Apples and Oranges
A comparative analysis of blogs and marketing texts which share an audience

Joseph McVeigh
Master’s Thesis
ENG800
Department of Modern Languages
University of Helsinki
April 2013
Contents
1. Introduction .......................................................................................................................... 1

2. Background .......................................................................................................................... 5
    2.1 Corpus linguistics and genre analysis ............................................................................... 5
    2.2 Webinars, email marketing, and blogs ............................................................................. 9

3. Materials ............................................................................................................................. 15
    3.1 Marketing sub-corpus ........................................................................................................ 16
    3.2 Blog sub-corpus ................................................................................................................ 19
    3.3 Dealing with a very specialized corpus ........................................................................... 20

4. Methodology ........................................................................................................................ 23
    4.1 A genre defined by users and a faceted classification ...................................................... 23
    4.2 Tagging and bagging: Steps in the corpus linguistic analysis ............................................ 29
    4.3 Steps in the genre analysis ............................................................................................... 32

5. Corpus Analysis ................................................................................................................... 34
    5.1 Nouns and adjectives ......................................................................................................... 35
    5.2 Verbs and adverbs ............................................................................................................. 38
    5.3 Pronouns .......................................................................................................................... 40

6. Topic analysis ....................................................................................................................... 41
    6.1 DISCRIMINATION at work – nouns and adjectives .......................................................... 42
    6.2 DISCRIMINATION at work – verbs and adverbs ............................................................... 44
    6.3 DISCRIMINATION at work – pronouns ............................................................................ 47
    6.4 Qualifying DISCRIMINATION at the workplace ............................................................... 49
        6.4.1 Register Now: Genre analysis of NCCC ................................................................. 51
        6.4.2 Listen: Genre analysis of Strafford ......................................................................... 55
        6.4.3 Cat’s paws and going fishing: Genre analysis of E&theL ........................................ 58
        6.4.4 Get a grip: Genre analysis of SYGIGH ................................................................. 63
        6.4.5 Short and sweet: Genre analysis of WM .............................................................. 66

7. Conclusion and Discussion .................................................................................................... 69

References ................................................................................................................................ 74

Appendix 1 ................................................................................................................................ 77
    Corpus figures and frequencies ............................................................................................. 77

Appendix 2 ................................................................................................................................ 89
    Examples of the webinar landing pages and blog posts ....................................................... 89
1. Introduction

Marketers create texts that they want people to read. Popular bloggers create texts that people read willingly. Searching for the reasons why people flock to one type of text instead of another could be like searching for a needle in a haystack. But what happens when a blog post and an advertisement share the same topic and are aimed at the same community? How then does the blogger’s language compare to the marketer’s? Is there a noticeable difference between the two genres?

Corpus linguistics and genre analysis are two ways to answer these questions, but neither method comes without problems. While some have noted the beneficial aspects of using an annotated corpus, such as offering a more representative sampling of language (Biber & Conrad 2001), others have discussed the limitations involved with annotations that are not sensitive to certain linguistic features (Archer 2012). This problem of missing crucial information that the annotation is not designed to recognize is not a minor one, the danger being that an analysis of just the lexico-grammatical properties of a corpus will only tell part of the story. Complementing the corpus analysis with genre analysis can help remedy this problem, but the research then threatens to become untenable since others have noted that genres are dynamic and full of variation along many parameters (Tardy 2011, Swales 1990).

In this thesis, I will analyze and compare the language of marketing texts and blogs. I have created a corpus which contains texts from both genres and every text in my corpus has the same general topic, which is labor and employment (L&E) law in the United States. I will start with a corpus linguistic analysis of the texts to give a general overview of their linguistic properties and to see how texts from the sub-corpora of marketing and blogs relate to each other. I will then identify certain texts with the same
specific topic from each of the sub-corpora and examine them using a combination of corpus linguistic and genre analysis methods.

I hope to gain a better insight into how marketing texts and blogs are directed to readers. My specific research questions are:

1. How do the marketing and blog texts differ in their lexico-grammatical features and rhetorical move structures, and how are these related to their goals?
2. Based on the analysis, is the comparison of marketing and blog texts akin to a comparison of apples and oranges? Are the texts simply too different despite the fact that they share some important aspects such as the discourse community and topic?

It is tempting to assume that the texts in my corpus would share many linguistic properties since they share a macro-topic (L&E law) and they discuss it with the same audience (L&E lawyers) and that any differences among them would be glaring.

According to Swales, however, this is not the case. A close reading using genre analysis is therefore not only called for but necessary to show specifically how the different genres (marketing and blogging) compose their texts differently.

Using genre analysis to find rhetorical moves that would be missed by the corpus linguistic analysis also solves another problem, which is how researchers can recognize the differences among the texts within each genre. Tardy (2011: 57) has noted that “as socially situated ways of communicating, genres must also be somewhat dynamic and changeable in relation to their users, uses, and other contextual factors.” This means that I can expect to see differences not only between the two marketing and blogging genres, but also between the sub-corpora in each genre. Biber and Conrad (2001: 1) similarly note that

Because registers vary in the extent to which they are specified situationally, the texts within registers also vary in the extent to which their linguistic characteristics are similar. At one extreme, texts from a specialized register tend to be very similar in their linguistic characteristics, corresponding to the extent
to which the register is specified situationally. In contrast, the texts in a general register, such as academic prose or fiction, tend to exhibit a wide range of linguistic variation.¹

Obviously the extent of this variation in the linguistic characteristics of texts from the same genre is only evident after the analysis has been done. It nevertheless can be expected. In describing the possible chaotic nature of genres, Bhatia (2004: 29) says that “although it is true that most genres have an identifiable ‘integrity’ of their own, it may often be seen as unstable, even ‘contested’ sometimes”. This “unstable” or “contested” nature should not, in my view, be seen as a problem. Rather than predicting that the analysis will be muddled, using the methods offered by genre analysis, in combination with the corpus linguistics methods, will give a clearer and more detailed picture of the texts in my corpus and, by extension, the genres they represent.

In a general sense related to my first research question, the topic of this thesis is interesting and important because it offers benefits to both linguists and marketers.

Herring and Paolillo (2006: 440) note that “weblogs deserve linguistic study, especially if language use differs in blogs compared to other genres”. As a linguist studying blogs, I know that the findings in my analysis section help to further define the sprawling nature of the blog genre, which can seem impenetrable to linguists venturing an analysis of it. The analysis not only describes a specific sub-genre of blogs, it can also be of future use to linguists studying similar single-topic blogs by offering a way to describe their linguistic properties.

For the marketing aspect, researchers such as Kim (2007) have noted how advertising provides vast possibilities for linguists to study discourse strategies and the ways that language is used for persuasion. As a former copy writer, I know that my analysis can

¹ I am aware that there is a debate about the differences between “register” and “genre” (and even “text-type”), but it is outside the scope of this thesis and will therefore not be discussed. What exactly I mean by “genre” will be made clear in a later section.
aid marketers in getting to know their genre better and, hopefully, in writing better marketing copy in the future. I believe that for anyone serious about marketing and/or linguistics, it is not enough to say that marketing is marketing and blogging is blogging, especially in cases where marketers and bloggers are discussing the same topic. What is needed instead is a comprehensive analysis of each genre to better understand the ways that they present their topics to their communities.

In a more specific sense related to my second research question, there is (to my knowledge) a gap in previous genre research in that it has not addressed the lexico-grammatical and rhetorical properties of texts on the same topic which are addressed to the same community, but which have very different goals. For the blogs, we can say their goal is mainly explanatory, while the marketing texts are obviously promotional. But how are the differences in the texts reflective of these goals? And if these genres are related in terms of their topics and communities, shouldn’t we expect them to have influences on each other?

Since I used to work as a webinar producer for one of the marketing companies that I drew texts from for this thesis, I can state for a fact that the blogs in my corpus do have an influence on the marketing texts. Webinar producers working in the L&E law area were encouraged to find topics that would interest L&E lawyers by reading the L&E law blogs, precisely because the audience for the webinars was the same as it was for the blogs.

