explicit reference in an excerpt of terms of use: *wh* relative clauses, nominalizations, and phrasal coordination.

1. *Furthermore, it is a violation of this Agreement to use the services of another company for the purpose of facilitating any of the activities which violate this Agreement. . . . Working relationships discussed in this material do not necessarily represent a reporting connection, but may reflect a functional guidance, stewardship, or service relationship. Where shareholder consideration of a local entity matter is contemplated by this material* (tou_16).

Example 2 (question and answers, dimension score −1.2) illustrates Situation-Dependent reference through the use of time adverbials (right now) and adverbs (really, fairly).

2. *Does cosmetic tattoo removal hurt? . . . It's not that bad, I'm in process of laser removal of one of mine right now, in a fairly delicate area, it does hurt more than the tattooing process but if you really hate it, it's livable* (queans_01).

Conclusion

The MD framework is a flexible method for the study of linguistic variation because the dimensions are open constructs to which more data can be added without the need

to repeat the full analysis. Additive MD Analysis exploits this open-ended nature of the MD framework by including registers in the dimensions. The addition of new registers enriches both the source and the additive analysis. The source analysis increases in scope and, therefore, gains more generalizability; the additive analysis also gains by drawing on powerful reference points for register analysis. Thus, an additive MD Analysis is a powerful tool for the analysis of register variations that more researchers should use as an end in itself. In addition, it can also be a point of entry into the MD framework, helping researchers familiarize themselves with the concepts and the fundamentals of the analysis, especially the qualitative, interpretive side of MD Analysis. As it does not require conducting a multivariate statistical analysis, which has often been pointed out as a major difficulty in conducting MD Analysis, an additive analysis can be a first step into the method.

Notes

1 In Biber (1988), the factor interpretation "on-line elaboration" means "real-time communication." Jonsson appears to have used it as "on the Web/Internet."
2 In later studies, this was named explicit reference.
3 See Veirano Pinto (this volume) for the corpus design.
4 The polarity of the dimension can be inverted. In this example, the polarity as shown in Biber (1988) is maintained, but in Nini (this volume) it was inverted.

References

Atkinson, D. (1992), "The Evolution of Medical Research Writing from 1735 to 1985: The Case of the 'Edinburgh Medical Journal'," *Applied Linguistics*, 13: 337–74.

Atkinson, D. (1996), "The Philosophical Transactions of the Royal Society of London, 1675–1975: A Sociohistorical Discourse Analysis," *Language in Society*, 25: 333–71.

Berber Sardinha, T., C. Kauffmann and C. Mayer Acunzo (2014), "A Multi-Dimensional Analysis of Register Variation in Brazilian Portuguese," *Corpora*, 9 (2): 239–71.

Biber, D. (1988), *Variation across Speech and Writing*, Cambridge: Cambridge University Press.

Biber, D. (1995), *Dimensions of Register Variation: A Cross-Linguistic Comparison*, Cambridge: Cambridge University Press.

Biber, D. (2006), *University Language*, Amsterdam/Philadelphia, PA: John Benjamins.

Biber, D., S. M. Conrad and R. Reppen (1998), *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press.

Biber, D. and E. Finegan (1988), "Drift in Three English Genres from the 18th to the 20th Centuries: A Multidimensional Approach," in M. Kytö, O. Ihalainen and M. Rissanen (eds), *Corpus Linguistics, Hard and Soft: Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora*, 83–101, Amsterdam: Rodopi.

Biber, D. and E. Finegan (2001), "Diachronic Relations among Speech-Based and Written Registers in English," in S. M. Conrad and D. Biber (eds), *Variation in English: Multi-dimensional Studies*, 66–83, Harlow: Longman.

