GIULIANA GARZONE

Investigating Blawgs through Corpus Linguistics: Issues of Generic Integrity

1. Introduction

This chapter focuses on law blogs, also called 'blawgs', i.e. blogs used by the legal and professional community to exchange scholarly opinions, to debate topical issues and discuss important legal cases. This is one of a diversified range of thematic blogs in domain-specific and professional communication into which weblogs have diversified.

The main purpose of this chapter is to explore some of the distinctive features of the law blog as a genre in order to determine to what extent this realization of the weblog has moved away in time from its original form, characterised by a personal, diary-like format into a type of academic, professional or journalistic expression, and whether this evolution has been so extensive as to jeopardise its generic integrity.[1] The analysis is essentially corpus-based and starts with a discussion of the criteria to be applied in the construction of a corpus of blogs.

---

[1]   Bhatia (2004: 123) defines 'generic integrity' as " a socially constructed typical constellation of form-function correlation representing a specific construct realizing a specific communicative purpose of the genre in question". This research is aimed at verifying whether as a result of its evolution the blawg today has preserved the core elements of the form-function correlation characterising the construct that realizes the weblog genre.

1.1. The weblog as a genre

Weblogs have been with us for less than two decades,[2] an astonishingly short stretch of time if one considers that the latest available reliable statistics indicate that as of 31 May 2011 there were over 162 million public blogs in existence.[3]

The blog is typically "located at the intersection of the private and public realms" (Miller/Shepherd 2004), offering individuals an extraordinary opportunity of self-expression, open to an audience that can be planetary. This is possible thanks to the affordances of the blog format, which is in an intermediate position between, on the one side, relatively static and asymmetrical HTML documents and, on the other side, interactive text-based Computer Mediated Communication (e.g. newsgroups, chatrooms, etc.) as well as social networks, to which blogs are more and more often linked. According to Herring/Scheidt *et al.* (2005: 162), blogs "allow authors to experience social interaction in ways that are otherwise difficult to achieve through web pages, while giving them ownership of, and control over, the communication space that is difficult to achieve in CMC".

Preliminary to this discussion, it is worth examining the distinctive traits as well as the main characteristics of the genre under investigation. A blog is "a website containing an archive of regularly updated online postings" (Grieve *et al.* 2010: 303), which has traditionally been described as characterised by three distinctive features: reverse chronology of its entries, frequent updating, and a combination of links with personal commentary (cf. e.g. Miller/Shepherd 2004; Herring/Kouper *et al.* 2005; Garzone 2012). These properties are accompanied by recurrent peripheral features that make blogs promptly recognizable: they present posts, which include

---

2   The term 'weblog' was first used by Jørn Barger on 17 December 1997 (Barger 2007), while its short form, 'blog', was coined in April/May 1999 by Peter Merholz, who jokingly broke the word *weblog* into the phrase *we blog* in the sidebar of his blog Peterme.com.

3   Statistics published by the Nielsen Company on its BlogPulse website <blogpulse.com>, last accessed 31.05.2011.

a date, a time stamp, and a permalink,[4] usually followed by a repertoire of links that mostly feature comment threads reacting to or commenting on the content of relevant posts (cf. Riboni 2010: 91ff.). The presence of comments responding to the posts, or in some cases to other comments, is also a recurrent feature of blogs, although it is not obligatory (as discussed in §2.1 below). In most cases each post is accompanied by a photograph of the writer, a specification of his/her name and, sometimes, even short biodata; this is sometimes also true for comments, which similarly feature the contributor's name or, more frequently, nickname. It has been suggested that a further distinctive trait is the presence of links to other blogs (Marlow 2004; cf. also Blood 2002), but this is not always the case (Herring/Kouper *et al.* 2005), as some blogs prefer linking to websites, mainstream media or social networks, e.g. Facebook or Twitter (Myers 2009: 28-37). The combination of posts, comments and links is characteristic of the genre, as it makes it possible to realize a peculiar kind of interpersonal interaction, favouring a sense of community among bloggers.

  Having originated mainly as individualistic forms of self expression, blogs have become greatly diversified in their realizations, also as a function of the topics they deal with, hybridizing with relevant traditional and online genres. For instance Herring/Scheidt *et al.* (2005) identify the influence of travelogues on travel blogs, and that of researchers' project journals on k-logs;[5] Garzone (2012) sees posts on 'news and blogging' websites (e.g. *The Huffington Post*) as the result of hybridization between blogs and news discourse (cf. also Bruce 2010), especially comment articles and editorials (cf. van Dijk 1992; Vestergaard 2000; Westin 2002). Finally, Mauranen (2013) identifies the conference paper as one of the ancestral genres of the science blog.

---

4 A permalink, or permanent link, is a URL that points to a blog post and remains unchanged, even when it passes from the front page to the archives.

5 'K-log' is an acronym for *knowledge blog*, which is an internal/Intranet blog, not accessible to the general public, which serves as a knowledge management system; the term 'klog' is also used for a blog that is technical-content oriented. Cf. <http://www.webopedia.com/TERM/K/klog.html>, retrieved 31/05/2011.

These considerations provide a general picture of the blogosphere as essentially divided into personal blogs and thematic blogs, as proposed by Krishnamurthy (2002). The same distinction has also been made by Grieve *et al.* (2010) through a factor analysis of functional linguistic variation in blogs, with the identification of thematic and personal blogs as the major blog text-types, followed by a marginal and rare blog type, the expert blog.

According to the BlogPulse statistics referred to above, in 2011 the personal blog was still by far the most popular compared with the other single most popular categories of blogs classified by topic – cultural issues, TV and movies, politics and sports. However, in the last few years thematic blogs have increased their relative importance, with many of them developing from single-author into multi-author blogs (MABs) where posts are written by many different authors and professionally edited, often hosted on the websites of institutions or organisations, such as companies, public departments, research bodies, newspapers or magazines, etc.