The influence, however, cuts both ways. While I cannot say that the marketing texts in my corpus influenced the blogs with the same assurance that I can say the blogs influenced the marketing texts, some researchers have noted the way in which the

---

2 “Webinars” are online seminars or learning courses offered to any sort of business professional. They will be explained in more detail below.
linguistic properties of advertising have invaded other genres. Bhatia (2004: 84) writes that

> Appropriation of lexico-grammatical and rhetorical resources from the discourse of corporate advertising has offered a number of very attractive options because of its innovative character and creative use of language. Advertising has turned the process of writing into an art form, where writers constantly compete for attention not only by innovative use of language but also by the creative use of traditional expressions and clichés, which are often shunned by good writers in other forms of discourse.

These cross influences, one very specific and the other more general, become heightened when we remember that the goals of the texts are different. While the blogs aim to explain or at the very least discuss L&E law, the marketing texts are designed specifically to sell a product. These goals are not a mystery to the community and they are ultimately the way in which the texts are judged. For the blogs, their high regard among the community shows that they are successfully reaching their goal. For the marketing texts, high sales of the webinar in one text mean that the linguistic properties – in some form or another – will be appropriated into other marketing texts.

### 2. Background

#### 2.1 Corpus linguistics and genre analysis

Corpus linguistics is the branch of linguistics that deals with analyzing data banks of texts, or observations of linguistic events. In general, corpus linguistics is an attempt to study language use empirically, as opposed to intuitively. Aarts (2011) writes that “corpus data reflect what people actually say and write, and as such provide the most appropriate data for linguists who want to investigate the *use* of language rather than linguistic competence or linguistic universals” (italics original). In this way, corpus linguistics is perfect for my study since I am analyzing how bloggers and marketers use
language when writing about the same topic and not which of them uses language better.

Aarts (2011) also defines what a corpus is by noting that:

for a collection of linguistic events to be a corpus, it has to meet minimally two conditions. The first is that it should present a faithful record of the utterances contained in running texts (rather than, say, a collection of examples of a particular linguistic phenomenon), the second is that it should give information about the questions by whom, where, when and why the texts were produced. (italics original)

The most common way corpora are studied is by annotating each word and text in the corpus with meta-information. On the meta-textual level this could include the author, the date each text was produced, or the type of text (sermon, news article, etc.). Another level of meta-information that is usually appended to the texts involves using a part-of-speech (POS) tagger, which is a type of software that automatically attaches the relevant part of speech to each word in the corpus. These are more or less the methods followed by popular large corpora such as the British National Corpus and the Corpus of Contemporary American English. They are also the methods I followed, not because they are also used by these large corpora, but because they fit my research questions well and facilitate the analysis of my corpus. The specific meta-information and POS tagger I used are described below in the Materials and Methods sections, respectively.

The reason that scholars have used the term genre since the time of Aristotle is because, as Solin (2011: 119) writes, “Genre interests scholars across disciplines because it allows looking at products of culture not merely as forms, but as social and symbolic action”. Genres can fluctuate and change as the practitioners expand or contract the limits of what is acceptable in the genre. As writers and orators and even everyday

---

3 For example, the word “boy” will come out of the tagger as “boy_NN1”, where “_NN1” is the tag for a singular common noun. This allows for easy retrieval of parts of speech using a program that can search for the tags. The specific tagger that I used will be described in the Methodology section (chapter 4) below.
speakers have expanded the realms of literature and speaking and discourse, so have scholars expanded the genres they invented to name and study them.

There is a problem, however, in just how to define the term “genre”. Devitt (2004: 8) opts for a less strict description and writes that “rather than making the concept so broad as to be useless, as some have argued, allowing such flexibility in the definition of genre for scholars keeps the concept fluid and dynamic, able to respond to scholars’ changing needs over time”. So the term genre does not need to be either unwieldy or constrictive, it merely needs to be flexible and adaptive to the study at hand.

But this does not solve the problem of whether or not such a definition can justifiably classify texts into a genre based on some necessary elements that must be present. For example, Swales (1990: 45) writes that a genre is “a class of communicative events”. This definition is perhaps too simple for my purpose here, but it does show that communication is essential in genres. Yates and Orlikowski (1992: 301) write that genres are “typified rhetorical action in the context of socially defined recurrent situations”. This definition forms a better basis on which to analyze the texts in my corpus because it recognizes both rhetorical actions and (recurrent) social situations.

What is still needed, however, is the notion that every text has a specific goal. Swales (1990: 46) anticipates this by writing that

The principle criterial feature that turns a collection of communicative events into a genre is some shared set of communicative purposes. Placing the primary determinant of genre-membership on shared purpose rather than on similarities of form or some other criterion […] is based on the assumption that, except for a few interesting and exceptional cases, genres are communicative vehicles for the achievement of goals.

It will be important to remember this “achievement of goals” aspect when comparing the use of language in each of the genres in this thesis since it is one of the main ways in which the genres in my corpus differ.
Another feature of genres is that the classification of texts based on their genre is not inherent to the texts themselves. On the contrary, as Biber et al. (1998: 170) note, “Genre categories are determined on the basis of external criteria relating to the speaker's purpose and topic; they are assigned on the basis of use rather than on the basis of form”. But this approach of imposing genre classifications on texts is not without its problems. Herring (2007) writes of one of these problems when she says that “the concept of genre can potentially be applied to communication at different levels of specificity”. So the question for researchers becomes how to assign the appropriate classification when dividing texts into genres. We can see how quickly this could become a problem just by asking whether this thesis should be classified as a linguistics paper, a corpus linguistics study, a genre analysis study, a Master’s thesis, or something else? Studies which attempt to group certain texts based on their genre, such as this one, are then faced with an even deeper problem. As DeVitt (2004: 7) writes, “how texts are grouped depends on which features the classifier has selected to observe […] Defining genre as a kind of text becomes circular, since what we call a kind of text depends on what we think a genre is”.

Fortunately, Swales (1990: 54-5) offers researchers a way out of this dilemma by writing that active discourse community members tend to have the greatest genre-specific expertise – as we often see in interactions between members of a profession and their client public. One consequence is that these active members give genre names to classes of communicative events that they recognize as providing recurring rhetorical action.

The problem of classifying the texts into one genre or another is therefore taken out of the researcher’s hands. The definition of genre is left to the discourse community who recognize the communicative purposes of a group of texts. The problem of circularity is avoided because, as Swales (1990: 54) notes “These purposes are recognized by the
expert members of the parent discourse community, and thereby constitute the rationale for the genre”.

A final problem involved with modern genre studies that needs to be solved is making sure the comparison is of two actual genres. Yates and Orlikowski (1992: 310) write that

the concept of medium has often been confused with that of genre. Confusion arises when researchers compare genres of communication (e.g., memos and bulletins) with communication media (e.g., electronic mail or fax). […] Thus, comparing memos with electronic mail, for example, confounds the concept of communication medium with that of communication genre.

It should therefore be pointed out that in this thesis I will be comparing the genre of L&E law webinar marketing, which is transmitted through email, to the genre of L&E law blogs, which are transmitted through the internet (web pages). Marketing and blogging are the genres; email and the internet are the media.⁴ I will show below, however, that the texts in my corpus share many medium factors as well and these will help in classifying them.

### 2.2 Webinars, email marketing, and blogs

Before going on, some descriptions of the terminology that will be used in this thesis is necessary. The term webinar is a portmanteau of “web” and “seminar” and it basically refers to an online course where audience members view a presentation on their computers while listening to an instructor over the phone or internet. The technology involved allows a moderator to restrict sound from the audience during the presentation, so that only the instructor’s voice can be heard. It also allows the moderator to set up a one-on-one discussion between audience members and the instructor during a question and answer session. These Q&As can be live and heard by the rest of the audience.

⁴ The blog texts can also be transmitted through email in the form of website subscriptions or Really Simple Syndication (RSS) readers, which present posts from web pages in an email client-like fashion.
These events are sometimes also called web seminars, online seminars, web conferences, and audio conferences.

Email marketing is simply a form of advertising done through email. Brownlow (2013) says about email marketing that “In its broadest sense, the term covers every email you ever send to a customer, potential customer or public venue.” Brownlow then lists three types of email marketing, the first of which, direct email, is the only one relevant to this study. Direct email, according to Brownlow (2013), “involves sending a promotional message in the form of an email.” This message is designed to sell a product or service and the text in the email can include various forms. The email will always have a hyperlink that will take readers to a webpage, called a landing page, where they will be able to purchase the product or service advertised in the email. The marketing texts in my corpus, for example, which are also marketing emails since that is how they are transmitted, include some (but not all) of the language that is on the landing page of the webinar that they are advertising. The extra language on the landing page, and my reasons for not including it, will be discussed below.

