Metadata of the chapter that will be visualized online

ChapterTitle	Variation Among Blogs: A Multi-Dimensional Analysis	
Chapter Sub-Title		
Chapter CopyRight - Year	Springer Science+Business Media B.V. 2010 (This will be the copyright line in the final PDF)	
Book Name	Genres on the Web	
Corresponding Author	Family Name	Grieve
	Particle	
	Given Name	Jack
	Suffix	
	Division	
	Organization	University of Leuven, QLVL Research Unit
	Address	Leuven, Belgium
	Email	Jack.Grieve@arts.kuleuven.be
Author	Family Name	Biber
	Particle	
	Given Name	Douglas
	Suffix	
	Division	English Department
	Organization	Northern Arizona University
	Address	Flagstaff, AZ, USA
	Email	douglas.biber@nau.edu
Author	Family Name	Friginal
	Particle	
	Given Name	Eric
	Suffix	
	Division	Department of Applied Linguistics and English as a Second Language
	Organization	Georgia State University
	Address	Atlanta, GA, USA
	Email	efriginal@gsu.edu
Author	Family Name	Nekrasova
	Particle	
	Given Name	Tatiana
	Suffix	
	Division	English Department
	Organization	Northern Arizona University
	Address	Flagstaff, AZ, USA
	Email	Tatiana.Nekrasova@nau.edu
Abstract	A blog, short for a weblog, is a website containing an archive of regularly updated online postings. The postings are generally made by one person and presented in reverse chronological order. The archive is generally made freely available to the public. The postings tend to consist primarily of raw text, but may also contain hyperlinks and other media, including picture, video and sound files. Often blogs allow for readers to post comments as well.	

Chapter 14

Variation Among Blogs: A Multi-Dimensional Analysis

Jack Grieve, Douglas Biber, Eric Friginal, and Tatiana Nekrasova

14.1 Introduction

A blog, short for a weblog, is a website containing an archive of regularly updated online postings. The postings are generally made by one person and presented in reverse chronological order. The archive is generally made freely available to the public. The postings tend to consist primarily of raw text, but may also contain hyperlinks and other media, including picture, video and sound files. Often blogs allow for readers to post comments as well. In terms of content, blogs appear to fall into one of two major types: personal blogs in which an author discusses their own life and thematic blogs in which an author discusses a topic other than themselves. Popular subjects for thematic blogs include current events, politics, arts, entertainment, sports and technology, though in principle any topic is permissible.

Excluding newsgroups, online archives of an individual's postings began to appear in the mid-1990s with two different types of websites: online diaries and commentary pages. Two of the earliest online-diaries are Claudio Pinhanez's diary, which was hosted by the MIT Media Lab website from 1994 until 1996, and Justin Hall's diary, which was updated regularly for 11 years, starting in 1994 [12]. Online diaries naturally conform to many of the key features now associated with blogs: they are regularly-updated and chronologically-ordered online archives of postings written by a single person. These same features also characterize portions of some mid-1990s websites that posted commentary on news and gaming, such as the website *Blue's News*, which featured Stephen Heaslip's commentary on the video game *Quake* starting in 1995. In these websites the postings of experts were logged and archived in a similar manner to online diaries. Over time it appears that these two forms of online writing merged to form the blog variety.

In the late 1990s, numerous factors contributed to a rise in the popularity of blogs. Free easy-to-use blog publishing software was made available, allowing people unfamiliar with web-design to create blogs. *Open diary*, first offered in 1998, was

J. Grieve (✉)

University of Leuven, QLVL Research Unit, Leuven, Belgium
e-mail: Jack.Grieve@arts.kuleuven.be

the first of these programs. Similar programs, such as *Live Journal* and *Diaryland*, became available in 1999. Blog hosting services were also introduced in 1999, most notably blogger.com, which would eventually be purchased by Google in 2003. These services became popular because they made it easy for potential bloggers to post their blogs and for blogs to be found by potential readers.

At the beginning of the millennium, blogging became an even more important influence on modern culture when numerous political blogs became major web destinations. Two of the earliest political blogs were Bob Somerby's *Daily Howler* and Mickey Kaus' *Kausfiles*, which were launched in 1998 and 1999. But it was not until the early 2000s that major news stories were actually being broken on the blogs. One of the first these of stories was the Trent Lott scandal in 2002. At a birthday party for Strom Thurmond, Lott implied that the United States would be better country today if Thurmond had won the presidency in 1948 on his platform of racial segregation. Mainstream media ignored this speech, but the political bloggers attacked Lott vigorously, bringing the issue to the attention of the public.

It is also during this time that the term *blog* first came to be used. The term is a shortened form of *weblog*, which was coined by Jorn Barger in 1997 to describe the list of web links he had assembled on his website *Robot Wisdom* [25]. The term *blog* was formed in 1999, when Peter Merholz split *weblog* into *we blog* in a play on words (The Economist, 2006). According to the Oxford English Dictionary, the first time this word appeared online was in May of 1999 on Brad Graham's weblog (www.bradlands.com), in a posting referencing Merholz: "Cam points out lemonyellow.com and PeterMe decides the proper way to say 'weblog' is 'wee'-'blog' (Tee-hee!)." Five days later, Peter Merholz used the word in context on his weblog (www.peterme.com): "For those keeping score on blog commentary from outside the blog community." By 2002, *blog* was voted as the new word most likely to succeed by the American Dialect Society.

While blogging has thus become an important, common and widely recognized variety of online language (the blog search engine Technorati.com currently searches over 100 million blogs), relatively little is known about its linguistic properties. There have been numerous studies of blogs that have taken a non-empirical and non-linguistic approach, focusing on such topics as the social import of blogging (e.g. [9–11, 18, 19, 22]), the content analysis of blogs [14–16, 21], and the rhetorical analysis of blogs [20]. Very few studies, however, have analyzed the linguistic properties of this variety of language. The major exceptions are Herring and Paolillo [13], which attempts to use selected linguistic features (the relative frequency of pronouns, determiners, and certain function words; cf. [1]) to analyze gender differences in blogs, and Pushmann [23], which examines variation in expressions of futurity in blogs.

There has also been research devoted to identifying the different sub-types of blogs. The main distinction that has been made in past research (e.g. [14, 17]) is between personal blogs and thematic blogs, as defined at the beginning of this section. Krishnamurthy [17], for example, identifies the personal vs. thematic dimension as the most important classifier of blog types. Krishnamurthy [17] also claims that whether a blog is a product of an individual person or a group of people is the

14 Variation Among Blogs: A Multi-Dimensional Analysis

second major determinant of blog type; however, as Herring et al. [14] note, group blogs are relatively rare and are probably best analyzed as distinct individual blogs. In addition, some researchers [13, 14] further divide thematic blogs into two categories: *filter blogs*, which contain a blogger's comments on newspaper articles and postings from other web sites that are either linked to or reproduced in the blog itself; and *k-logs*, short for *knowledge-logs*, which are essentially informational web-sites on a particular topic written by an expert. Pushmann [24] also identifies a *corporate blog* type, where corporations "blog" about their products, though corporate blogs seem to be best classified as a type of thematic blogs, given their narrow focus. While insightful, none of these classifications are based on linguistic data.