From the viewpoint of genre categorization and genre analysis, blogs have been perceived as problematic. In this respect they are similar to other web-mediated genres (cf. e.g. Askehave/Ellerup Nielsen 2004, 2005; Garzone 2007), but they pose further problems due to their rapid and extensive diversification which – it has been argued – undermines the shared rhetorical action that originally made it possible to classify all blogs as belonging to a single genre (Miller/Shepherd 2009). Mauranen (2013) also points out that in the case of blogs a pre-existing discourse community of users can hardly be identified, and this makes it impossible to meet one of the criteria traditionally used in genre analysis for the purpose of genre identification (i.e. recognition and understanding of the communicative purpose of the relevant discursive events *by the members of the professional or academic community in which it occurs*, Bhatia 1993; cf. also Swales 1990), given that in web-mediated communication "it is the context that seems to create genres, and communities emerge around them" (Mauranen 2013: 30).

The approach adopted in this chapter prefers to consider the blog as a macro-genre (Garzone 2012), as proposed by Bhatia (2004) for

letters,[6] comprising a whole range of 'hyphenated' (i.e. bi- or poly-rhematic) genres (public affairs blog, corporate blog, political blog, etc.), each with its own peculiarities but also sharing some distinctive features and part of its communicative purpose with all other blogs.

2. Method

The aim here is to assess the generic integrity of blawgs and their degree of differentiation from their parent genre. Given that the original purpose of blogs in general was self-expression – an element that is still considered to be distinctive of the macro-genre and therefore shared by the (micro-)genres it includes as a pre-requisite for membership – the analysis will focus on the degree of persistence of personal/existential elements.

The linguistic/discursive realization taken to be an indicator of such an elements is the frequency of interactional metadiscourse, and in particular of self-mention (Hyland 2001). Accordingly, the analysis will focus on personal pronouns and other forms of self-mention, comparing results with those obtained from two comparable corpora, also observing variations across posts and comments. It will also consider the thematic status of first-person pronouns in a systemic-functional perspective (Halliday 1994: 36ff.; 109ff) as well as the lexical verbs associated with such pronouns.

The methodological toolbox relied on in this study is essentially based on discourse analysis, and especially on metadiscourse research (Crismore 1989; Hyland 2005), using Hyland's (2001) work on self-mention in academic research articles.

It also considers studies on blogs from various disciplinary perspectives (e.g. Blood 2002; Krishnamurthy 2002; Herring/Kouper

---

6   Bhatia (2004: 34) describes the 'letter' as a macro-genre, while he considers the 'business letter' as a 'genre colony'. It is to be noted that among scholars there is hardly any agreement as to how to refer to groupings of a number of (sub)genres that serve – at least partially, if not exclusively – the same purpose.

et al. 2005; Herring/Paolillo 2006; Puschmann 2009; Giltrow/Stein 2009; Myers 2009; Grieve *et al.* 2010; Mauranen 2013) and more specifically on law blogs, mainly by legal scholars (e.g. Caron 2006; Kerr 2006). Finally I will take account of the literature on the suitability of traditional analytical tools for the analysis of Web-mediated communication from various various methodological perspectives (i.e. text linguistics: Garzone 2002; genre analysis: Askehave/Ellerup Nielsen 2005, Garzone 2007; Giltrow/Stein 2009; Santini/Meheler/Sharoff 2010; argumentation theory: Carter 2000; Lewiński 2010; Degano 2012).

The investigation will be carried out on *ad hoc* corpora of the relevant micro-genre, using results from automatic interrogation routines (Wordsmith Tools 5.0) as a basis for the discussion.

2.1. Compiling a web-derived corpus: the case of blogs

Constructing a Web-derived corpus raises a number of issues, not only when it is based on automatic web genre recognition (Sharoff 2010).[7] According to Santini *et al.* (2010) such issues include the presence of centrifugal forces at work on the Web and the indeterminacy of document boundaries (with special regard for granularity)[8] in computer-mediated environments, linked to the "composite and diversified characterization of web documents", ad enhanced by "the

---

7   Web-derived corpora are collected by using either search engines or web crawlers (i.e. Internet robots, or 'bots', software applications that run automated tasks over the Internet), which means that they are very large in size as there can be no pre-planning of the type of documents included. For corpus content categorization, the huge size of the corpora makes it necessary to make recourse to automatic classification, a complex operation which is usually performed by comparing their composition against traditional corpora (e.g. the BNC). This is inherently problematic both in terms of accuracy and of preliminary procedures (supervised machine learning). Cf. Sharoff (2010).

8   Already in 2007 Garzone included granuality in the list of features that distinguish web-mediated genres from traditional genres, together with *multimodality, extension of participation framework, hypertextuality/hyper-mediality, co-articulation,* and *interactivity* (cf. Garzone 2007: 24-25; cf. also Garzone 2012: 39).

tendency towards rapid change and evolution of genre patterns" (Santini *et al.* 2010:13). These challenges, however, are hardly relevant to this study, which can rely on a clear preliminary definition of the target genre, and of the documents to be included in the corpus.

A further set of more specific issues derives from the fact that Internet pages contain markup, and more specifically according to Santini (2010): navigation frames enabling navigation on a complex website, text-internal hyperlinks, and non-hypertextual markup, such as explicit formatting of headings, etc. In consideration of the linguistic and discursive focus of this study, markup elements were not retained in the compilation of the law blog corpus, which consists only of running text. However, the need to take account of the contextual, multimodal and interactional aspects of the documents analysed was accommodated by integrating automatic interrogation routines with 'manual' analysis, also examining hypertextual, graphic and visual elements that contribute to a text's semiotic configuration.