Email marketers often place a very high premium on the subject line (also called headline or heading) of their marketing texts. The idea is that this is the line of text that will get potential customers to open the emails and follow the hyperlink (or “clickthrough”) to the product’s landing page. According to Brian Clark (2013), founder of Copyblogger, a company which instructs copywriters how to compose better content, “On average, 8 out of 10 people will read headline copy, but only 2 out of 10 will read the rest. This is the secret to the power of your title, and why it so highly determines the effectiveness of the entire piece.” This emphasis on the subject line, however, is usually tempered with the acknowledgement that the main body of the marketing text is just as important and that the two sections actually complement each other. The subject line
piques readers’ interest and the marketing body fulfils their desire for the product offered. For example, McGlaughlin et al. (2012: 11) say that “we must understand the connection between our subject line and the value proposition [or, body of the marketing text].” Although these claims come from normative literature instead of academic research, they are nevertheless important because they show the value that marketers place on both the subject lines and the bodies of the texts. As such, I have not analyzed the subject lines of the marketing texts separately from their corresponding bodies.

While there was a high premium placed on the language used in the marketing copy at the webinar company I worked for, there were never any statistics on what type of language worked best. Of course, producers knew which webinars had sold best and they would sometimes try to imitate the language used, but such practice is more emotional than empirical. Likewise, the supervisors who acted as editors of the marketing texts often had years of experience writing copy, but their ideas of what language worked best were simply based on intuition. Fortunately, the empirical nature of corpus linguistics can help marketers base their claims about copy on hard evidence, in much the same way that corpus linguistics helps keep linguists out of their proverbial armchairs.

Discourse in advertising is frequently studied by first explaining whether the ads in question are soft-sell ads, which appeal to a potential customer’s emotions, or hard-sell ads, which offer a more direct approach to presenting the product. Kim (2007: 95) says that “Reason (or hard-sell) advertising attempts to persuade the consumer by providing direct and easily decipherable reasons for the purchase.” Similarly, Bhatia (2004: 63) says that “advertisements of hard-sell type make use of a number of typical rhetorical moves to persuade potential customers to buy the product or service they promote, the
most common of them being ‘product appraisal’.” All of the ads in my corpus are examples of hard-sell advertising and an example of their move structure will be given in the Methodology section (chapter 4) below, while specific examples from the corpus will be discussed in the Analysis section (chapters 5 and 6).

Bhatia makes a further distinction between high- and low-involvement goods. He writes

High-involvement goods are those which generally tend to be higher in price, are purchased relatively infrequently, and require some research and reflection on the part of the consumer before making a purchase. Low-involvement goods tend to be packaged goods of a relatively low price, which are purchased by the consumer in supermarkets and convenience stores. (2004: 98)

The marketing texts in my corpus, therefore, are examples of advertising of high-involvement goods since webinars are neither relatively cheap nor are they purchased in places like a supermarket.

A final note about advertising discourse is related to the cross-influences and goals of the texts mentioned above. Bhatia (2004: 63) claims that

Advertising, on the one hand, is the most traditional form of promotional activity, which is intended to inform and promote in order to sell ideas, goods or services to a selected group of people; on the other hand, it is also one of the most dynamic generic forms exhibiting some of the most innovative uses of lexico-grammatical and discoursal forms and rhetorical strategies.

But innovation in genre is relative. As mentioned above, the marketing texts in my corpus draw specifically on the blogs in the corpus. Any discussion of innovative lexico-grammatical, discoursal, or rhetorical form in them should therefore be moderated by the fact it may have been taken from the blogging genre. In the case of the blogs in my corpus, as will be explained below, they are similar to reporting genres, which means they may include linguistic forms from that genre. Since the marketing texts in my corpus are promotional, they may lift any lexico-grammatical, discoursal, or
rhetorical forms from the blogs, but these forms will be repurposed to fit the goal of the texts, which is to sell a L&E law webinar.

Blogs, short for weblogs, are ubiquitous across the internet. Herring et al. (2004: 1) define blogs as “frequently modified web pages in which dated entries are listed in reverse chronological sequence.” This very basic definition is a good starting point, but it is perhaps too general for this thesis. The problem is that the varied nature of the web pages that people call blogs does not lend itself to a clear definition. Blood (2003) and Myers (2010) note that hyperlinks in the text and the ability to comment on a post are central to any website that researchers wish to label as a blog, but even these aspects are not absolute requirements. In the next chapter I will show why the blogs in my corpus are correctly identified as blogs. For now, instead of trying to define what is and what is not a blog, I will instead define what the term means in this thesis by explaining a few special characteristics of the texts that I am applying it to.

First, the blogs in my corpus all include the possibility for readers to comment and they all have hyperlinks in at least some of their texts. They are arranged in the reverse-chronological format that is so definitive of blogs (Blood 2003, Herring et al. 2004). Furthermore, they incorporate two types of blogs that Blood (2003) defines: they are (1) “written by professionals about their industry” and (2) they “link primarily to news about current events”. The case of linking to news about current events is not as prevalent as the first case, since although almost every post in my blogs is about a current event in the L&E law field, the authors do not always link to a news story or report of it. The hyperlinking and commenting aspects of the blogs will not be discussed in this thesis because the focus is on the rhetorical nature of the blog texts and their titles in comparison to the marketing texts.
Perhaps more important than the hyperlinking and commenting aspects is the way in which the blogs in my corpus are single-topic. This cannot be overstated. The authors of the blogs write about labor and employment law in the United States and only labor and employment law in the United States. They naturally incorporate other fields or social areas into their discussion (for example, baseball or literature), but their discussion is firmly focused on law. This is an important aspect of the blogs in my corpus not only because it defines them, but because it is an aspect that is (to my knowledge) left unaddressed by researchers. Whereas previous literature focuses on the “blogs at large” genre, this thesis can perhaps best be described as focusing on the “single-topic blog” sub-genre. The position of the single-topic L&E law blogs, the specific sub-genre that I am analyzing, is shown in Figure 1.

Figure 1. The position of the L&E law blogs as a sub-genre.
The figure shows that L&E law blogs are a sub-genre of law blogs, which is a sub-genre of blogs, which is a genre in online communication.

3. Materials

The texts in my corpus are blog posts and marketing texts. Although these two types of texts make up the corpus as a whole, it is perhaps better to view them as two separate sub-corpora since the research here is comparative and since they will be treated as separate in my analysis. The reason they will be treated separately is because, as mentioned above, I am expecting there to be variation both between the genres and between the texts in the genres. This means that I am expecting both of the marketing sub-corpora to show variation in comparison to each other and in comparison to the blog sub-corpora. Likewise, I expect each blog sub-corpus to show variation in comparison to the marketing sub-corpora and the other blog sub-corpora. When creating a corpus, the aspects that affect the decision of which texts to include are, among other things, quantity, diachrony vs. synchrony, and text. These aspects are not mutually exclusive. Instead, they are addressed together by deciding which texts will best help the researcher answer their questions. The goal is to construct a corpus that is representative of the subject that is being studied, and representation does not always mean the most texts from the longest time period and widest origin.

I decided to focus on one year’s worth of texts for two reasons. First, although a calendar year may be arbitrary, it is relevant in the business world, especially when companies analyze their success and report their profits. Second, I reviewed the sources and recognized that a year’s worth of text from both the marketing companies and the blogs would be a sufficient amount of language to base a study upon. This is not to say
that more texts would not be helpful (and that is certainly open for future research), just that drawing the line at one year seemed both logical and justified.

Finally, it is important to state that my corpus is specialized. It includes only marketing and blog texts from certain sources and it is being used to make a comparison. This means that concerns about representativeness, balance, and size do not bear on my corpus as much as they would on other general reference corpora. For example, a corpus being used to make claims about language per se would need to include enough different kinds of texts and be large enough that it could be considered as representative of the language in question. My corpus, on the other hand, is specialized in that it is only being used to make claims about the texts which it includes. These texts are examples of blogging and email marketing, but they do not represent either as a whole.