While blogs have thus been the subject of previous research, very little is currently known about the overall linguistic characteristics of blogs or the linguistically defined sub-types of blogs. Given the importance of this variety, a linguistic analysis of blogs is clearly of descriptive value. But because the blog is a new variety of language that is still under development the results of this study could be the basis for future diachronic linguistic investigations as well. In addition, the study of linguistic variation in blogs has clear applications in the domain of web-searching: blog search engines could be improved by linguistic-based searches.

The goals of the present study are to identify the principal dimensions of linguistic variation in the blog variety of the English language and to use these dimensions to identify the primary sub-types or *text types* of the blog variety. In order to answer these research questions a 2 million word corpus of blogs was compiled. The values of numerous functional linguistic features were measured for each blog in the corpus and subjected to a factor analysis in order to identify the principal dimensions of linguistic variation for the blog variety. The dimensions produced by the factor analysis were then used as predictors in a cluster analysis, in order to identify the primary blog text types. We hypothesize that these text types will correspond to the categories of personal and thematic blogs.

14.2 Corpus Compilation and Analysis

The corpus compiled for this study represents the modern (2003–2005) American blog variety of the English language. The corpus contains 500 sub-corpora each containing text from a single blog, usually spanning numerous postings. These 500 sub-corpora are the basic unit of analysis in this study. Each sub-corpus contains between 9,864 and 1,099 words. On average, each sub-corpus contains 4,500 words. In total, the corpus contains 2,261,520 words.

The corpus was compiled using globeofblogs.com as an index to locate blogs. Blogs were selected for inclusion in the corpus based on three criteria. First, 10 blogs were selected from each state, using information which was provided by globeofblogs.com. Second, blogs were selected so as to obtain as even a distribution as possible across age and gender; as such, many blogs were excluded when their authors did not provide demographic information. These first two criteria were put in

place because originally this corpus was used to investigate social dialect differences within blogs, but demographic differences are not considered here. Third, blogs were selected that contained as much text as possible. Topic was not controlled.

Once a blog was identified, its contents were then manually downloaded by copying and pasting from the internet browser into text files. In particular, blog posts were copied, while all other text included on the blog websites – including navigation panels, headers and boilerplate text – was excluded from the corpus. In some cases the extraneous text was avoided simply by not copying the extraneous text; in other cases, the extraneous text was deleted from the text files by searching and replacing.

Once compiled, the blog corpus was automatically tagged using the Biber grammatical tagger [4].¹ The current version of this tagger incorporates the corpus-based research carried out for the *Longman Grammar of Spoken and Written English* [8]. The tagger identifies a wide range of grammatical features, including word classes (e.g. nouns, modal verbs, prepositions), syntactic constructions (e.g. WH relative clauses, conditional adverbial clauses, *that*-complement clauses controlled by nouns), semantic classes (e.g. activity verbs, likelihood adverbs), and lexical-grammatical classes (e.g. *that*-complement clauses controlled by mental verbs, *to*-complement clauses controlled by possibility adjectives).

Overall, the tagger performed successfully over the blog corpus, at approximately 95% accuracy. Nonetheless, the tagger was not quite as successful as it would have been if more standardized texts had been analyzed, due primarily to the presence of non-standard spellings in some of the blogs. The tagger, however, is designed to assign tags to words based on grammatical context when those words (or spellings) are not included in its dictionary. The presence of creative spellings therefore caused minimal error in tagging. It is therefore assumed that the results of the study are valid and would be the same if the corpus had been manually tagged without error.

14.3 Factor Analysis

In order to identify the main dimensions of linguistic variation in the blog variety of the English language, the values of numerous linguistic features were computed for each blog sub-corpus and subjected to a factor analysis. This procedure identified the major patterns of linguistic co-occurrence across the data set. The resultant factors were then interpreted as underlying dimensions of functional linguistic variation. This section describes the factor analysis technique, presents the results of the factor analysis, and interprets these factors so that the major patterns of functional linguistic variation in the blogs register can be determined.

¹ The Biber Tagger is available at the Corpus Linguistics Laboratory at Northern Arizona University.

14.3.1 Method

A factor analysis is a multivariate statistical procedure that identifies systematic co-occurrence patterns in a set of variables. Essentially, a factor analysis is a method of data reduction: it reduces a large set of variables into a smaller set of aggregated factors by determining which of these variables pattern similarly across the dataset.

The application of factor analysis here is based on the theoretical assumption that register differences involve underlying linguistic co-occurrence patterns. When a speaker or writer shifts from one register to another, they naturally shift from one set of co-occurring linguistic features to a different set of co-occurring features. When applied to linguistic data, factor analysis can therefore be used to identify sets of linguistic features that tend to co-occur across the texts of a corpus. This approach was originally developed to analyze the range of spoken and written registers in English [2, 4], and is applied here to uncover the underlying parameters of linguistic variation among blogs. Similar approaches are adopted in Chapter 1 by Santini (this volume) and Chapter 13 by Paolillo et al. (this volume).

For the present study, rates of occurrence were computed for 131 functional linguistic features (see Appendix) across the 500 sub-corpora, based on tag counts for each sub-corpus. The values of these variables were then subjected to a factor analysis (in SAS, using Varimax rotation). Only 54 linguistic features were retained in the final analysis. Several features were dropped because they were redundant or overlapped to a large extent with other features. For example, the counts for common verbs, nouns, and adjectives overlapped extensively with the semantic categories for those word classes, even though the counts were derived independently. In other cases, features were dropped because they were rare in these web documents or because they did not co-occur significantly with other features in these texts (e.g. semantic classes of phrasal verbs). Finally, some features were combined into more general classes. For example, three features that incorporate a passive verb phrase were originally distinguished: agentless passives, passives with a *by*-phrase, and non-finite passives as nominal post-modifiers. These features were combined into the general category of passive verbs in the final analysis.

14.3.2 Results

The solution for four factors was selected as optimal. Taken together, these factors account for 40% of the shared variance and are readily interpretable. Subsequent factors accounted for relatively little additional variance. The features with loadings over 0.30 (positive or negative) are listed in Table 14.1; these are the features used to compute factor scores for the individual blog sub-corpora, where factor scores are computed for a text by summing the frequency of the features loading on that factor.