As yet another problem is the composite nature of the blog genre, which necessarily features posts, while comment threads are mostly present, but are not constitutive of the genre. This is why some researchers have chosen to collect only posts: for instance, in their multi-dimensional factor analysis Grieve *et al.* (2010) only used posts and excluded all other kinds of text found on the blog websites. This choice, albeit justified in theory by the non-mandatory nature of comments, excludes an important component present in the overwhelming majority of blogs, which in most cases have the possibility of inserting comments, at least as a potential option for readers. Even in the few blogs which do not allow direct insertion of comments, as a rule readers' response is not totally excluded. A case in point is the *Daring Fireball* blog,[9] which originally had no option for public comments and encouraged readers to e-mail them, but has now added third-party Chrome/Safari extensions ("Daring Fireball with Comments" (<http://daringfireballwithcomments.net/>) 'retrofitting' (i.e. restoring) the ability to comment back into the site.

---

9   See <daringfireball.net>, run by John Gruber. Unless otherwise specified, all websites referred to in this chapter were last accessed 10/05/2013.

So the standard format of a blog is usually considered to be 'posts plus comments plus peripheral features'.

In light of these considerations, posts and comments are assumed here to be part of one and the same communicative event, although asymmetrical and with different functions.

This is why both the posts and comments of the blogs under investigation are included in my corpus, although they are organised into two separate sub-corpora on account of their discursive differences: posts are pre-planned (sometimes quite long and complex) textual realizations, pointing to a problem, or expressing opinions or commenting on facts, while comments are contingent, often spontaneous, responses to posts or (more rarely) to other comments, and are part of an adjacency pair, the first element of which is the post, or a previous comment. This will make it possible to analyse the texts as one corpus, or – when expedient – to look at Posts and Comments separately.

2.2. *Corpus description*

Blawgs are quite numerous today being used for a variety of purposes, e.g. academic and professional communication, scholarly conversations, and instant academic publication (Berman 2006, 2007; Kerr 2006; Solum 2006; Volokh 2006). They also enable writers in this area of specialization to reach much larger audiences than those traditionally addressed, thus contributing to the dissemination of legal knowledge and the discussion of topical legal cases.

The corpus compiled for this study comprises texts from four blawgs, two from the US and two from the UK: the Wall Street Journal Law Blog, the Above the Law Blog, The Magistrates' Blog and the UK Human Rights Blog. These are examined on the basis of computer-generated textual evidence though the use of WordSmith Tools 5.0 (Scott 2011). The contents of each blawg were copied and pasted into separate files for Posts and Comments, constructing two sub-corpora whose overall details are given in Table 1:

	Main corpus	Posts	Comments
Tokens	153,548	79,666	73,882
Types	12,971	8,764	8,804
STTR	43.37	44.44	44.55

Table 1. Details of Blawg corpus and sub-corpora.

The Wall Street Journal Law Blog (WSJ, <http://blogs.wsj.com/law/>) is part of the website of a major US financial newspaper. It is presented as "an online publication that covers hot cases, emerging trends and big personalities in law […] brought to you by lead writer Joe Palazzolo, with contributions from The Wall Street Journal's legal reporters." Its home page betrays the obvious effort to make it immediately recognizable as a blog on account of its peripheral features, with a photograph of Palazzolo and of the eight main contributors, their names and positions (e.g. "Joanne Chung, Law Bureau Chief").

The other three blawgs are maintained by law professionals. Each of them has its own specific focus.

The Above the Law blog (AL <www.abovethelaw.com>), addressed to law professionals, is run by members of law firms in the US, and is owned by Breaking Media. In *The Washington Post* it has been defined 'a must-read legal blog' (Shapira 2007). It is a site of exchange of views among lawyers, also dealing with business aspects of the legal profession. Occasionally it publishes gossip and rumours about law schools, small legal practices and the salaries and bonuses at many of the large firms. Its homepage, which does not feature contributors' photographs, starts with a Top Stories section, with the headlines of each "Top Story" linking to relevant texts and illustrations.

The Magistrates' Blog (MAG <http://magistratesblog.blogspot.it>), a British website, introduces itself as "written by a team, who may or may not be JPs, but all of whom are interested in the Magistrates' Courts". It purports to offer "Musings and Snippets from a group of people interested in Magistrates' Courts and their work. All cases are based on real ones, but anonymised and composited". It posts experts' personal reflections and thoughts on legal cases and the profession,

making it clear that individual opinions are not to interfere with law enforcement ("JPs swear to enforce the law of the land, whether or not they approve of it").

The fourth source, the UK Human Rights Blog (HR <http://ukhumanrightsblog.com/>) is "written by members of 1 Crown Office Row barristers' chambers," and edited by Adam Wagner, Rosalind English and Angus McCullough QC. It discusses courts' decisions on human rights and topical cases. Quite interestingly, when a legal case is discussed, the post is preceded by its details (e.g. "Association for Molecular Pathology *et al* v Myriad Genetics Inc, *et al*, United States Supreme Court 13 June 2013"), a feature that confers on these posts a typically legal flavour, and is accompanied by a link to the original judgment.

The three professional blogs exhibit most of the peripheral features of the genre – the reverse chronology of entries, frequent updating, and links to other blogs or websites, date and time stamp, but do not feature the writers' photographs.

As this study seeks to identify the peculiarities of the target texts within the more general picture of legal communication, blawg corpus data are compared with data from two comparable corpora consisting of (meta)legal texts, representing both the academic and the professional register: research papers in the law section of the CADIS corpus (courtesy of Prof. M. Gotti, University of Bergamo), hereafter mentioned as 'CLC' (i.e. Cadis Law Corpus), and arbitration awards from Kluwer Bank, issued between 1998 and 2002 and delivered by sole arbitrators, which will be referred to as 'AAC' (Arbitration Award Corpus). Table 2 shows the characteristics of these two corpora:

	CLC	AAC
Tokens	368,980	349,381
Types	18,447	18,635
STTR	40.64	34.95

Table 2. Characteristics of corpora used for comparison.