### 3.1 Marketing sub-corpus

The texts in my marketing sub-corpus are drawn from two webinar firms, National Constitution Center Conferences (NCCC) and Strafford. NCCC was chosen because I used to be a webinar producer there. Strafford was chosen because it is one of NCCC’s biggest competitors. There are eleven texts from NCCC and nineteen from Strafford. The texts constitute every webinar advertisement that was aimed directly at L&E lawyers from both companies in 2011, the same year that my blog posts are from. When I say that they were aimed directly at L&E lawyers, this means that each text was sent as an email, either in whole or in parts, at least once and that L&E lawyers made up the bulk of the email recipients. As these are email marketing examples, customers interested in the marketing email were sent via a hyperlink in the email to the webinar’s landing page, which is where I drew the texts from. The companies know which potential customers in their email list are L&E lawyers based on such factors as
previous attendance in an L&E law webinar, information provided when signing up to receive email alerts, and email lists that the companies may have purchased. Obviously, who exactly the emails were sent to is a company secret, and other types of lawyers would have also received the same marketing emails. But the emails were targeted in that more emails were sent to those who were more likely to purchase the webinar based on their job title, i.e. L&E lawyers made up the majority of recipients of the marketing texts in my sub-corpus. For each webinar, hundreds of thousands of emails were sent. Most importantly, the same marketing text was sent to everyone.

Furthermore, the marketing texts from both companies are included on their websites under the appropriate labor and employment law headings ("Employment and ERISA" for Strafford and "Employment" for NCCC). That is, each company posts the marketing texts for their webinars at least one month in advance of the live webinar date (usually when the emails start) and they have a certain page which lists all of their previous and upcoming L&E law webinars.

As noted earlier, I used to work as a webinar producer for the company which runs NCCC. The construction of the NCCC marketing texts works like this. The first part of a L&E law webinar producer’s job is to research a field of business and come up with a topic for a webinar that they think L&E lawyers would purchase. After the topic for a webinar is approved by their supervisor, the producer writes copy for the marketing. The marketing copy goes through many revisions. Since the marketing is done entirely through email, the language in the copy is pored over. For example, when submitting copy for the first time, webinar producers are required to write seven to ten phrases to appear in the email subject line. After that, it is not unusual to rewrite the marketing
copy (the email body) three to four times. The idea is to not only use language that would tell the customer what material would appear in the webinar, but obviously to also use language that would encourage customers to feel the need to purchase the webinar, which is the style common of hard-sell advertisements. Sometimes this means using drastic warnings to scare the customer into thinking there might be legal ramifications if they do not make changes based on the information provided in the webinar and sometimes this simply means using hot button terminology, such as the names of recent court cases or updates to federal law.

Each marketing text in my sub-corpora (from both NCCC and Strafford) also includes a section that describes the speaker, who is a practicing lawyer who does not work at the company which markets the webinar. The marketing copy is influenced by the webinar speakers only to the extent that the webinar producer and speaker would agree on exactly what would be covered in the webinar. Webinar producers, who had done the initial research on the topic, were encouraged to find speakers who could both present the issues that the producer wanted covered and who were “celebrities” in their fields. In the case of L&E law, this often meant that the speaker worked for a big name law firm that L&E lawyers would recognize. Other than that, the marketing copy is left to the webinar producer and their supervisor(s). In terms of content, which obviously would be reflected in the copy, producers did not just allow the speaker to dictate what would be in the webinar. Usually, the producer simply told the speaker what to cover and the speaker agreed since both wanted to cover the pressing issues in the allotted time. This process is more straightforward than it seems. If the speaker does not agree to cover what the producer wants covered, another speaker is found. The benefit to the

---

5 Obviously, the amount of rewrites I performed may just show that I am a poor copywriter.

6 The notion of “celebrity lawyer” or big name law firm will be addressed below in the discussion of rhetorical moves.
speaker is that they could be seen as an expert in the field and add the webinar presentation to their resume, two things which may lead to more speaking positions (and higher speaking fees) in the future.

The final aspect of the content to discuss is “CLE”, a term which is common in the marketing texts in my corpus. It stands for “continuing legal education” and it refers to the mandatory requirements of attorneys in the United States to maintain their license to practice law. The amount of CLE credits needed differs by state, but every lawyer is required to gain a certain number of CLE credits every year. It is important in this thesis because in almost every state, lawyers can receive CLE credits by attending webinars and the companies whose marketing I will analyze offer CLE credits to lawyers who purchase and attend their webinars. This offer is a rhetorical move that I will discuss below.

3.2 Blog sub-corpus

The texts in my blog sub-corpus come from four different blogs aimed at L&E lawyers. The blogs are Employment and the Law (E&theL), FMLA Insights (FMLAI), Screw You Guys, I’m Going Home (SYGIGH), and Work Matters (WM). I was fortunate in being able to avoid bias on which blogs to include by referring to lists of L&E law blogs that were generated by the discourse community. Two separate organizations, LexisNexis and the American Bar Association (ABA) Journal release lists of the best L&E law blogs every year. Most importantly, both the LexisNexis and ABA Journal lists of L&E law blogs are voted on by readers. In the case of the ABA Journal, which listed vote tallies for its blogs, WM and SYGIGH placed first and second. E&theL was chosen as the “Top Labor & Employment Law Blog of 2011” by LexisNexis readers, while FMLA Insights was on both lists (second on the LexisNexis list and third on the ABA list).

7 The ABA Journal calls these “blawgs”, but I’ll refrain.
Each blog is written by a practicing attorney and each is aimed at a national audience. I included every post from 2011 from each blog, with the exception of eight posts from *FMLA Insights* because they were written by a different author and I wanted to keep my data uniform. The exclusion of these texts will be discussed in further detail in the Methodology section below.

In order to give an idea of the size of the corpus, the number of texts and word counts of each blog and marketing text are in Table 1 below.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of texts</th>
<th>Word count (tokens)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCCC</td>
<td>11</td>
<td>4,759</td>
</tr>
<tr>
<td>Strafford</td>
<td>19</td>
<td>9,431</td>
</tr>
<tr>
<td>E &amp; the L</td>
<td>32</td>
<td>14,526</td>
</tr>
<tr>
<td>FMLAI</td>
<td>46</td>
<td>23,275</td>
</tr>
<tr>
<td>SYGIGH</td>
<td>64</td>
<td>28,398</td>
</tr>
<tr>
<td>WM</td>
<td>77</td>
<td>19,679</td>
</tr>
<tr>
<td>Totals</td>
<td>249</td>
<td>100,068</td>
</tr>
</tbody>
</table>

Table 1. General stats for my corpus. The thick line divides the corpus between the marketing and blog genres.

### 3.3 Dealing with a very specialized corpus

The first step I took in creating the corpus was to gather the texts by visiting the websites of the companies and the blogs I chose to analyze and copying the appropriate texts. For the web pages of the marketing companies, I took only the text that was relevant to the marketing. For example, and similar to any web page, the pages for the marketing texts in my corpus had sidebars and banners with information about the company and links that take users to different locations in the website. There is also an explanation of what a webinar is and instructions on how to register for one. Most importantly for the webinars aimed at lawyers, there is information on how attending the webinar can count as a CLE credit. I did not include any of this information since it was neither written by the marketers, nor would it have been in the emails sent to
potential customers. Rather, the emails include a link that takes interested customers to the landing page set up exclusively for the webinar advertised in the email, which is also the page where I took the text for each webinar. On that page, customers can find a link to the page where they can register (and pay) for the webinar. The landing pages for both of the companies naturally vary a bit, but they follow this general structure. The most important thing to note here is that the marketing text for the webinars from both companies is placed front and center on their respective landing pages. Examples of webinar marketing landings pages from both companies, complete with the sidebars and banners, are in Appendix 2.

I also excluded some other information from the marketing text I analyzed. First, I removed the price from the NCCC texts. Although price is an obvious factor involved in whether someone buys a product, it is not pertinent to this study, even though it is placed between the title and the marketing text body on the webpage. Second, where NCCC placed the price of the webinar or recording between the title and marketing body, Strafford placed the phrases “Recording of a 90-minute CLE webinar/teleconference with Q&A” and “Conducted on Wednesday, September 21, 2011 Recorded event now available”. I removed each of these, which was the same across each marketing text (with the exception of differing dates), because, like the price for the NCCC marketing texts, it is not important for the linguistic analyses I will make; it would not have been included in the emails sent to potential customers. Another reason for excluding these phrases is to keep them from skewing the quantitative analysis of the texts. It would be easy in a close reading to glance over them

---

8 If I had the sales figures for the webinars, then I could include their prices in my analysis, for example, by ranking each webinar based on its sales and then comparing the linguistic properties of the most successful webinars to the least successful ones.
and focus on the main body of the marketing text, but a word counting program would not be able to do so. It therefore makes more sense to exclude them from the start.