Table 14.1 Factor loadings

Factor	Loading	Features
1	Positive	Prepositions, attributive adjectives, nominalizations, passives, WH relative clauses, <i>that</i> relative clauses, post nominal <i>to</i> clauses, post nominal <i>that</i> clauses
	Negative	Emphatics, first person pronouns, discourse particles, hedges, past tense, time adverbials, place adverbials, progressive verbs, <i>to</i> clauses with desire/intent/decision verbs, quantity nouns, activity verbs
2	Positive	Present tense, second person pronouns, <i>do</i> as PRO-verb, demonstrative pronouns, <i>be</i> as main verb, indefinite pronouns, WH questions, possibility modals, predictive modals, conditional subordination, necessity modals, mental verbs
	Negative	Prepositions, past tense
3	Positive	Demonstrative pronouns, emphatics, pronoun <i>it</i> , hedges, clausal coordination, adverbs, conjuncts, predicative adjectives, factive adverbs, likelihood adverbs
	Negative	Second person pronouns, nouns
4	Positive	<i>That</i> deletion, past tense, third person pronouns, adverbial subordination (other), <i>that</i> clauses with factive verbs, <i>to</i> clauses with speech act verbs, <i>to</i> clauses with modality/cause/effort verbs, communication verbs
	Negative	Nouns, attributive adjectives

14.3.3 Interpretation of Factors

Once the factors have been extracted, it is necessary to interpret each factor in order to explain why particular features co-occur. This is accomplished by considering the function of the linguistic features loading on each factor and the style of blogs with high positive and negative factor scores.

Factor 1: Informational vs. Personal Focus. To interpret Factor 1, it is first necessary to consider the functional significance of the combination of features with large positive and negative loadings. Positive features on Factor 1 are prepositions, attributive adjectives, nominalizations, passives, and various post-nominal modifying clauses. The negative features on Factor 1 are emphatics, first person pronouns, discourse particles, hedges, past tense, quantity nouns, progressive verbs, activity verbs, *to* clauses with desire/intent/decision verbs, and time and place adverbials.

Most of the positive features on Factor 1 are associated with nouns and noun modification: nominalizations are derived nouns, attributive adjective and post-nominal modifying clauses are noun modifiers, and prepositions often function as the head of post-nominal preposition phrases. Even passives are associated with a highly nominal style: the passive emphasizes the patient noun phrase. The features on the positive end of Factor 1 are therefore quite homogeneous in terms of their linguistic function: all are associate with a highly nominal style. The functional significance of this grouping of features in actual texts tends to be an information-focused style: in particular, high frequency of nouns and noun modifiers in a text

14 Variation Among Blogs: A Multi-Dimensional Analysis

indicates a high informational density, because noun modification is a strategy for packing as much information into a text as possible.

This highly informational language is exemplified by texts from the corpus that score high on Factor 1. Samples of texts from three of the highest scoring blogs are provided in Table 14.2, where pre- and post-nominal modifiers have been highlighted in italics. All three text samples exhibit a highly informational style: clearly, the goal of each text is to inform the reader about a particular topic. Notice that while the texts all have a similar informational focus, they all discuss different topics: economics, technology and world affairs. This informational style can also be characterized as being quite formal. This is because informational style is usually associated with relatively formal registers like newspaper reportage and academic writing.

While the features with high positive loadings on Factor 1 are nominal, many of the features with high negative loadings are verbal, most notably various subtypes of verbs and adverbs. Several negatively loaded features on Factor 1 are also associated with a high degree of personal involvement, including first person pronouns, *to* clauses with desire/intent/decision verbs, emphatics, discourse particles, and hedges. Many of the negative features on Factor 1 are therefore associated with an involved and informal style. Overall, this finding is similar to most multi-dimensional register studies (e.g. [3, 4, 6] Biber, 2006), where the first factor reflects an opposition between nominal/informational discourse and verbal/involved discourse.

However, there is an added consideration with Factor 1 in the present analysis: the negative features loading on Factor 1 do not correspond exactly with previous multi-dimensional studies. Some features normally associated with involvement,

AQ2

Table 14.2 Text samples from blogs with highly positive Factor 1 scores

File	Factor 1 score	Text sample
TX02	38.43	Even before <i>bankruptcy</i> filings began rising this spring, an <i>American Bankers Association</i> survey of 350 member institutions found that <i>credit card loan</i> delinquencies had been increasing when measured by the number of <i>accounts past due</i> . When measured by <i>dollars</i> lost, it has declined. In September, it reported that the rate rose to a record of 4.81% during the <i>second</i> quarter, driven in large part by the <i>higher</i> price of gasoline
NM06	38.04	It's already come up for discussion with the <i>PeopleFinder</i> project and will probably continue to be an area of <i>discussion</i> in the various <i>Recovery 2.0</i> efforts. The <i>cryptorighs.org</i> project called <i>Highfire</i> is dealing with it from a <i>different</i> approach as well
ND05	36.63	In any event, it is impossible to talk about the <i>CIA leak</i> investigation without getting into a <i>discussion about the justification for the War in Iraq</i> . Libby, Cheney, Rove, and others in the <i>Administration</i> were outspoken about their support of the statements made by the president in his <i>State of the Union Address</i> prior to the <i>Iraq</i> invasion which claimed <i>Saddam</i> had attempted to purchase uranium yellowcake from Niger

such as present tense and 2nd person pronouns, are missing here. In their place we find several features that can be associated with a personal narrative style: first person pronouns, past tense, activity verbs, progressive aspect. The frequency of first person pronouns is the most obvious indication that Factor 1 is associated with a personal style. Texts with high negative scores on Factor 1 have one major topic: their author. The remaining features offer evidence of the nature of the personal discussion in which these texts are engaged. In particular, numerous features are associated with a narrative style (past tense, activity verbs, progressive aspect) and one feature is associated with expressing personal plans or desires (*to* clauses with desire/intent/decision verbs). In other words, these personal texts seem to both recount and comment on the lives of their authors.

Table 14.3 presents blog excerpts that illustrate the function of the negative features on Factor 1, where discourse particles and first person pronouns have been italicized.

The personal nature of these texts is immediately noticeable; further analysis reveals that these texts are both personal narratives and personal commentaries. Compared to the texts presented in Table 14.2, the difference is clear: texts with highly positive Factor 1 scores are far more impersonal, informative and formal than texts with highly negative Factor 1 scores, which tend to be personal, narrative/reflective and highly involved. Overall, Factor 1 therefore seems to reflect a contrast between impersonal/informative writing and personal narrative/commentary writing, and is therefore labeled the *Informational vs. Personal Focus Dimension*.