3. Analysis: individualistic/existential elements

3.1. Pronominal reference

The first step in the analysis focuses on first-person pronouns as forms of self-mention, i.e. as indicators of the textual presence of an individualistic/existential component. Explicit non-pronominal author self-references were also considered, but they are virtually absent in all the corpora, with the exception of one occurrence of 'this writer' and four of 'this author' in CADIS (*this author views, this writer believes*, etc.). Second-person pronouns were also taken into account as 'engagement markers' (Hyland 2005) signalling strong writer involvement when a potential interlocutor is addressed. Table 3 shows the results of the comparison; it is to be noted that the first column features data extracted from the whole Blawg Corpus, comprising both Posts and Comments:

	Blawgs %	CLC %	AAC %
I	0.69	0.04	0.02
Me	0.11	<0.01	0.04
My	0.16	0.03	<0.01
myself	<0.01	<0.01	<0.01
We	0.25	0.07	0.07
Us	0.08	0.03	0.01
Our	0.09	0.02	0.02
ourselves	<0.01	<0.01	<0.01
You	0.52	<0.01	0.01
Your	0.13	<0.01	<0.01
Yourself	0.01	<0.01	<0.01

Table 3. Self-mention in the Blawg Corpus vs CLC and AAC (percentages).

The figures show that first-person singular reference in blawgs, including posts and comments, is much more frequent than in the other two corpora. This difference is especially marked for the first-person singular, as 'I' is over 17 times more frequent than in CLC and

34 times more frequent than in AAC. The divergence is even more striking for second-person pronouns/possessives, which hardly occur at all in CLC and AAC.

Table 4 below shows the same comparison, but data regarding Blwags are broken down analytically distinguishing between Posts and Comments:

	Blawg Posts %	Blawg Comments %	All Blawgs %	CLC %	AAC %
I	0.39	**1.10**	0.69[10]	0.04	0.02
me	0.06	0.18	0.11	<0.01	0.04
my	0.11	**0.25**	0.16	0.03	<0.01
myself	0	<0.01	<0.01	<0.01	<0.01
we	0.25	0.23	0.25	0.07	0.07
us	0.06	0.04	0.08	0.03	0.01
our	0.11	0.07	0.09	0.02	0.02
ourselves	<0.01	<0.01	<0.01	<0.01	<0.01
you	0.37	**0.75**	0.52	<0.01	0.01
your	0.07	0.22	0.13	<0.01	<0.01
yourself	<0.01	0.01	0.01	<0.01	<0.01

Table 4. Self-mention in blawg posts and comments vs CLC and AAC.

The main difference between Posts and Comments in terms of frequency, is in first-person singular and second-person pronominal reference, especially as regards 'I' and 'you', with Comments featuring a much more prominent personal and interpersonal component. However, when compared with the two control corpora also Posts show a much higher frequency of linguistic elements compatible with the original existential dimension of the genre.

The difference between Posts and Comments is further highlighted by searching their respective Wordlists in order to obtain keywords, i.e. the words that in statistical terms are abnormally frequent in one of two sub-corpora in comparison with the other. The results obtained are shown in Table 5:

---

10   Please note that the datum relative to the whole corpus is not necessarily the mathematical average between the figure relative to Posts and that relative to Comments, as the size of the two sub-corpora is uneven.

Comments			Posts		
Keyword	*Rank*	*Keyness*	*Keyword*	*Rank*	*Keyness*
I	3	291.36	Said	3	104.30
September	4	241.81	court	4	65.12
guest	5	192.06	Firm	7	60.30
reply	6	168.59	federal	10	49.07
you	7	87.86	Dewey	11	46.04
I'm	8	86.42	disclosure	12	43.90
don't	9	84.85	justice	13	43.19
it's	10	75.28	Thursday	14	42.50
just	11	62.91	patent	16	36.74
October	12	60.86	settlement	17	36.26
your	13	59.66	judge	18	36.20
you're	14	54.97	lawsuit	22	33.82

Table 5. Keywords in Blawg Comments vs Posts.[11]

It is interesting that the Posts Keywords are mainly legal terms, like 'court', 'disclosure', 'justice', which indicate a specifically domain-oriented focus, as well as words like 'guest' and 'reply' which are part of the functional metalanguage of the genre (the list also includes the day of the week when much of the material contained in the corpus was downloaded). This, together with the presence in Comments Keywords of names of months, is simply evidence of the date specified in every single entry.

As regards the Comments Keywords, it is extremely meaningful that most of those at the top of the list are linguistic elements connected with the interactional dimension, and in particular first- and second-person pronouns and verb forms, 'I' and 'you' being among those with a high keyness index. The presence in the list of contracted verb forms (four in the ten top keywords) is evidence of the more colloquial character of comments, as is the frequent use of 'just', which is typical of spoken language, and confirms the nature of speech-like interactive written discourse that blog comments share

---

11 It is to be noted that only meaningful keywords have been included in the Table, while entries like "S", "T", "U" and "RE", which mainly result from occasional elements that are part of the graphic structure of blog entires, have been omitted.

with many other web genres (cf. Cherney 1999; Baron 2000; Pistolesi 2004).

3.2. Self-mention and theme

As Hyland (2001) points out for academic discourse, writers project their views and confer a personal dimension on discourse not only by choosing to use first-person pronouns, but also by thematising such pronouns, which in English involves fronting them in sentence initial position. This emphasises the writer's presence and the individual character of the views or information put forth in the ensuing sentence.

While Hyland takes into account all exclusive first-person pronouns (i.e. 'I' and exclusive 'we'), in this study only first-person singular pronouns are considered, as their occurrence provides indisputable evidence of a substantial component of self-expression in blawgs.

As regards the criteria used here to identify thematised pronouns, the items counted as thematic are first-person singular pronouns in absolute sentence initial position, thus having the role of Topical theme, as well as first-person singular pronouns appearing at the beginning of a sentence preceded only by a Thematic theme – continuative, conjunction or modal adjunct (cf. Halliday 1994).