I did not exclude the information about the speaker (or presenter) of the webinar since, unlike the price and general information about webinars that I did remove, the speaker bio works as marketing copy. From my experience as a webinar producer, I know that this text often comes from the speaker themselves and it is always in the form of a short biography. It may have been written by the speaker for the specific webinar, but for lawyers it is more likely to have been pulled directly from the speaker's personal website or their personal page on the website of the law firm where they work. Each time a speaker is chosen to present a webinar, they are asked to submit a short bio, unless they have spoken for the company before, in which case it is common to see the same bio appear in the marketing for each webinar that the speaker presents, perhaps with slight modifications. Although the text is not written by the webinar producers, it is often edited by them (usually to cut down the length), which is one of the reasons why I included the speaker bios in my analysis. The other reason, as previously mentioned, is because the speaker bio is a marketing text in itself, designed in this case to sell the webinar by selling the speaker, which could mean either name-dropping a well-known law firm or government agency that the potential customers would recognize, or by mentioning that the speaker has experience presenting the topic of the webinar.

The marketing texts are grouped simply into two sub-corpora based on where they came from, either NCCC or Strafford, and then organized by date from the oldest to the most recent. The exact date of when each webinar was run is in the file name of each text.

For the blogs, I followed roughly the same process. After deciding which blogs I would draw the texts from, I took the title and body of every blog post that was published in
2011 on each blog. As there was on the webpages for the marketing texts, the blogs also had sidebars, banners, etc. to remove (or not take in the first place). I also did not include any comments because analyzing them is outside the scope of this thesis.

One common practice among blogs is to place publishing information between the title and the body of each post. The blogs in my corpus are no different and I removed this information, which for every blog post was either the author, the date the text was posted, or both. Three of the four blogs in my corpus have only one author, while the fourth, *FMLA Insights*, had two authors. I felt it was important to keep constant as many variables as possible in my analysis and so I only used posts from one of the authors of *FMLA Insights*. The author I used posted to the blog 46 times and posted every month of the year, compared with the author I excluded, who posted only eight times and only in the first four months of the year. The only other language removed from the blog posts was text related to images, such as captions and alt-text, which is a kind of second caption that appears when readers hold their mouse pointers over an image, an aspect made possible by HTML. Some of the blog texts included images, but they will not be discussed here except for mentioning that each and every one of them was a stock photo or image. The final aspect to mention is that the blog sub-corpus is divided into four parts, one for each blog.

4. Methodology

4.1 A genre defined by users and a faceted classification

Grouping the texts in a corpus can involve many considerations, but for me the process was relatively straightforward. The division basically came down to grouping the texts based on which blog or webinar company they came from since they otherwise share

---

9 The law firm that these two authors work for actually started another law blog in 2011, with the author I excluded as its main author and one of the posts in my corpus is a promotion for this new blog.
properties such as time period of production (2011) and topic (L&E law). I made a further distinction based on the genre of communication and this divided my corpus into two parts: blogs and advertisements. Thus the division based on genre is the major one made in this thesis, while the division based on origin (which specific blog or webinar company) is minor but also important for comparing sub-corpora within genres.

In section 2.2 above, I touched on the problems associated with using the term genre, most notably that of falling into circular reasoning by defining a genre with the texts which are being analyzed and then using this definition to divide the texts into different genres. But as Devitt (2004: 7) notes, “that conundrum does not mean that genres do not involve classification nor that devising a classification system is necessarily a waste of time. There are purposes for which classifications systems are helpful”. Fortunately for this analysis, a definition of genre and a defense of it will not be necessary, since there is a simple way around the circular reasoning problem that affects researchers who attempt to divide their data into genres based on their intuitions. However exhaustive a researcher might be in defining and defending the genres they place their data into, they can never reach the legitimacy of genres that have been defined by their community of users. This is the mainstream approach to using genre. Devitt (2004: 8) writes that

Most current rhetorical genre scholars base their analyses of genre not on the classifications of critics and analysts but on the ways people classify texts into genres as they use them. Carolyn Miller argues for analyzing the everyday genres that people use, and by 1997 David Russell takes as a given that participants’ recognition of a genre is what rightly determines whether one genre is distinct from another […] Genre has been redefined, then, from a classification created by critics to a classification that people make as they use symbols to get along in the world.

This is precisely the basis upon which I will classify my texts. Not only have the users created the genres for me, but they have also already qualitatively classified the texts I am using into those genres. The fact that the texts in my blog corpus come from blogs
which were voted as the best L&E law blogs means that they are construed as L&E law blog texts. Such a definition may sound redundant, but it is important to realize the way it helps this analysis avoid the circularity problem. The users of the L&E law blog genre have defined the genre. If a blog was not a part of that genre, it would not have been considered in either of the best L&E law blog lists because it would have been rejected by the users. So instead of collecting texts and then grouping them based on their genre and defending my reasons for grouping them, the process is inverted by collecting texts which have already been grouped by the users of the genre I wish to analyze.

The grouping of my marketing texts into the genre of L&E law webinar marketing is slightly different, but no less objective. While the genre has not been explicitly defined by the users through a ranking system such as “best L&E law webinar marketing texts”, it has been explicitly defined by its creators. The marketers placed the texts in the “L&E law webinars” section on their websites. These are the marketing texts that are sent to people who register to receive emails about upcoming L&E law webinars (or to those who have purchased an L&E law webinar, since submission of an email address is required to attend the webinar). Furthermore, the inclusion of CLE credits makes the intended audience lawyers in a broad sense, but when added to the classification of the marketing texts into the L&E sections on the websites, the intended audience becomes specifically L&E lawyers.

Even though the grouping of my texts has been justified by the users of the genres which I am analyzing, it is prudent to add an extra layer of justification for the grouping based on what Herring (2007) has called “facets”. Herring’s classification is especially suited to this thesis because it focuses specifically on computer-mediated communication, an area which includes both blogs and email. This faceted grouping aids in collecting the extra-linguistic information associated with each text, a practice
which is common in both corpus linguistics and genre analysis. It strengthens the previous grouping since the texts fall into the same categories based on their facets and by offering another way around the problems associated with using the term *genre*. In discussing computer-mediated discourse, Herring bases her classification of the texts on facets, which she says “cut across the socio-technical modes, and combine to allow for the identification of a more nuanced set of computer-mediated discourse types, while avoiding the imprecision associated with the concept of *genre*” (2007: 4). Herring used a faceted classification scheme for computer-mediated discourse as a whole and while the facets she chose are open-ended and inductive, they nevertheless serve just as well to classify my texts. Using the faceted method is particularly applicable in this thesis because, as Herring (2007: 4) says, “It is intended to complement genre or mode-based analyses, which can provide a convenient shorthand for categorizing [computer-mediated discourse] types, but are less precise and flexible”.

Herring (2007: 10) says that “computer-mediated discourse is subject to two basic types of influence: medium (technological) and situation (social)”. Since these influences are both “unordered” and in a “non-hierarchical relationship”, I will only focus on the medium factor as it is the best suited for my analysis. I will simply adopt some of the facets which Herring used. This will show just how similar my texts are. Table 2 below lists the facets shared by the texts in my sub-corpora.
Table 2. Facets shared by the texts.