Factor 2: Addressee Focus. The positive features on Factor 2 are present tense, *do* as pro-verb, *be* as main verb, mental verbs, conditional subordination, WH questions, second person, demonstrative and indefinite pronouns, and possibility, pre-

Table 14.3 Text samples from blogs with highly negative Factor 1 scores

File	Factor 1 score	Text sample
NB09	-30.28	<i>So</i> in high school, ur a loser when you dont have a boyfriend right? <i>well</i> if one of my friends asked me this <i>i</i> would tell them no your not a loser you dont need a guy to complete you. But when <i>i</i> think about myself there is always a question in my mine.. why cant <i>i</i> get a guy, why does he like her more than me.. if <i>i</i> was skinnier maybe he'd go out with me. <i>Well i</i> didnt used to feel that way
AL02	-28.26	<i>I</i> am sick and tired of the girl who does nothing at work. The other day she left without out doing half of the things we are supposed to do everyday. <i>I</i> got fed up of her actions and <i>I</i> sent an e-mail to the boss about it. <i>I</i> put it under the guise of concern for her, <i>I</i> really truly am. Kinda, sorta, maybe . . . <i>Anywho</i> , the next day she still hadn't done her work or put away her stuff!!!! <i>Now</i> she has two full buggy's full of things to put away
MI03	-27.35	No, <i>I</i> didn't quite kill anyone (yet :P) my little sister came in my room without knocking and . . . saw . . . <i>well</i> . <i>I</i> 'm not trying to sound gross to anyone, but the fact was my boyfriend was over. It was horrible . . . <i>I</i> think <i>I</i> 'm scared for life! <i>I</i> havn't even spoken to her since the scenario. <i>I</i> think <i>I</i> walked in on my older sister and her boyfriend once, but <i>I</i> was only five . . .

14 Variation Among Blogs: A Multi-Dimensional Analysis

dictive and necessity modals. The negative features on Factor 2 are prepositions and past tense.

The positive features on Factor 2 are associated with interactive discourse and addressee focus. The high frequency of second person pronouns is most obviously associated with an interactive style: *you* is a direct reference to the audience of the text. Additionally, the high frequency of other pronouns indicates an interactive style because the frequent use of pronouns implies that the reader is in the same basic frame of reference as the author. Similarly, WH-questions are directly related to interaction between interlocutors, and other features, such as modals, mental verbs, and conditional subordination are used in blogs for giving advice or instruction to the reader. Finally, the high frequency of present tense verbs reflects a focus on current events, in contrast to the past tense verbs that have a negative loading on Factor 2.

Variation in interaction across Factor 2 can easily be seen if some of the most highly positively and negatively scoring blogs are considered. Table 14.4 presents texts samples with highly positive Factor 2 scores and Table 14.5 presents texts samples with highly negative Factor 2 scores. Second person pronouns, WH questions and modals are highlighted in italics in each table.

The blogs in Table 14.4 are clearly far more interactive and addressee-focused than the blogs in Table 14.5: the first blog is basically an advertisement, the second is written in a conversational style, and the third discusses the relocation of a blog. The blogs in Table 14.5, on the other hand, discuss a variety of topics in a variety of styles, but none refer directly to the readership. These differences result in clear differences in the highlighted features: there are 12 highlighted features in Table 14.4 but only 1 highlighted feature in Table 14.5.

Overall, Factor 2 therefore seems to reflect functional variation in blogs associated with the degree of interaction and the focus on the addressee: blogs scoring highly positively on Factor 2 directly refer to and interact with their readership, whereas blogs scoring highly negatively on Factor 2 essentially ignore their audience, choosing instead to simply express information. This factor is therefore labeled as the *Addressee Focus Dimension*.

Table 14.4 Text samples from blogs with highly positive Factor 2 scores

File	Factor 2 score	Text sample
MN07	19.76	Anybody interested in any custom sewing needs (a friend of mine <i>will</i> be hitting Sharon up for a kilt, I know) <i>should</i> get ahold of her Even for clothing for the Mundane . . . :) She does that stuff, too. Thank <i>you</i> , Sharon, it's beautiful!
MT08	28.35	Hopefully when I'm on a new shift and not dog-ass tired, I'll be angry/opinionated again. I can't foresee this apathy being permanent. Maybe in the mean time I'll post poetry (fortunately for <i>you</i> , not mine!), or other meaningless drivel. I'm also open to suggestions . . .
ND07	22.56	I hope <i>you</i> find the new place to your liking. And if not it's my blog and the fact that I like it is all that matters . . . And those of <i>you</i> who <i>may</i> have me on your blogroll, make sure <i>you</i> change it to the new URL or <i>you'll</i> never know

Table 14.5 Text samples from blogs with highly negative Factor 2 scores

File	Factor 2 score	Text sample
AL09	-22.87	The trip downstream was beautiful as the sun rose to burn off the early morning misty haze hanging over the water. Once I reached the three mile mark (right at 30:03), the water opened up a little and the breeze started to stir up the surface. It actually helped push me along and my speed improved by 0.1 mph, as I passed Buzbee's
MI08	-21.68	For many reasons, the Caesars made their mark on professional sports in Detroit. It introduced the town to Mike Illitch, who <i>would</i> become a player in Detroit with his ownership of the Tigers, Red Wings, Joe Louis Area, and with ventures like the Fox Theater. It brought what amounted to, in baseball terms, a minor league team to the suburbs of Detroit with crowds of 5,000-plus . . .
WI01	-19.05	Reading on the internet indicates well water with minerals is best for plants, instead of treated city water, or bottled water. Perhaps this switch in water is what caused the African Violet to bloom, perhaps it is the time of year, perhaps because I had started to rotate the plant so different areas receive light from the North window

Factor 3: Thematic Variation. The positive features on Factor 3 are pronouns (demonstratives, *it*), emphatics, hedges, predicative adjectives, various adverbs (including conjuncts or linking adverbials), and clausal coordination. The negative features on Factor 3 are second person pronouns and nouns.

At first glance, many of the features loading positively on Factor 3 appear to be associated with a spoken and conversational style: clausal coordination and predicative adjectives are frequent in spoken discourse because they are often the product of unplanned language production; demonstrative pronouns and *it* are associated with generalized and inexplicit references and reduced lexical content, indicative of a shared context of communication; and hedges, emphatics, and factive and likelihood adverbs are all used to express stance, a common function of spoken language. Of course, no blog is spoken, but blogs that have large positive scores on Factor 3 (see Table 14.6) do impart a conversational tone nonetheless. In particular, these blogs shift rapidly from one topic to the next, as is common in spoken discourse. For example, in the first sample, the blogger discusses myspace, golfing and his job in the space of a single one page posting. In contrast, the text samples with highly negative Factor 3 scores (presented in Table 14.7) focus on one topic – not just across these samples, but across all the postings contained in the sub-corpus for that blog. Associating Factor 3 with thematic variation also offers further explanation for the high loading of clausal coordination and conjuncts: if a text discusses numerous disparate ideas and topics, perhaps it is particularly necessary to explicitly connect these parts through grammatical links.