Biber, D. and M. Hared (1992a), "Dimensions of Register Variation in Somali," *Language Variation and Change*, 4 (1): 41–75.
Biber, D. and M. Hared (1992b), "Literacy in Somali: Linguistic Consequences," *Annual Review of Applied Linguistics*, 12: 260–82.
Biber, D. and M. Hared (1994), "Linguistic Correlates of the Transition to Literacy in Somali: Language Adaptation in Six Press Registers," in D. Biber and E. Finegan (eds), *Sociolinguistic perspectives on register*, 182–216, New York: Oxford University Press.
Bick, E. (2014), "PALAVRAS, a Constraint Grammar-Based Parsing System for Portuguese," in T. Berber Sardinha and T. S. B. Ferreira (eds), *Working with Portuguese corpora*, 279–302, London: Bloomsbury.
Caldicott, H. (1984), *Missile Envy: The Arms Race and the Nuclear War*, New York: William Morrow.
Connor-Linton, J. (2001), "Author's Style and Worldview: A Comparison of Texts about Nuclear Arms Policy," in S. M. Conrad and D. Biber (eds), *Variation in English: Multi-dimensional Studies*, 84–93, Harlow: Longman.
Conrad, S. M. (2001), "Variation among Disciplinary Texts: A Comparison of Textbooks and Journal Articles in Biology and History," in S. M. Conrad and D. Biber (eds), *Variation in English: Multi-dimensional Studies*, 94–107, Harlow: Longman.
Conrad, S. M. (2014), "Expanding Multi-Dimensional Analysis with Qualitative Research Techniques," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 344–411, Amsterdam/Philadelphia, PA: John Benjamins.
Dyson, F. (1984), *Weapons and Hope*, New York: Harper and Row.
Forchini, P. (2012), *Movie Language Revisited: Evidence from Multi-Dimensional Analysis and Corpora*, Bern: Peter Lang.
Friginal, E. and A. J. Hardy (2014), "Conducting Multi-Dimensional Analysis Using SPSS," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 297–316, Amsterdam/Philadelphia, PA: John Benjamins.
Gayler, N. (1984), "The Way Out: A general Nuclear Settlement," in G. Prins (ed), *The Nuclear Crisis Reader*, 234–43, New York: Vintage.
Jonsson, E. (2016), *Conversational Writing: A Multi-Dimensional Study of Synchronous and Supersynchronous Computer-Mediated Communication*, Frankfurt am Main/New York: Peter Lang.
Kahn, H. (1961), *On Thermonuclear War*, New Jersey: Princeton University Press.
Quaglio, P. (2009), *Television Dialogue: The Sitcom Friends vs. Natural Conversation*, Amsterdam/Philadelphia, PA: John Benjamins.
Souza, R. C. (2014), "Dimensions of Variation in *TIME* Magazine," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis 25, Years On: A Tribute to Douglas Biber*, 177–93, Amsterdam/Philadelphia, PA: John Benjamins.
Veirano Pinto, M. (2014), "Dimensions of Variation in North American Movies," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 109–47, Amsterdam/Philadelphia, PA: John Benjamins.
Zuppardo, M. C. (2103), "A linguagem da aviação: Um estudo de manuais aeronáuticos baseado na Análise Multidimensional" [Aviation Language: A Study of Aeronautical Handbooks Based on Multi-Dimensional Analysis], *Revista Virtual de Estudos da Linguagem*, 11: 6–25. Retrieved January 17, 2018, from http://www.revel.inf.br/files/4b416887c9e8c51b14c95daccc4f39d5.pdf.

Appendix 1

Table 1 Descriptive statistics of the 1988 features

Linguistic feature	Mean	Standard deviation
Past tense	40.1	30.4
Perfect aspect verbs	8.6	5.2
Present tense	77.7	34.3
Place adverbials	3.1	3.4
Time adverbials	5.2	3.5
First person pronouns	27.2	26.1
Second person pronouns	9.9	13.8
Third person pronouns	29.9	22.5
Pronoun IT	10.3	7.1
Demonstrative pronouns	4.6	4.8
Indefinite pronouns	1.4	2.0
DO as proverb	3.0	3.5
Wh questions	.2	.6
Nominalizations	19.9	14.4
Gerunds	7.0	3.8
Nouns	180.5	35.6
Agentless passives	9.6	6.6
BY passives	.8	1.3
BE as main verb	28.3	9.5
Existential THERE	2.2	1.8
THAT verb complements	3.3	2.9
THAT adjective complements	.3	.6
Wh clauses	.6	1.0
Infinitives	14.9	5.6
Present participial clauses	1.0	1.7
Past participial clauses	.1	.4
Past participial WHIZ deletions	2.5	3.1
Present participial WHIZ deletions	1.6	1.8
THAT relatives: Subject position	.4	.8
THAT relatives: Object position	.8	1.1
Wh relatives: Subject position	2.1	2.0
Wh relatives: Object position	1.4	1.7
Wh relatives: Pied pipes	.7	1.1
Sentence relatives	.1	.4
Adverbial subordinator—cause	1.1	1.7
Adverbial subordinator—concession	.5	.8
Adverbial subordinator—condition	2.5	2.2