The first facet, “synchronicity”, refers to whether users must be logged on to receive messages. In the case of email or blogging, both are asynchronous systems since the sender and addressee(s) do not need to be logged on simultaneously. The second facet, “message transmission”, refers to whether it is possible for “the receiver to interrupt or otherwise engage simultaneously with the sender’s message” (Herring 2007: 14). For the texts in my sub-corpora, the answer is no. Thus they share this facet as well. They also share the facet of “persistence of transcript” since, as Herring (2007: 15) says, “Email is persistent by default, remaining in users’ email queues or files until deleted” and since the blogs are also persistent. The “size of message” facet is not as strictly shared as the previous facets in that it refers to the number of characters the system allows in a single message. This does not refer to “how long most people are willing to type and others are willing to read” (Herring 2007: 15), but to whether the system places a limit on the length of the message. In general, the lengths of emails and blog texts are essentially limitless. It should be noted, however, that there are practical limits on the marketing texts in my corpus which are more restrictive than they are for the blogs. The

<table>
<thead>
<tr>
<th>Facet</th>
<th>Shared by the marketing and blog genres?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchronicity</td>
<td>Yes</td>
</tr>
<tr>
<td>Message transmission</td>
<td>Yes</td>
</tr>
<tr>
<td>Persistence of transcript</td>
<td>Yes</td>
</tr>
<tr>
<td>Size of message buffer</td>
<td>Yes (?)</td>
</tr>
<tr>
<td>Channels of communication</td>
<td>Yes</td>
</tr>
<tr>
<td>Anonymous messaging</td>
<td>Yes (?)</td>
</tr>
<tr>
<td>Message format</td>
<td>Yes</td>
</tr>
</tbody>
</table>

10 All of the posts are still up at the time of this writing and will be available thanks to RSS feeds and Internet archive sites.
marketing texts are also more uniform in their length, which shows that there is
definitely a length restriction being imposed. This restriction will be addressed below
in the corpus linguistic analysis for its effect on the texts’ informational focus. But
going strictly by Herring’s definition of the “size of message buffer”, we can say that
the sub-corpora share this facet.

The “channels of communication” facet refers to whether the system makes other
multimedia available to the addressee(s), such as audio or video, in addition to the
textual channel. While some of the blog posts in my sub-corpus included images, these
were always stock images that were only generally related to the topic of the post. They
could hardly be seen as offering another channel of communication. The channel used
by the texts in my corpus is therefore textual. “Anonymous messaging” refers to
whether users have the option to conduct discourse anonymously or under a
pseudonymous username. It is unclear whether the blogs and marketing texts truly share
this facet. On the one hand, the bloggers’ names are clearly appended to each text, while
the marketers’ names are never included in the marketing texts. On the other hand, the
marketing texts are a collaborative creation and the company’s name is always included
in the marketing texts. So it is uncertain whether the use of the company name, instead
of the marketers’ actual names qualifies as “anonymity”, hence the question mark in the
table. Although Herring (2007: 16) says that “anonymity has been found to have
important effects in online discourse”, I would argue that the marketing texts are not
anonymous in the same way that blogging or communicating online under a pseudonym
is.

11 The headline and bullet points for the NCCC texts, for example, are all under eighty characters, a limit
that is imposed by the marketers so that the subject line will not get cut off when viewed in an email
client.
The final facet, “message format”, simply means “the order in which messages appear, what information is appended automatically to each and how it is visually presented, and what happens when the viewing window becomes filled with messages” (Herring 2007: 17). It is unfortunately impossible to know how the messages appear because of the viewing possibilities offered by email clients and RSS readers. Both of these technologies allow users to sort their messages in almost any order they choose, as well as the ability to automatically filter messages as they come in. The traditional or default format of email clients and RSS readers, however, has no filters and sorts the messages in a reverse chronological format with the most recent on top. Also, a date and author (or company) is automatically appended to each text. They are presented in a similar manner in that the main (and one could argue the only) sections of both are a title and a body. Finally, unlike in an online forum, nothing happens when the viewing window becomes filled with messages, since each text has its own web page or email. I would therefore argue that the texts share the “message format” facet.

It would be tempting to assume that the texts in my corpus would be very similar based on the way that they share many facets. Herring (2007: 9) even says that the facets are used to see which features of computer-mediated discourse “most directly affect users’ linguistic choices”. To this I would argue that since the facets are open-ended and essentially infinite, there will always be facets to add which the texts do not share.

4.2 Tagging and bagging: Steps in the corpus linguistic analysis

The first step was to tag the texts using a part-of-speech (POS) tagger. Archer (2012) says that “Part-of-speech annotation is particularly useful because it increases the specificity of data retrieval from corpora, and also forms an essential foundation for further forms of analysis”. Similarly, Biber (1988: 65) says that “the use of a
computerized corpus also enables automatic identification of linguistic features in a very large collection of texts”. I chose to use the CLAWS (Constituent Likelihood Automatic Word-tagging System) tagger, which was developed at Lancaster University by Paul Rayson. The CLAWS POS tagging software analyzes the text and assigns a part of speech tag to each word. According to its webpage, the software “has consistently achieved 96-97% accuracy” (CLAWS). The tagset I used was CLAWS7, which has 137 tags. After the texts have been tagged with their appropriate part-of-speech, the corpus can then be searched using a program that will retrieve whichever tags (parts of speech) that the researcher wishes to analyze. For example, someone researching verb use in a corpus using CLAWS7 could search for “_V*” and retrieve every word that the CLAWS software tagged as a verb. They could also perform more nuanced searches by retrieving only instances of the verb “has” since this verb has its own tag (_VHZ).

Analyses which are neither as broad nor as narrow as the examples above are also possible through more complex search terms and/or combining the results of multiple searches.

The decision to use the CLAWS software and the CLAWS7 tagset was motivated by several factors. First, the accuracy rate and extensive nature of the tags was obviously desirable to me, with the former offering a measure of insurance that I could trust the tagger and the latter allowing me to perform complex searches in my analyses. Second, CLAWS7 is the tagset now used by the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). Although my specialized corpus does not come anywhere near the scope of these 100 million word behemoths, it does share the same time period, or the time period of English that CLAWS has been proven to correctly tag. Finally, CLAWS is an accepted standard in the field of linguistics and Paul Rayson was kind enough to offer me access to it.
The program I used to analyze my corpus was AntConc, developed by Lawrence Anthony. AntConc is a freeware concordance program which allows for all of the searches I mentioned above as well as many more (AntConc 2011). More information on the specific analyses I performed can be found in the Analysis sections below.

The next step was to perform a quantitative analysis of my corpus by comparing the frequencies of certain linguistic properties of the texts in my sub-corpora. Since the texts are of unequal length, it is necessary to normalize the frequencies. The normalization is done by dividing the raw count by the length of the text and then multiplying by 1,000. This will give the normalized frequency per 1,000 words for the linguistic feature being analyzed.

The motivation of which linguistic properties to analyze comes from previous work by Biber and others. In an exhaustive analysis, Biber (1988) identified linguistic dimensions on which texts differed in comparison to each other. These dimensions were defined by the consistent co-occurrence of linguistic features (Biber 1988: 13). One of the linguistic features that Biber analyzed and that is adopted here is the frequency of nouns and adjectives in the texts compared to the frequencies of verbs and adverbs. The reason this particular comparison is made in this thesis is because it should be indicative of a major feature in which texts differ, which is the relative formality of the texts in each of the sub-corpora. Biber and Conrad (2001: 6) say that genres “differ from one another by being more or less formal”. Heylighen and Dewaele (2002: 320) are more direct by saying “It is hard to avoid the conclusion that a dimension similar to [the relative formality of the text] appears as the most important and universal feature distinguishing styles, registers or genres in different languages” (italics original).¹² The previous studies are in agreement that the frequencies of nouns and adjectives should be

¹² Heylighen and Dewaele use the term “contextuality” to refer to the informality of a text.
higher in more formal texts, while the frequencies for verbs and adverbs should be higher in more informal texts. Moreover, what is important is the co-occurrence of these features, since as Biber (1988: 101) notes, the co-occurrence of linguistic features is indicative of texts which share communicative functions or goals. Since the texts in my corpus make up two different genres, i.e. two separate groups of texts with different sets of communicative functions or goals, we should see differences in the co-occurrence of linguistic features. Specifically, the frequencies of nouns and adjectives compared to the frequencies of verbs and adverbs should indicate which genre is relatively more formal. In performing such an analysis, we should expect to see a difference between the genres because, as Conrad and Biber have noted, “when analyses are based on the co-occurrence and alternation patterns for groups of linguistic features, important differences across registers are revealed”.