Factor 3 is therefore labeled as the *Thematic Variation Dimension*, where blogs with highly positive scores present information on a variety of topics, while blogs with highly negative scores are focused on a single issue.

14
Variation Among Blogs: A Multi-Dimensional Analysis

Table 14.6 Text samples from blogs with highly positive Factor 3 scores

File	Factor 3 score	Text sample
PA04	21.45	I just spent the entire evening re-updating my Myspace. What a waste of my life. Oh well, it doesn't matter Today I had my first ever golf outing. You see, it was the first practice for the golf team [...] Anyways ... the days before that, on Saturday and Sunday, I worked. I made \$140 this weekend, \$70 each day
MO01	19.23	I signed a lease last wednesday for my firstest apartment, which is pretty scary. Basically, I agreed to part with a lot of my very own hard earned dollars a month for a year [...] I can't remember what my last entry says, but basically Mark is an idiot, and I haven't spoken to him [...] At any rate, it looks like they're going to finally sentence jody
ND02	18.90	Student Congress was fine, I think that speaking infront of those people about my opionons on an issue was possibly one of the scariest things I've ever done though [...] I purchased the new Broken Social Scene CD [...] So, today, Lucy and I went to Wal-Mart

Table 14.7 Text samples from blogs with highly negative Factor 3 scores

File	Factor 3 score	Text sample
MI08	-18.22	For many reasons, the Caesars made their mark on professional sports in Detroit. It introduced the town to Mike Illitch, who would become a player in Detroit with his ownership of the Tigers, Red Wings, Joe Louis Area, and with ventures like the Fox Theater
NM05	-16.70	The Braves had little chance of retaining his services as Mazzone has had a long-standing agreement with Sam Perlozzo, his best friend since childhood, that he would serve as his pitching coach if Perlozzo got a permanent job
MO09	-14.32	On the various New Testament lists of the Twelve Apostles (Matthew 10:2-4; Mark 3:16-19; Luke 6:14-16; Acts 1:13), the tenth and eleventh places are occupied by Simon the Zealot (also called Simon the "Cananean," the Aramaic word meaning "Zealot") and by Judas of James

Factor 4: Narrative Style. The positive features on Factor 4 are past tense, third person pronouns, *that* deletion, factive verbs with *that* clauses, certain forms of adverbial subordination (e.g. since, while), communication verbs, *to* clauses with speech acts verbs, and *to* clauses with modality/cause/effort verbs (e.g. *allow*, *leave*, *order*). The negative features on Factor 4 are nouns and attributive adjectives.

Most of the features that load positively on Factor 4 are associated with a narrative style. The two clearest markers of narration are past tense verbs and third person pronouns. In addition, communication verbs are used to report the speech of others (common in narrative), and the adverbial subordinators that load on Factor 4 (e.g. since, while) are used to make temporal reference – a way to mark time in a

narrative. In contrast, the negatively loaded features on Factor 4 are associated with high information density, similar to the positive features on Factor 1.

The co-occurrence of these features in blogs is illustrated in Tables 14.8 and 14.9. The texts scoring highly positive on Factor 4 are all clearly narratives, whereas the texts scoring highly negative on Factor 4, while accomplishing various communicative goals, do not attempt to tell a story. The differences in pronoun usage, use of communication verbs, and tense are particularly clear (third person pronouns and communication verbs are highlighted in the two tables; neither of these features are found in the second table). Factor 4 has thus been labeled as the *Narrative Style Dimension*.

Table 14.8 Text samples from blogs with highly positive Factor 4 scores

File	Factor 4 score	Text sample
GA06	19.50	Well folks I stopped at the natural health store and picked up something to help me sleep. I'm hoping it works and I won't have to go the prescription drug route. Wish me luck Spoke to PD today. <i>He called</i> while I was at work and kept me on the phone for hours. At one point <i>he asked</i> if I was still attracted to <i>him</i> . I <i>told him</i> unfortunately, yes
MO06	17.06	There wasn't much we could do, in fact <i>he</i> was highly agitated when we were there and being that my mom was exhausted we decided to go home and give the nurses a rest. Later that night my mom and I went back to see <i>him</i> he was sitting up without restraints but he still didn't know anyone. <i>He was talking</i> more <i>he was saying</i> "I need . . . , I"
VA08	16.95	After the meeting I wanted to <i>speak</i> to the 2 teachers, to see how my boy was doing in his class, but <i>everyone</i> swarmed around <i>them</i> and some of the mothers really got chit chatty with the two teachers and it made it very difficult to <i>speak</i> with <i>them</i>

Table 14.9 Text samples from blogs with highly negative Factor 4 scores

File	Factor 4 score	Text sample
NM06	-18.21	I've begun to seriously question the viability of the World Future Society. Most of the interesting news I read comes from specialty websites, blogs, and press releases. And with all the material being produced on a daily basis, a monthly publication that devotes the majority of its' article space to a quick review of "Future Jobs" isn't meeting much of a need
TX02	-15.45	Even before bankruptcy filings began rising this spring, an American Bankers Association survey of 350 member institutions found that credit card loan delinquencies had been increasing when measured by the number of accounts past due. When measured by dollars lost, it has declined
NC01	-14.62	According to Heise Online, during a London demonstration of the shipping version of the XBox 360 hardware, Microsoft showed the console's ability to play nice with Apple's venerable iPod devices. iPods are able to connect with standard Apple cables using the USB ports of the 360

14.4 Text Type Analysis

The previous section discussed the major dimensions of linguistic variation in blogs, which were extracted using a factor analysis. These dimensions will now be used to identify the main text types of the blog variety.

A text type is a variety of language that is defined exclusively by linguistic properties [5, 6]. In other words, a text type is a variety of language that is composed of texts that are maximally similar in terms of their linguistic characteristics. Our definition of a text type therefore differs from our definition of a register, which is a variety of language defined by situational (i.e. non-linguistic) characteristics. Text types therefore do not necessarily correspond to registers: text types can be composed of texts from different registers if these texts are linguistically similar. However, while registers are defined based on situational characteristics, empirical analysis has demonstrated that registers are usually characterized by pervasive linguistic features as well [4]. This is because situational context tends to exert functional pressures on linguistic output. Text types and registers thus represent complementary ways to dissect the textual space of a language.²

In order to identify the major text types of the blog register, the dimensions of linguistic variation produced by the factor analysis will be used as predictor variables in a cluster analysis, which will group blogs into clusters based on the similarity of their scores across the four factors. These clusters will then be interpreted as text types and interpreted functionally by considering the thematic domains and communicative purposes of the blogs grouped into each type.