According to these criteria, in the Blawg corpus 'I' is used in thematic position in 42.29% of occurrences, while in CLC and AAC the percentage is respectively 33,33% and 35.48%. These data confirm the greater emphasis laid on the individual and personal character of the views put forth in Blawgs as compared to traditional academic and professional genres.

In light of this analysis, it can thus be stated that, although they have developed within a domain-specific context thanks to the efforts of academic and professional writers in the legal field, blawgs have not lost the personal individualistic character that is distinctive of the weblog genre, although this is much more evident in Comments than in Posts.

3.3. Lexical verbs with first-person singular pronouns

In order to gain a better understanding of the role that blawg writers construct for themselves in a text, in this section the verb forms that collocate with first-person singular pronouns are analysed across the three corpora.

In this comparison, the first element that stands out is the variety of lexical verbs used in the text. A further finding is the higher frequency in Blawgs of first-person singular verb forms in the simple past, which provides evidence of a strongly personal narrative component. This is all the more meaningful given that, as Walker (2003: 45) points out, "many weblog entries are shaped as brief, independent narratives, and some are explicitly or implicitly fictional, though the standard genre expectation is non-fiction."

Each of these two aspects – variety of lexical verbs collocating with the first-person singular pronoun and presence of a personal narrative component – will now be discussed in more detail.

The verbs used in the present tense in the Blawg Corpus are mostly verbs of cognition (*think, know, understand*), metadiscursive verbs (*assume, suggest, notice*) mostly realizing speech acts, and attitudinal verbs (*fear, feel, worry*). In contrast, the range of verbs in the two control corpora is much more limited. In CLC there is a prevalence of metadiscursive verbs, while attitudinal verbs are virtually absent, thus indicating a more epistemic rather than affective approach in presenting information and arguments. In AAC, verbs in the present tense are used to put forth the arbitrators' views (*I agree, I hold, I find*), which often take on a performative value as they set out the arbitrator's findings, or illustrate his/her reasoning (*I assume, I consider*).

As regards the Blawg Corpus, it is interesting that first-person singular forms in posts tend to involve mostly verbs of cognition (*believe, think, doubt, hope, worry*, etc.), for example:

(1)   *I doubt* the lenders will fund chapter 11 too much longer (WSJ, post)
(2)   The new law allows the police to deal with squatters, but *I don't anticipate* too
          many prosecutions arising as a result (MAG, post).

Cognitive verbs are also extensively present in comments, but an equally numerous range of attitudinal and metadiscursive verbs is also present (e.g. *assume*, *claim*, *suggest*), which seems to provide evidence of a more explicit argumentative vein. In the following examples, an attitudinal and a cognitive verb are still in the service of argumentation, but cast in an evidently personal dimension:

(3)   However, *I feel* that this woman was – and probably still is – highly unstable
        and not in a rational frame of mind […] (HR, comment)

(4)   *I assume* her legal advisers have advised her to appeal against the length of the
        sentence (HR, post).

In the next example, approval for the interlocutor's humour and the expression of disagreement are expressed in even more personal terms:

(5)   *I love* his wry humour and his quirky knack of capturing the gamut of the
        human experience as seen from the bench. But I reserve the right not to agree
        with everything he says (MAG, comment)

In many cases, the personal component takes on an existential dimension, both in posts and in comments. In the following examples, professional information about the writer is provided, as well as reference to personality traits:

(6)   I do not sit in family court (MAG, post)

(7)   I am not one for ruining careers (AL, post)

(8)   I am a loyal follower of the UK Human Rights Blog. (HR, comment)

In (9) below, the present progressive is used to express the commenter's personal choices:

(9)   I am considering this school since it MIGHT be the only law school I can
        attend, I cannot move down south. (WSJ, comment; capitalization for
        'flaming' in the original)

It is to be noted that blawgs contain, albeit rarely, also other types of verbs which do not occur at all in the control corpora. See the following examples featuring a behavioural/physiological verb (*live*) and verbs of doing (*blog, cook*):

(10)   And hey, *I live* in San Francisco and *I BLOG* for a living. (AL, comment; capitalization for 'flaming' in the original)

(11)   Now *I can cook* a bit, even if I wouldn't get far on Masterchef. (MAG, post)

In both cases it is evident that the presence of some verbs that do not belong to the metadiscursive, cognitive or attitudinal categories typically used in professional and academic communication is strictly connected with the projection of the writer's existential dimension into the text.

The difference between blawgs and the control corpora is even more marked if first-person singular verb forms in the present perfect are examined, as in CLC and in AAC they include exclusively metadiscursive or metacognitive verbs (e.g. *I have already quoted, I have noted, I've wondered*), while these are extremely rare in the Blawg Corpus (e.g. *I have explained, I have written*), where this tense more frequently refers to personal existential or professional experience (e.g. *I have matured, I have received*, etc.), bringing it to bear on the problem being discussed in each case, as in the following example:

(12)   *I have spent* most of the Summer reading this and half a dozen other reports which either support or criticise the Human Rights Act. In not one single case, in any of the reports *I have read*, nor indeed, on this blog *have I seen*, anyone address the 'elephant in the room'. It is a subject which has become almost invisible in legal academic literature on the subject. The current paper argues that the 1998 Act is instrumental in protecting minorities. Is it? (HR, post)

This example shows very well how the discussion of a general issue (the actual effectiveness of the Human Rights Act 1998 to "ensure equal access to its provisions by the working poor") is cast in terms of personal experience. The author's point of view could have been formulated impersonally in the second sentence, e.g. *In none of the*

many reports which either support or criticise the Human Rights Act,
nor indeed, in entries to this blog addresses the 'elephant in the
room'. And it is all the more meaningful that the writer refers to the
Post itself as 'the current paper', thus betraying his perception of posts
on this website as professional/academic contributions, while at the
same time his choices in terms of language and register confer upon it
a personal/existential slant.