4.3 Steps in the genre analysis

Whereas the corpus linguistic analyses will serve to give an overview of the texts being analyzed, the genre analysis will serve to present a more nuanced analysis of the texts. Biber and Conrad (2001: 4) say that “qualitative techniques are required to interpret the functional bases underlying each set of co-occurring linguistic features”. Despite the power offered to corpus linguists by modern computers and advanced tagging software, close readings of important texts are necessary. In fact, corpus studies usually aid researchers in identifying which texts require close readings and prevent them from having to attempt the impossible by reading every text in their corpus. Close readings can help researchers get a better feel of the texts and place their findings from corpus searches in context. In this way, the genre analysis section of this paper will not merely be an added bonus, nor will the corpus linguistics section be simply scratching the
surface. Instead, they will complement each other. I will use the strengths of each to offer a more complete analysis of my corpus.

The fact that genre analysis requires close readings of the texts makes a close reading of every text in my corpus impossible (or at least beyond the scope of this thesis). Even if I could read and analyze all 249 texts, it would not be beneficial unless I grouped them according to the language they contained. Only then would a comparison be permissible. In that case, I decided to group each text according to the topic it covered. This was relatively easy to do because almost every marketing and blog text has a descriptive title.

Thus the genre analysis will focus on a close reading of the texts that present the same topics. The topic I chose to analyze was DISCRIMINATION, a common topic in L&E law. This topic was chosen for two reasons. First, I wanted the topic analysis to draw from as many texts as possible. Second, and more importantly, I wanted the topic analysis to include as many sub-corpora as possible. I decided that at least both of the marketing sub-corpora and three of the four blog sub-corpora should be included in the topic analysis. Furthermore, I wanted to have at least two texts from each sub-corpus. After determining the topic of each of the texts in my corpus, these restrictions left DISCRIMINATION as the only topic to analyze.

Before moving on the specific steps in the genre analysis, a word on the purposes of each genre is in order, since genres are primarily defined by their purposes. The communicative goals of the marketing and blog genres are not the same in my corpus. The goal, or purpose, of the marketing texts is promotional; they are designed to sell a product to their community. The goal of the blog texts in my corpus, on the other hand, is best seen as expository; their purpose is to report events, explain their significance to the L&E law community, and possibly offer commentary or advice based on them. The
different goals of these genres are important because they have an impact on the linguistic structure that the texts will take. According to Swales (1990: 54), a text’s purpose “shapes the schematic structure of the discourse and influences and constrains choice of content and style”. Tardy (2011: 54) is more exhaustive in describing the ways that genres influence texts when she writes that they “embody a social group’s expectations not just for linguistic form, but also for rhetorical strategies, procedural practices and subject-matter or content, among other dimensions, and the unique ways in which these dimensions intersect within a genre”. Therefore, although the differing goals or purposes of the texts in my corpus might seem like a minor element in which they diverge, especially when compared to the fact that they share topics, a community, and many medium facets, the goals have the possibility to profoundly influence the content of the texts. In fact, the different goals or purposes lead to an assumption that the marketing and blog genres not only will but must be different. In a rather deterministic way, the marketers and bloggers do not have a choice in creating lexicogrammatically and discoursally different texts since the purposes of their texts, or what could be seen as the end goals of their texts, call for it.

5. Corpus Analysis

I will start with an analysis of the texts using corpus linguistic methods. This will give us a general overview of the linguistic properties of the texts before moving on to closer readings of certain texts and using genre analysis in chapter 6. As said before, the writers of the texts in my corpus have different goals, so we should expect their language to be different. But we can take the highly regarded nature of the blogs and (at the very least) the continued business of the webinar companies to say that in 2011 this is how the best L&E law blogs and webinar marketing companies discussed their topics
with the same audience. For better clarity, and because they show conflicting results, I separated the noun and adjective analyses from the pronoun analyses.

5.1 Nouns and adjectives

Figure 2 shows the normalized frequencies (per one thousand words, ptw) of the mean, standard deviation (a measure of the spread of the distribution), minimum, maximum, and range of noun use in each sub-corpus. These scores, Biber (1988: 180) says, “enable comparison of the mean scores for different sub-genres as well as consideration of the internal coherence of the sub-genre categories”. As can be easily seen, the two genres differ greatly in their use of nouns. While the mean frequencies for nouns in the blogs are around 300 or below, the means for the marketing sub-corpora are as high as 390 and 439. Biber (1988: 227) says that “a high nominal content in a text indicates a high (abstract) informational focus, as opposed to primarily interpersonal or narrative foci”. So we can say that in relation to each other, the marketing texts are more information oriented than the blogs, which is perhaps not surprising when we remember that the marketing texts have stricter length constraints.

The higher noun frequencies of the marketing texts also show, according to Heylighen and Dewaele (2002), that the marketing texts are more formal than the blogs. They claim (2002: 308) that more formal language features a higher use of nouns, while more informal language has a higher frequency of verbs. The use of verbs in the corpus will be analyzed in the next section (5.2).

13 Tables with the specific scores can be found in Appendix 1, except for where they are explicitly referred to in the text, in which case they will be included in their corresponding figures.
Heylighen and Dewaele (2002: 308) claim that as the frequency of nouns in a text increases, so will the frequency of adjectives, while on the other hand an increase in verbs will lead to an increase in adverbs. Figure 3 shows the normalized frequencies of adjectives in each sub-corpus and the numbers do indeed comply with Heylighen and Dewaele’s assertion. The mean adjective frequencies are nearly identical for the marketing sub-corpora (94.70 for NCCC and 92.47 for Strafford) and they range from 16 to 30 points higher than the blogs. It is also interesting to see that while the standard deviations of adjective frequency for each of the sub-corpora are relatively close, the ranges vary greatly, from FMLAI’s 117.22 and WM’s 112.97 to E&theL’s 60.12 and
Strafford’s 60.49. This shows that the use of adjectives varies internally to a greater extent among the texts in the FMLAI and WM sub-corpora than it does for NCCC, Strafford, and E&theL, possibly because the lengths of the texts also vary to a greater extent.
5.2 Verbs and adverbs

The next step then is to compare the use of verbs and adverbs in the sub-corpora. If we accept Heylighen and Dewaele’s claims, we should expect that the frequencies of verbs and adverbs will be lower in the marketing sub-corpora than the blog sub-corpora, since the use of nouns and adjectives was higher. Figure 4 shows the normalized frequencies (per one thousand words) of the mean, standard deviation, minimum, maximum, and range of verb use in each sub-corpus. It is easy to see that indeed the marketing texts use fewer verbs relative to the blog texts. Even beyond that, Strafford, the marketing sub-corpus which used more nouns compared to NCCC, uses fewer verbs, while the blog

<table>
<thead>
<tr>
<th>Adjectives</th>
<th>NCCC</th>
<th>Strafford</th>
<th>E&amp;theL</th>
<th>FMLAI</th>
<th>SYGIGH</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>94.70</td>
<td>92.47</td>
<td>76.94</td>
<td>64.64</td>
<td>62.44</td>
<td>60.11</td>
</tr>
<tr>
<td>SD</td>
<td>18.79</td>
<td>15.77</td>
<td>15.02</td>
<td>21.02</td>
<td>18.93</td>
<td>22.68</td>
</tr>
<tr>
<td>Min</td>
<td>66.01</td>
<td>70.56</td>
<td>44.48</td>
<td>25.64</td>
<td>23.81</td>
<td>18.18</td>
</tr>
<tr>
<td>Max</td>
<td>128.15</td>
<td>131.05</td>
<td>104.60</td>
<td>142.86</td>
<td>105.84</td>
<td>131.15</td>
</tr>
<tr>
<td>Range</td>
<td>62.14</td>
<td>60.49</td>
<td>60.12</td>
<td>117.22</td>
<td>82.03</td>
<td>112.97</td>
</tr>
</tbody>
</table>

Figure 3. Normalized adjective frequencies (ptw) in the corpus.
sub-corpora which had higher mean noun frequencies (E&theL and FMLAI) actually have lower mean verb frequencies, just as Heylighen and Dewaele led us to expect.

Figure 4. Normalized verb frequencies (ptw) in the corpus.

The adverb frequencies mimic this divide. Figure 5 shows the normalized frequencies for adverbs in each sub-corporus. Biber (1988: 224) writes that place and time adverbials “mark direct reference to the physical and temporal context of the text […] their distribution as marking situated, as opposed to abstract, textual content”. So the frequencies here bring us to the same conclusion that the analysis of nouns and adjectives did, i.e. that the marketing texts are more abstract and informational in relation to the blog texts, which are more informal and situated or contextual.
Figure 5. Normalized adverb frequencies (ptw) in the corpus.