14.4.1 Method

In the multi-dimensional approach, text types are identified quantitatively using a cluster analysis. A cluster analysis is a multivariate statistical technique used to classify objects into groups based on the values of numerous predictor variables. To identify text types, a cluster analysis is used to group sub-corpora into clusters based on shared linguistic characteristics: the texts grouped together in a cluster are maximally similar linguistically, while the different clusters are maximally different linguistically. This approach has been used to identify the general text types in English and Somali [5, 6], and has also been used to identify text types among web sites [7].

² In addition, while the concept of a *genre* is not as important in our system as the concepts of *text type* and *register*, we define a *genre* in a very similar manner to how we define *register* – i.e. as a variety of language defined by the external situation in which it is produced. However, while a register is characterized by pervasive linguistic features, a genre is characterized by conventionalized linguistic features.

In the present study, a cluster analysis was used to determine the main text types of the blog register, with the four dimensions of variation produced by the factor analysis used as predictor variables. The resultant groups identified by the cluster analysis are interpreted as representing blog text types.

The FASTCLUS procedure from SAS was used for the cluster analysis. Disjoint clusters were analyzed because there was no theoretical reason to expect a hierarchical structure. Peaks in the cubic clustering criterion and the pseudo-F statistic (produced by FASTCLUS) were used to determine the optimal number of clusters. These measures are heuristic devices that reflect goodness-of-fit – the extent to which the texts within the clusters are similar and the extent to which the clusters are maximally distinguished.

In order to define the text types produced by the cluster analysis, two types of information are considered: the four dimension scores for each cluster, and detailed consideration of prototypical blogs from each cluster, where blogs closest to a cluster centroid are deemed to be prototypical members of that cluster.

14.4.2 Results

Based on an analysis of peaks in the clustering criterion and the pseudo-F statistic, three clusters were identified as the optimal solution. Table 14.10 presents the basic descriptive statistics for the three clusters identified by the cluster analysis: the number of blogs in each cluster, the dispersion within each cluster (maximum distance to the cluster centroid), the nearest cluster, and distance between neighboring cluster centroids.

The vast majority of the blogs (94.6%) were classified into either Cluster 1 or 3, which are also the two clusters that are closest together. Cluster 2, which contains comparatively few blogs, is therefore the most distinctive.

Table 14.11 presents the values for each of the four factors across the three clusters, averaged across the blog sub-corpora in each cluster.

Cluster 1 is characterized by a negative score on Factor 1 and positive scores on Factors 2, 3, and 4, whereas Clusters 2 and 3 are characterized by positive scores on Factor 1, and negative scores on Factors 2, 3, and 4. Clusters 2 and 3 are distinguished by the relative strength of the scores across the factors, especially on Factor 1: blogs in Cluster 2 are characterized by far larger positive values on Factor 1 than blogs in Cluster 3.

Table 14.10 Cluster analysis results

Cluster	Frequency	Max distance	Nearest cluster	Cluster distance
1	236	29.30	3	21.75
2	27	27.79	3	25.67
3	237	25.65	1	21.75

14 Variation Among Blogs: A Multi-Dimensional Analysis

Table 14.11 Average factor scores across clusters

Cluster	Factor 1 informational vs. personal	Factor 2 addressee focus	Factor 3 thematic variation	Factor 4 narrative style
1	-11.45	2.15	4.14	3.49
2	32.01	-5.24	-7.55	-8.69
3	7.75	-1.54	-3.26	-2.48

14.4.3 Interpretation of Clusters

Based on the factor averages, blogs in Cluster 1 should be more highly personal, both in terms of topic and voice, and relatively more narrative, addressee focused, and thematically variable than blogs in the other two clusters. This interpretation of Cluster 1 can be validated by considering samples from blogs that are prototypical of Cluster 1 (i.e. which fall closest to the centroids of Cluster 1). Samples from the three most prototypical blogs are presented in Table 14.12.

In all cases, these blogs have a clear personal focus – not only in terms of the style of the blog, but also in terms of the thematic focus of the blog. In the first blog sample, the author narrates an event from his or her life. In the second blog sample, the author opines on his or her favorite scary movies and books. In the third blog sample, the author discusses balancing his or her family’s budget. These blogs also appear to be both fairly informal and narrative in their style. This first text type therefore seems to correspond roughly to the personal blog type or the online diary blog type usually identified in past research as one of the major types of blogs.

Table 14.12 Text samples from prototypical Cluster 1 blogs

File	Dist. to centroid	Text sample
VA06	2.26	As always, Red Lobster was gorgeous and tasted even better knowing that I wasn’t going to have to pay for any of it. I’m terrible like that, I know. It was good conversation, good food, and a good time. K entertained us with his random knowledge of how people can be buried and Minnie told us that she was pulled over by a cop for speeding again on Wednesday, but after some hardcore flirting only got the police officer’s cell phone number
WV07	2.65	Scary movies are so hard to make good through the end. Usually I find them lame, or too gory to deal with. I’d have to say, The Ring came close. As far as books go, it’s about the same deal, so I’m just going to pick the one that freaked me out the most: Whispers, by Dean Koontz freaked me out in ways I never thought possible
MT09	3.52	Balancing the budget here at home is really easy – there is a lot of toys that we want and when there are funds available to meet that end we do it. We never give up our savings right off the top of the payroll. Hopefully after a long time doing it this way – payday can come and we won’t even really notice – nor look forward to. I soooo look forward to that day

Based on cluster averages, blogs in Cluster 2 should be highly impersonal and informational, non-interactive, thematically focused and non-narrative. This assumption can be validated by considering samples from blogs that are prototypical of Cluster 2. Samples from the three most prototypical blogs are presented in Table 14.13. As expected, the three blog samples provided in Table 14.13 are all characterized by a highly informational/impersonal style, clearly presenting the author’s strong opinion on a particular subject in an formal style reminiscent of newspaper articles and academic writing.

Similar to Cluster 2, blogs in Cluster 3 are characterized by an informational style as well – although the average Cluster 3 scores for Factor 1 are far weaker than the average Cluster 2 scores for Factor 1. The same is true of the other three dimensions, which are relatively unmarked for Cluster 3, though Cluster 3 is somewhat more addressee focused than Cluster 2. Cluster 3 therefore seems to be an intermediate type between Cluster 1 and Cluster 2. This pattern can be better understood by considering samples from blogs that are prototypical of Cluster 3, which are presented in Table 14.14. These examples reveal the difference between the blogs in Cluster 3 and Cluster 2: while blogs in Cluster 2 are completely impersonal and informational, blogs in Cluster 3 often use a personal voice to discuss and offer opinions on impersonal topics – in all of these most prototypical cases, surprisingly, storms and hurricanes. These blogs appear to be more informal and conversational than blogs in Cluster 2 as well. Cluster 3 therefore falls in between Cluster 1 and Cluster 2, by using a personal voice (like those blogs in Cluster 1) to discuss impersonal topics (like those blogs in Cluster 2).