    In some other cases a relational/existential verb in the first-
person singular present perfect is used to construct an authoritative
image of the writer:

(13)   *I have been* a member of the MA for more than a quarter-century (MAG, post)

But probably the most meaningful element that emerges from the
comparison between blawgs and the control corpora is the
considerable number of first-person singular verb forms in the simple
past appearing in the former, which by contrast are quite sparse in the
latter.

    CLC has only five verb forms in the simple past (i.e. 3.40% of
all first-person singular forms). Since they all belong to the
metadiscursive and cognitive types, they all contribute to the
construction of the textual image of the author as researcher, as an
epistemic subject, mainly referring intertextually to his/her previous
research (e.g. "This international legal 'primitivism', *I argued*,
embodied the same kind of fear and fascination" (CLC jil47(4)06), "*I*
benefitted immensely from Jerry Frug's and David Barron's work on
similar issues" (CLC, hilj47(1)06).

    Quite significantly, in AAC first-person singular verbs in the
simple past are absent, with the only three instances embedded in
language reports. This – also in light of general considerations put
forth above on the kind of lexical verbs used – suggests that in awards
the focus is specifically on reasoning and deliberating. The self-image
that the arbitrator constructs in the text is that of the producer of the
final decision.

    These findings contrast with results from the Blawg Corpus, in
which 255 out of 1,045 first-person singular verb forms (i.e. 24.40%)
are in the simple past, thus providing evidence of a stronger personal

narrative component, which is slightly more marked in posts (27.88%) than in comments (23.19%).

Only in very few cases is the simple past utilized for metatextual/metadiscursive purposes (e.g. "As *I made* abundantly clear in my introduction, …"; " As *I pointed out* in my last post on the case….."), and always in posts, while this use is absent in comments.

Otherwise, in comments verbs in the first-person singular simple past are state or action verbs referring to the blogger's existential experience, which becomes part of the message. The beginning of a post in *Above the Law* on the real usefulness of law schools illustrates very well how the switch to personal narrative shifts the discussion of general issues to a personal perspective:

(14)   Is law school worth the tuition? Should I take out loans to go to a highly-ranked school, or accept a scholarship to a lower-ranked school? These are the burning questions that this website loves to pose.
I have opinions on these subjects like everyone else, but honestly, what do I know? The legal market was very different when I went to law school.
I attended The University of Pennsylvania Law School from 1996 through 1999. *I loved my* classes, my professors and my friends. Sure, law school was stressful, but, as I frequently quipped, it was better than work. (AL, post, Tom Wallerstein, 27/09/2012)

Here questions concerning the choice of the law school to go to, and the possible return on investing in a prestigious law school, are reinterpreted and discussed in light of personal experience. The author tells the story of his own professional life and considers the careers of his fellow students, to produce a general conclusion, again in personal terms:

(15)   I never bought into the idea that a law degree is a ticket to do anything you want. Obviously, that is even less true today than it was in the late '90s. But at least back then, if you went to what U.S. News deemed to be a top school, you had a pretty decent chance of ultimately guiding your career in the direction you wanted it to go. (AL, post, Tom Wallerstein, 27/09/2012)

This intersection between professional and personal emerges clearly also in the following example from the UK Human Rights Blog:

(16) Freemen on the Land are "parasites" peddling "pseudolegal nonsense":
Canadian judge fights back
September 30, 2012 by Adam Wagner
Meads v. Meads, 2012 ABQB 571 (Canadian) – read judgment / PDF

Almost a year ago, I and some other legal bloggers wrote about a phenomenon known as the Freemen on the Land movement. I called the post Freemen of the dangerous nonsense, for that is exactly what the movement is, for those desperate enough to sign up to it. Now a Canadian judge has done many judges around the world a huge favour by exploding the movement's ideas and leaders (or "gurus") in a carefully referenced and forensic 192-page judgment, which should be read by anyone who has ever taken a passing interest in this issue, and certainly by any judge faced by a litigant attempting the arguments in court. [..]
This is a long judgment, on the scale of a reasonably sized book. I will try my best to point out a few interesting bits but I would recommend that you read it. It is well set out and easy to follow. My numbered references are to paragraphs. […] (HR, post)

It is meaningful that the case is introduced by a personal narrative that highlights its potential interest, but also implies the blogger's ability to predict its importance. Introducing the actual discussion of the case (I will try my best …), he presents his report as a personal commented version of a lengthy text. The Post itself is long and complex, totaling 1,737 words. It is followed by a list of links for further reading and – as of October 3 at 9:07 pm – by 57 responses. Here is an example of a complete comment:

(17)   on September 30, 2012 at 1:32 pm | Reply Tom (iow)

Very welcome indeed, and although I'm sure that freemen true believers will dismiss it with the usual handwaving about how the judge did not have authority, it is useful to educate third parties with.
Also interesting to me is that the section on "foisted agreements" has some relevance to quasi-legitimate commercial practices as well, such as software "licence agreements", private parking companies, and some rogue employers. Some examples of their practices are not a whole lot different, except that in a few cases the courts seem to have approved of it.

Again, self-mention occurs at the beginning of the comment, presented as a personal contribution, although in the second paragraph

'the relevance of the section on foisted agreement to certain commercial practices' is linguistically presented as objective, by making recourse to presupposition: thus "*Also interesting to me is that the section on 'foisted agreements' has some relevance to quasi-legitimate commercial practices as well*" presupposes "*the section on 'foisted agreements' has some relevance to quasi-legitimate commercial practices as well*".

4. Conclusions

This chapter used corpus linguistics to investigate the generic integrity of blawgs (i.e. law blogs), one amongst many thematic blogs that have evolved in the past few years out of the (macro-)genre of the weblog. Originally introduced as a diaristic genre aimed at self-expression, self-disclosure and, more indirectly, community building, blogs are now also used for scholarly, professional and popularization purposes by various domain-specific discourse communities.