Table 1 (Continued)

Linguistic feature	Mean	Standard deviation
Adverbial subordinator—other	1.0	1.1
Prepositions	110.5	25.4
Attributive adjectives	60.7	18.8
Predicative adjectives	4.7	2.6
Adverbs	65.6	17.6
Type/token ration	51.1	5.2
Word length	4.5	0.4
Conjuncts	1.2	1.6
Downtoners	2.0	1.6
Hedges	.6	1.3
Amplifiers	2.7	2.6
Emphatics	6.3	4.2
Discourse particles	1.2	2.3
Demonstratives	9.9	4.2
Possibility modals	5.8	3.5
Necessity modals	2.1	2.1
Predictive modals	5.6	4.2
Public verbs	7.7	5.4
Private verbs	18.0	10.4
Suasive verbs	2.9	3.1
SEEM/APPEAR	.8	1.0
Contractions	13.5	18.6
THAT deletion	3.1	4.1
Stranded prepositions	2.0	2.7
Split infinitives	.0	.0
Split auxiliaries	5.5	2.5
Phrasal coordination	3.4	2.7
Non-phrasal coordination	4.5	4.8
Synthetic negation	1.7	1.6
Analytic negation	8.5	6.1

Appendix 2

Table 2 Descriptive statistics of the 1988 dimensions

	Biber (1988) Dimensions									
	Dim. 1		Dim. 2		Dim. 3		Dim. 4		Dim. 5	
Register	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Biographies	−12.4	7.5	2.1	2.5	1.7	3.5	−.7	1.6	−.5	2.5
Personal letters	19.5	5.4	.3	1.0	−3.6	1.8	1.5	2.6	−2.8	1.9
Professional letters	−3.9	13.7	−2.2	3.5	6.5	4.2	3.5	4.7	.4	2.4
Face-to-face conversations	35.3	9.1	−.6	2.0	−3.9	2.1	−.3	2.4	−3.2	1.1
Telephone conversations	37.2	9.9	−2.1	2.2	−5.2	2.9	.6	3.6	−3.7	1.2
Popular lore	−9.3	11.3	−.1	3.7	2.3	3.5	−.3	4.8	.1	2.3
Official documents	−18.1	4.8	−2.9	1.2	7.3	3.6	−.2	4.1	4.7	2.4
Press editorials	−10.0	3.8	−.8	1.4	1.9	2.0	3.1	3.2	.3	2.0
Interviews	17.1	10.7	−1.1	2.1	−.4	4	1.0	2.4	−2.0	1.3
Science fiction	−6.1	4.6	5.9	2.5	−1.4	3.7	−.7	1.7	−2.5	.8
Adventure fiction	−.0	6.3	5.5	2.7	−3.8	1.7	−1.2	2.8	−2.5	1.2
Mystery fiction	−.2	8.5	6.0	3.0	−3.6	3.4	−.7	3.3	−2.8	1.2
General fiction	−.8	9.2	5.9	3.2	−3.1	2.3	.9	2.6	−2.5	1.6
Romantic fiction	4.3	5.6	7.2	2.8	−4.1	1.6	1.8	2.7	−3.1	.9
Humor	−7.8	6.7	.9	1.8	−.8	2.6	−.3	2.7	−.4	1.4
Spontaneous speeches	18.2	12.3	1.3	3.6	1.2	4.3	.3	4.4	−2.6	1.7
Prepared speeches	2.2	6.7	.7	3.3	.3	3.6	.4	4.1	−1.9	1.4
Hobbies	−10.1	5.0	−2.9	1.9	.3	3.6	1.7	4.6	1.2	4.2
Academic prose	−14.9	6.0	−2.6	2.3	4.2	3.6	−.5	4.7	5.5	4.8
Broadcasts	−4.3	10.7	−3.3	1.2	−9.0	4.4	−4.4	2.0	−1.7	2.8
Religion	−7.0	8.3	−.7	2.7	3.7	3.3	.2	2.7	1.4	2.4
Press reportage	−15.1	4.5	.4	2.1	−.3	2.9	−.7	2.6	.6	2.4
Press reviews	−13.9	3.9	−1.6	1.9	4.3	3.7	−2.8	2.0	.8	2.1