To summarize, blogs in Cluster 1 are written in a very personal voice and are concerned primarily with the blogger’s own life. Given this subject matter it is not surprising that these blogs also tend to be narratives and tend to vary thematically.

Table 14.13 Text samples from prototypical Cluster 2 blogs

File	Dist. to centroid	Text sample
ND05	3.08	The bottom line is the Administration could have simply corrected Wilson by pointing out that the Vice President’s office did not recommend his trip to Niger, and that the CIA was solely responsible for the recommendation. There was no reason to mention that Wilson had a wife who worked in the CIA, whether or not she was in fact a covert agent
KY09	5.12	Supporters of the war in Iraq, and many of President Bush’s neighbors in Crawford, Texas, must surely be casting about for some present-day equivalent to the then ratings-challenged Alphabet Network. Cindy Sheehan and Camp Casey apparently have frayed the nerves of both groups
WV06	7.33	Politics and religion, in their most extreme forms, have taken insight and inspiration and systemized them into imperious and unyielding masters. These masters are ideology and dogma. They appeal to the simple-minded, who take comfort in dualistic absolutes. There is no place for subtlety or nuance. Neither is there openness to change or growth

14 Variation Among Blogs: A Multi-Dimensional Analysis

Table 14.14 Text samples from prototypical Cluster 3 blogs

File	Dist. to centroid	Text sample
VA04	2.04	Stories I am hearing form the locals are almost unimaginable. One woman was telling me as she stood in her house at the start of the storm she had a few feet of water flood into her house. Then the eye came through and there was quiet. Then the other side of the storm came over and her house flooded so high she, her mom, and daughter had to stand on stools to keep their heads above water
MA04	2.63	Of course, I also saw signs on Topsail that it isn't just storms that can damage beach areas. I was astounded to find one morning that the normal surf was eroding the beach as the tide came in. There had been no overnight storm that I knew of, and the waves didn't seem particularly intense at the time. Still, the beach that had looked like this the day before
VA07	3.73	Don't get me wrong. I am delighted to see what a good job FEMA and all the other agencies seem to have done with Hurricane Rita. It's nice to know that these agencies can do what they are supposed to be doing – providing, of course that you get the political appointees the hell out of the way and let trained people do their jobs!

These blogs also tend to be relatively addressee focused – a functional pattern that seems to reflect the conversational style of these blogs and their author’s desire to have their blogs read, enjoyed and commented on by their readers, and perhaps in particular by their friends. Cluster 1 is therefore labeled the *personal diary blog type*. This blog type is very common: out of the 500 of the blogs contained in the corpus, 236 blogs fall under Cluster 1.

Blogs in Cluster 2, on the other hand, are written in very formal and impersonal style and are used by their authors to convey information on a particular topic. These blogs read like newspaper and academic articles because of their similar communicative goals. Cluster 2 is therefore labeled as the *expert blog type*. This blog type, however, is very uncommon: out of the 500 blogs contained in the corpus, only 27 blogs fall under Cluster 2. Cluster 2 is also the most distinct of the three clusters.

The remaining 237 blogs fall under Cluster 3. These blogs, like those in Cluster 2, are informational; however, these blogs are characterized by a relatively personal and addressee focused tone, like those in Cluster 1. In other words, in terms of style, they read much like the personal diary type blogs found under Cluster 1. These blogs are therefore labeled as the *commentary blog type*: they are used by their authors to convey their opinions on one or more topics, but unlike expert blogs they are written in a more personal tone.

Because most of the blogs in the corpus are written in a personal and conversational style that characterizes Clusters 1 and 3 (which contain 473 of the 500 blogs and are also the two closest clusters), it appears that this style is the standard blog voice. Given the nature of the medium – a personal web site which is read and commented on by others – the fact that this tone is generally adopted is not surprising. The major division in blog writing is therefore based on topic: blogs that

focus on their authors' lives (Cluster 1) are distinguished from blogs that focus on impersonal topics (Cluster 3). The expert blog type (Cluster 2) is then seen as a marginal third blog type which is informational, like those blogs in Cluster 2, but which is written in a distinctly non-blog-like style, more consistent with standard forms of informational writing. The basic division between personal blogs and thematic blogs, posited in the introduction and past research, therefore seems to be correct.

14.5 Summary of Findings

In conclusion, based on a factor analysis of functional linguistic variation across a 2 million word corpus of blogs, four principal dimensions of linguistic variation were identified, which represent significant patterns of functional linguistic variation for this variety of language. These four principal dimensions are the informational vs. personal focus dimension, the addressee focus dimension, the thematic variation dimension, and the narrative style dimension. The four dimensions produced by the factor analysis were also used as predictors in a cluster analysis in order to identify the major linguistically-defined categories of blogs. Two major blog text-types were identified by the cluster analysis: personal blogs and thematic blogs. Both of these blog types are characterized by a highly personal and conversational style, which appears to be the standard blog voice. The difference between these two types involves the content of the blogs: blogs that focus on their authors' lives are distinguished from blogs that focus on impersonal or informational topics. In addition, a marginal third blog type – labeled the expert blog – was identified, although this blog type appears to be quite rare. The expert blog is informational, like the thematic blog, but is written in a distinctly non-blog-like style, similar to standard informational writing. It was therefore concluded that there are two basic types of blogs: personal blogs and thematic blogs. This finding confirms common assumptions about blog registers.