The analysis of self-mention in the texts of four law blogs, also comparing them for control with other two legal corpora, respectively academic and professional, has confirmed that, in spite of the profound evolution that has characterized blogs in the last two decades, their original distinctive feature, i.e. individualistic self-expression, has been preserved. This clearly appears from the texts as, in spite of the domain-specific character of the topics they deal with, law bloggers tend to avoid the de-personalized approach of legal academic and professional discourse, constructing instead a textual image of themselves by means of self-reference, and in particular first-person pronouns. At the same time, they also show a higher frequency of the second-person pronoun to address their interlocutors than the two corpora used for comparison, and this can be interpreted as an interactional device aimed at building a sense of community.

This interpersonal component characterizes posts and comments alike, although it is more marked in the latter owing to their semiotic configuration.

The analysis of the verb forms with which first-person singular pronouns collocate suggests that self-mention in blawgs is mainly used not only metadiscursively (as is mostly the case in the other two corpora), with reference to stages and stances in the text or to the blogger's previous writings, but also to express his/her views or existential experience. The discussion of legal issues is approached not only in intellectual or professional terms, but also in light of the individual's experiences and opinions.

This draws a picture of the blawg as a hyphenated micro-genre, instrumental to the needs of a domain-specific community, which at the same time maintains some of the original generic characteristics of the blog (macro-)genre. In some cases (e.g. in the Wall Street Journal Law Blog) this also means deliberately preserving some of the genre's peripheral features, so as to be promptly recognizable by users, reassuring them the blawg has not lost its original generic integrity and continues to be a blog.

References

Askehave, Inger / Ellerup Nielsen, Anne 2004. *Webmediated Genres – A Challenge to Traditional Genre Theory. Working Paper nr. 6.* Aarhus: Center for Virksomhedskommunication.

Askehave, Inger / Ellerup Nielsen, Anne 2005. What are the Characteristics of Digital Genres? – Genre Theory from a Multi-modal Perspective. *Proceedings of the 38th Hawaii International Conference on System Sciences*. Los Alamos: IEEE Press, 1-8.

Barger, Jorn 2007. Top 10 Tips for New Bloggers from Original Blogger Jorn Barger. <http://www.wired.com/culture/lifestyle/news/2007/12/blog_advice>.

Baron, Naomi 2000. *Alphabet to Email. How Written English Evolved and Where It's Heading*. London/New York: Routledge.

Berman, Doug A. 2006. Scholarship in action: The power, possibilities and pitfalls for law professor blogs. *Public Law*

and Legal Theory Working Series n. 6, Center for Interdisciplinary Law and Policy Studies, the Ohio State University Mortz College of Law. <http://ssrn.com/abstract= 898174>.

Berman, Doug A. 2007. More grist for the blog-scholarship debate [Blog post]. *Sentencing Law and Policy*, *March 19, 2007*. <http://sentencing.typepad.com/sentencing_law_and_policy/20 07/03/more_grist_ for_.html>.

Bhatia, Vijay K. 1993. *Analysing Genre: Language Use in Professional Settings*. Harlow/New York: Longman.

Bhatia, Vijay K. 2004. *Worlds of Written Discourse. A Genre-based View*. London: Continuum.

Blood, Rebecca 2002. Weblogs: A History and Perspective. In Rodzvilla, John (ed.) *We've Got Blog. How Weblogs are Changing our Culture*. Cambridge, MA: Perseus, 7-16.

Bruce, Ian 2010. Evolving Genres in Online Domains. The Hybrid Genre of the Participatory News Article. In Mehler/Sharoff/ Santini (eds), 323-348.

Caron, Paul L. 2006. Are Scholars Better Bloggers? Bloggership: How Blogs are Transforming Legal Scholarship. *Washington University Law Review*, 84/5: 1025-1042.

Carter, Locke M. 2000. Argument in hypertext: Writing strategies and the problem of order in a nonsequential world. *Computers and Composition.* 20, 3-22.

Cherney, Lynn 1999. *Conversation and Community: Chat in a Virtual World.* Stanford Ca.: CSLI Publications.

Crismore, Avon 1989. *Talking with Readers. Metadiscourse as Rhetorical Act.* New York: Peter Lang.

Degano, Chiara 2012. Argumentative Genres on the Web: The Case of Two NGOs' Campaigns. In Campagna, Sandra / Garzone, Giuliana (eds) *Evolving Genres in Web-mediated Communication*. Bern: Peter Lang, 99-129.

Garzone, Giuliana 2002. Describing E-commerce Communication. Which Models and Categories for Text Analysis?. In Evangelisti, Paola / Ventola, Eija (eds). *TEXTUS* (*English in Academic and Professional Settings*). 14/2, 279-296.

Garzone, Giuliana 2007. Genres, Multimodality and the World-Wide Web: Theoretical Issues. In Garzone, Giuliana / Catenaccio, Paola / Poncini, Gina (eds), *Multimodality in Corporate Communication. Web Genres and Discursive Identity.* Milan: FrancoAngeli, 15-30.

Garzone G. 2012. Where Do Web Genres Come from? The Case of Blogs. In Campagna, Sandra / Garzone, Giuliana (eds) *Evolving Genres in Web-mediated Communication.* Bern: Peter Lang, 227-253.

Giltrow, Janet / Stein, Dieter 2009. Genres in the Internet. Innovation, Evolution and Genre Theory. In Giltrow, Janet / Stein, Dieter (eds) *Genres in the Internet: Issues on the Theory of Genres.* Amsterdam: John Benjamins, 1-25.

Grieve, Jack / Biber, Douglas / Friginal, Eric / Nekrasova, Tatiana 2010. Variations Among Blogs: A Multi-dimensional Analysis. In Mehler/Sharoff/Santini (eds), 303-322.