Appendix

List of 131 Initial Features

Verbs, private verbs, public verbs, mental verbs, activity verbs, persuasive verbs, communication verbs, occurrence verbs, causative verbs, existence verbs, aspectual verbs, common verbs, pro-verb do, auxiliary have, be as main verb, transitive phrasal verbs, intransitive phrasal verb, mental phrasal verb, communication phrasal verb, occurrence phrasal verb, copular phrasal verb, aspectual phrasal verb, activity phrasal verb, modals, predictive modals, possibility modals, necessity modal, present tense, past tense, perfect aspect, progressive aspect, infinitives, passives, agentless passives, by passives, post-nominal passive, prepositions, pronouns, first

14 Variation Among Blogs: A Multi-Dimensional Analysis

person pronouns, second person pronouns, third person pronouns, demonstrative pronouns, pronoun it, nouns, nominalizations, group nouns, place nouns, quantity nouns, technical/concrete nouns, abstract nouns, cognitive nouns, process nouns, human nouns, adjectives, attributive adjectives, predicative adjectives, color attributive adjectives, evaluative attributive adjectives, time attributive adjectives, size attributive adjectives, topical attributive adjectives, relational attributive adjectives, particles, adverbs, non-factive adverbs, factive adverbs, likelihood adverbs, attitudinal adverbs, adverbials, time adverbials, place adverbials, conjuncts, clausal coordination, phrasal coordination, downtoners, amplifiers, general emphatics, general hedges, conjunctions, conditional subordinators, causative subordinators, concession subordinators, other subordinators, wh words, wh questions, wh clauses, wh relative clauses, object relatives, subject relatives, that relatives clauses, all that clauses, all that clauses with verbs, all that clauses with nouns, all that clauses with adjectives, that clause with non-factive verbs, that clause with factive verbs, that clause with attitudinal verbs, that clause with likelihood verbs, that clauses with factive adjectives, that clauses with attitudinal adjectives, that clauses with likelihood adjectives, that clauses with non-factive nouns, that clauses with factive nouns, that clauses with attitudinal nouns, that clauses with likelihood nouns, all to clauses, to clauses with all adjectives, to clauses with all verbs, to clauses with mental verbs, to clause with desire/intent/decision verbs, to clause with effort verbs, to clause with probability verbs, to clause with speech act verbs, to clause with modality/cause/effort verbs, to clause with all nouns, to clauses with certainty adjectives, to clauses with ability/will adjectives, to clauses with personal affect adjectives, to clauses with ease/difficulty adjectives, to clauses with evaluative adjectives, verb complements, adjective complements, type token ratio, average word length, text length, that deletion, contraction, stranded prepositions, split auxiliaries, post nominal to clauses, post nominal that clauses.

References

1. Argamon, S., M. Koppel, J. Fine, and A.R. Shimoni. 2006. Gender, genre, and writing style in formal written texts. *Text* 23(3):321–346.
2. Biber, D. 1986. Spoken and written textual dimensions in English: resolving the contradictory findings. *Language* 62:384–414.
3. Biber, D. 1987. A textual comparison of British and American writing. *American Speech* 62:99–119.
4. Biber, D. 1988. *Variation across speech and writing*. Cambridge, MA: Cambridge University Press.
5. Biber, D. 1989. A typology of English texts. *Language* 27:3–43.
6. Biber, D. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge, MA: Cambridge University Press.
7. Biber, D., and J. Kurjian. 2007. Towards a taxonomy of web registers and text types: A multi-dimensional analysis. In *Corpus linguistics and the web*, eds. M. Hundt, N. Nesselhauf, and C. Biewer, 109–132. Amsterdam: Rodopi.
8. Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

9. Blood, R. 2002. Introduction. In *We've got blog*, ed. J. Rodzwillz. Cambridge, MA: Perseus.
10. Delwiche, A. 2004. Agenda setting, opinion leadership, and the world web logs. Presented at the Annual Conference of the International Communication Association. New Orleans, LA.
11. Gilmore, D. 2003. Moving toward participatory journalism. *Nieman Reports* 57(3):79–80.
12. Harmanci, R. 2005. Time to get a life—pioneer blogger Justin Hall bows out at 31. *San Francisco Chronicle*, 20 Feb 2005.
13. Herring, S.C., and J.C. Paolillo. 2006. Gender and genre variation in weblogs. *Sociolinguistics* 10(4):439–459.
14. Herring, S.C., L.A. Scheidt, S. Bonus, and E. Wright. 2004. Bridging the gap: A genre analysis of weblogs. In *Proceedings of the 37th Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.
15. Herring, S.C., L.A. Scheidt, S. Bonus, and E. Wright. 2005. Weblogs as a bridging genre. *Information, Technology and People* 18(2):142–171.
16. Herring, S.C., L.A. Scheidt, I. Kouper, and E. Wright. 2007. A longitudinal content analysis of weblogs: 2003–2004. In *Blogging, citizenship and the future of media*, ed. M. Tremayne. London: Routledge.
17. Krishnamurthy, S. 2002. The multidimensionality of blog conversations: The virtual enactment of September 11. Presented at Internet Research 3.0. Massricht, The Netherlands.
18. Lasica, J.D. 2002. Blogging as a form of journalism. *USC Annenberg Online Journalism Review*.
19. Lasica, J.D. 2003. Blogs and journalism need each other. *Nieman Reports* 57(3):70–73.
20. Miller, C.R., and D. Shepherd. 2004. Blogging as social action: a genre analysis of the weblog. In *Into the Blogosphere: Rhetoric, Community and the Culture of Weblogs*, eds. L. Gurak, S. Antonijevec, L. Johnson, C. Ratliff, and J. Reyman. Minneapolis, MN: University of Minnesota Press.
21. Papcharissi, Z. 2004. The blogger revolution? Audience and media as producers. Presented at the Annual conference of the International Communication Association. New Orleans, LA.
22. Park, D. 2003. Bloggers and warbloggers as public intellectuals: Charging the authoritative space of the weblog. Presented at Internet Research 4.0. Toronto, ON, Canada.
23. Pushmann, C. 2007a. Corpora, blogs and linguistic variation – arguments for using structured web data in corpus development. Presented 8 Nov 2007, at the University of Paderborn, Germany. <http://www.slideshare.net/coffee001/corpora-blogs-and-linguistic-variation-paderborn>
24. Pushmann, C. 2007b. Blogs or flogs? Genre conventions and linguistic practices in corporate web logs. Presented 31 Aug 2007, at the Telematica Instituut, Enschede, The Netherlands. <http://www.slideshare.net/coffee001/blogs-or-flogs-genre-conventions-and-linguistic-practices-in-corporate-web-logs>
25. Wortham, J. 2007. After 10 years of blogs, the future's brighter than ever. *Wired Magazine*. Retrieved 14 Jan 2008 from http://www.wired.com/entertainment/theweb/news/2007/12/blog_anniversary

Chapter 14

Query No.	Query
AQ1	“The Economist, 2006” is not listed in the reference list. Please provide.
AQ2	“Biber, 2006” is not listed in the reference list. Please provide.
AQ3	Please provide volume and page range for the reference [18].

AQ 1

Kluth, A. 2006. It's the links stupid. The Economist, 379 (April 22, 2006): 5.

AQ2

Biber, D. 2006. University language: A corpus-based study of spoken and written registers. Amsterdam: John Benjamins.

AQ3

No page range, as it was published online.
However, there is an error: year of publication is 2001, not 2002.
Also, URL is: www.ojr.org/ojr/workplace/1017958873.php
Retrieved January 14, 2008