Halliday, Michael A.K. 1994. *An Introduction to Functional Grammar*, 2nd ed. London, Arnold

Herring, Susan C. / Kouper, Inna / Paolillo, John C. / Scheidt, Lois Ann / Tyworth, Michael / Welsch, Peter / Wright, Elijah / Yu, Ning 2005. Conversations in the Blogosphere: An Analysis 'From the Bottom Up'. *Proceedings of the 38th Hawaii International Conference on System Sciences. Proceedings of the 38th Hawaii International Conference on System Sciences.* Los Alamos: IEEE Press, 1-11.

Herring, Susan C. / Paolillo, John C. 2006. Gender and Genre Variation in Weblogs. *Journal of Sociolinguistics* 10/4, 439-459.

Herring, Susan C. / Scheidt, Lois Ann / Wright, Elijah / Bonus. Sabrina 2005. Weblogs as a bridging genre. *Information Technology & People* 18/2, 142-171.

Hyland, Ken 2001. Humble Servants of the Discipline? Self-mention in Research Articles. *English for Specific Purposes.* 20/3, 207-226.

Hyland, Ken 2005. *Metadiscourse. Exploring Interaction in Writing.* London: Continuum.

Kerr, Orin S.   2006. Blogs and the Legal Academy, George Washington University Public Law and Legal Theory Working Paper No. 203, Available at <http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per _id=328150>, accessed 1 October 2012.

Krishnamurthy, Srikanth 2002. The multidimensionality of blog conversations: The virtual enactment of September 11. Paper presented at Internet Research 3.0, Maastricht, The Netherlands. October.

Lewiński, Marcin. 2010. *Internet political discussion forums as an argumentative activity type. A pragma-dialectical analysis of online forms of strategic manoeuvring in reacting critically*. Amsterdam: SicSat. Available online: at <http://dare.uva.nl /en/record/340631>, accessed 1 October 2012.

Marlow, Cameron 2004. Audience, structure and authority in the weblog community. Paper presented at the 54th Annual Conference of the International Communications Association, May, New Orleans <http://rockngo.org/wp-content/uploads/mt/ archives/ICA2004.pdf>.

Mauranen, Anna 2013. Hybridism, Edutainment, and Doubt: Science Blogging Finding its Feet. *Nordic Journal of English Studies* 13/1: 7-36.

Mehler, Alexander / Sharoff, Serge / Santini, Marina (eds) 2010. *Genres on the Web: Computational Models and Empirical Studies*. Dordrecht: Springer.

Miller, Carolyn R. / Shepherd, Dawn 2004. Blogging as Social Action: A Genre Analysis of the Weblog. In Gurak, Laura *et al.* (eds) *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*. <http://blog.lib.umn.edu/blogosphere/blogs_as_vir-tual.html>, retrieved 04/08/2010.

Miller, Carolyn R. / Shepherd, Dawn 2009. Questions for Genre Theory from the Blogosphere. In Giltrow, Janet / Stein, Dieter (eds) *Genres in the Internet: Issues on the Theory of Genres*. Amsterdam: John Benjamins, 263-290.

Myers, Greg 2009. *The Discourse of Blogs and Wikis*. London: Continuum.

Pistolesi, Elena 2004. *Il parlar spedito. L'italiano di chat, e-mail e SMS*. Padua: Esedra.

Puschmann, Cornelius 2009. Lies at Wal-Mart. Style and the Subversion of Genre in the *Life at Wal-Mart* blog. In Giltrow, Janet / Stein, Dieter (eds) *Genres in the Internet: Issues on the Theory of Genres*. Amsterdam: John Benjamins, 49-84.

Riboni, Giorgia 2010. *Comunicazione e interazione discorsiva in Rete: Il caso dei blog politici*. Doctoral dissertation, supervised by Giuliana Garzone. Dottorato di Ricerca in Anglistica. Università degli Studi di Milano.

Santini, Marina 2010. Cross-Testing a Genre Classification Model for the Web. In Mehler/Sharoff/Santini (eds), 87-128.

Santini, Marina / Mehler, Alexander / Sharoff, Serge. 2010. Riding the Rough Waves of Genre on the Web. In Mehler/Sharoff/Santini (eds), 3-30.

Scott, Mike 2011. *Wordsmith Tools 5.0.* Oxford: Oxford University Press.

Shapira, Ian 2007. No Objections Here. Supply-and-Demand Has Top Law Firms' 'Summer Associates' Hitting Pay Dirt Without Breaking Much of a Sweat. *The Washington Post*, Tuesday, 24/07/2007. <http://www.washingtonpost.com/wp-dyn/content/article/2007/07/23/AR2007072301890_pf.html>.

Sharoff, Serge 2010. In the Garden and in the Jungle. In Mehler/Sharoff/Santini (eds), 149-166.

Solum, Lawrence B. 2006. Blogging and the transformation of legal scholarship. Paper presented at *Bloggership: How blogs are transforming legal scholarship* <http://ssrn.com/paper=898168>.

Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

van Dijk, Teun A. 1992. Racism and argumentation: 'Race Riot' Rhetoric in Tabloid Editorials. In van Eemeren, Frans H. *et al.* (eds). *Argumentation Illuminated.* Dordrecht: Foris, 242-259.

Vestergaard, Torben 2000. That's not news: persuasive and expository genres in the press. In Trosborg, Anna (ed.) *Analysing Professional Genres*. Amsterdam: John Benjamins, 97-111.

Volokh, Eugene 2006. Scholarship, blogging and trade-offs: On discovering, disseminating and doing. Paper presented at

Bloggership: How blogs are transforming legal scholarship. *Washington University Law Review*, 84/5, 1089-110 <http://digitalcommons.law.wustl.edu/lawreview/vol84/iss5/5>.
Walker, Jill 2003. Blog (Weblog). *Routledge Encyclopedia of Narrative Theory*. London: Routledge, 45.
Westin, Ingrid, 2002. *Language Change in English Newspaper Editorials*. Amsterdam: Rodopi.