5.3 Pronouns

Referring to a study by Biber, Hudson (1994: 332) says that “The balance of common nouns to pronouns is already well-established as one of the most important variables on which genres vary”. Figure 6 below presents the normalized frequencies for all pronouns in each sub-corpus. The frequencies mirror the use of nouns, where a higher frequency of nouns in a sub-corpus correlates with a lower frequency of pronouns in that sub-corpus (in relation to the other sub-corpora).
A discussion of which exact pronouns are used is beyond the scope of this thesis, but Biber (1988: 225) has noted that studies which have grouped pronouns together as one class, such as this one, have found that a higher frequency in their use is indicative of “a less formal style” and a more “interpersonal focus”. So the pronoun frequencies show precisely what the other frequencies have, i.e. that the blogs are more informal in relation to the marketing texts. The only discrepancy, however, is very interesting in that the pronoun frequencies for E&theL show that it is much more similar to the marketing texts than the other blogs. The noun frequencies did not show this.

6. Topic analysis
As noted above, the topic analysis will focus on DISCRIMINATION in L&E law. This topic was chosen because it represented the widest range of texts from the corpus. It includes both of the marketing sub-corpora, three of the four blog sub-corpora (FMLAI is excluded), and at least two texts on DISCRIMINATION from each of the sub-corpora included in this analysis. From the marketing sub-corpora, there are two texts from NCCC and three from Strafford. From the blog sub-corpora, there are six texts from E&theL, thirteen from SYGIGH, and five from WM.

6.1 DISCRIMINATION at work – nouns and adjectives

Figure 7 shows the normalized frequencies of nouns in the texts about DISCRIMINATION from each sub-corpus. The frequencies and their relations to each other are similar to those that were found in the analysis of the entire corpus (section 5.1). Again we see that the blogs feature much fewer nouns than the marketing texts, where the maximum frequency of nouns in the blogs (WM: 355.56) is still 19 points lower than the minimum frequency of nouns in the marketing sub-corpora (NCCC: 374.41). In fact, with the exception of WM, the mean frequencies of nouns in the DISCRIMINATION texts from each sub-corpus are all within eight points of their respective total mean noun frequencies. So when discussing the topic of DISCRIMINATION in the workplace (or trying to sell a webinar that will discuss it), the marketing texts are far more abstract and information oriented than the blogs, which could be said to be more informal in comparison.

The other interesting aspect to note about these figures is the differences in ranges for the marketing texts and blogs. Whereas the marketing texts have much smaller ranges in noun frequencies, the blogs vary quite a bit, from 69.32 in WM to 133.61 in SYGIGH.
This shows that the bloggers write with a freer style, another indication of their relative informality.

![Box plot showing noun frequencies for different sub-corpora](image)

**Figure 7. Normalized noun frequencies (ptw) of the texts that discuss DISCRIMINATION.**

<table>
<thead>
<tr>
<th>Nouns</th>
<th>NCCC</th>
<th>Strafford</th>
<th>E&amp;theL</th>
<th>SYGIGH</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Mean</td>
<td>385.22</td>
<td>431.51</td>
<td>298.30</td>
<td>250.38</td>
<td>308.51</td>
</tr>
<tr>
<td>SD</td>
<td>10.82</td>
<td>13.95</td>
<td>24.31</td>
<td>34.87</td>
<td>24.64</td>
</tr>
<tr>
<td>Min</td>
<td>374.41</td>
<td>417.27</td>
<td>256.10</td>
<td>203.47</td>
<td>286.23</td>
</tr>
<tr>
<td>Max</td>
<td>396.04</td>
<td>450.45</td>
<td>331.95</td>
<td>337.08</td>
<td>355.56</td>
</tr>
<tr>
<td>Range</td>
<td>21.63</td>
<td>33.18</td>
<td>75.85</td>
<td>133.61</td>
<td>69.32</td>
</tr>
</tbody>
</table>

The frequencies of adjective use in the DISCRIMINATION texts in each sub-corpus also follow the total frequencies found in the analysis of the entire corpus. Figure 8 shows that the blogs, which featured less nouns, also feature less adjectives, a result that we would expect based on the claims by Heylighen and Dewaele (2002) and the results
from the analysis of the entire corpus. It is again interesting to see the differences in the ranges. While the marketing sub-corpus NCCC shows considerable range in adjective frequency, even greater than one of the blogs (E&theL), its range is much smaller than the two blogs with the highest ranges, SYGIGH and WM. The ranges of adjective frequencies of these two blogs also dwarf the range of Strafford. A comparison of the adjective frequencies again leads to the conclusion that the blogs are more informal and less information oriented than the marketing texts.

Figure 8. Normalized adjective frequencies (ptw) of the texts that discuss DISCRIMINATION.

6.2 DISCRIMINATION at work – verbs and adverbs
While the analysis of nouns and adjectives in the texts on DISCRIMINATION in the workplace showed that the marketing sub-corpora were more formal and informational than the blog sub-corpora, the frequencies for the verbs and adverbs are not as clear cut. Figure 9 shows the normalized verb frequencies for the DISCRIMINATION texts in each sub-corpus. Here the differences in usage between the marketing and blog sub-corpora are not nearly as large as they were for the nouns. The difference in means between the blog sub-corpus WM and the marketing sub-corpus NCCC, for example, is less than three points, while the difference between the blog sub-corpus E&theL and NCCC is only ten points. These frequencies and the differences between them do not easily fit with the idea that the blogs are more informal than the marketing texts, at least not in terms of their use of verbs. It would be safer to say instead that based on these frequencies, the marketing sub-corpus NCCC and the blog sub-corpora E&theL and WM are relatively similar to each other and that they are less formal than the marketing sub-corpus Strafford and more formal than the blog sub-corpus SYGIGH.
Figure 9. Normalized verb frequencies (ptw) of the texts that discuss DISCRIMINATION.

The use of adverbs backs up this statement. Figure 10 shows the adverb frequencies for the DISCRIMINATION texts for each sub-corpus. From the mean frequencies, we can see that NCCC, E&theL, and WM are relatively similar to each other, but different than both Strafford and SYGIGH. This is what we would expect from the analysis of verbs in the DISCRIMINATION topic, but like that analysis, it is also in direct contrast to what we would expect from the analysis of nouns and adjectives in the DISCRIMINATION topic. It is not so easy to divide the two genres based on formality. Rather, when we try to conclude the relative formality of these sub-corpora based on their use of nouns/adjectives and verbs/adverbs, we get conflicting results. Thus, Heylighen and Dewaele’s (2002: 335) claim that formality “is the most fundamental and most universal
dimension of stylistic variation” may be correct, but it is not apparent here, at least not when the other analyses have shown a clear distinction between the genres.

6.3 DISCRIMINATION at work – pronouns

The frequencies of pronoun usage in the DISCRIMINATION texts are similar to the verb and adverb frequencies. Figure 11 shows the normalized pronoun frequencies and we can see that again the mean frequencies for NCCC, E&theL, and WM are similar. If a higher frequency of pronouns indicates “a less formal style” and a more “interpersonal focus” (Biber 1988: 225), then these three sub-corpora are somewhat less formal than the marketing sub-corpus Strafford and very much more formal than the blog sub-corpus SYGIGH. This is the same conclusion that was drawn from the verb and adverb

Figure 10. Normalized adverb frequencies (ptw) of the texts that discuss DISCRIMINATION.
frequencies and it is different than what the noun and adjective frequencies told us. The frequencies do, however, fit with Heylighen and Dewaele’s (2002: 308) assumption that in less formal texts, an increase in verb frequencies “will be reinforced by the fact that the more formal noun phrases […] will tend to be left out completely or replaced by pronouns”.

Figure 11. Normalized pronoun frequencies (ptw) of the texts that discuss DISCRIMINATION.

In terms of the relative formality of the sub-corpora then, we are left with, on the one hand, noun and adjective frequencies that make a clear divide between the genres and, on the other hand, verb and adverb frequencies, backed up by the pronoun frequencies, that group two of the blog sub-corpora (E&theL and WM) with one of the marketing sub-corpora (NCCC). The verb/adverb and pronoun frequencies further set this group
off as being more formal than the other blog sub-corpus (SYGIGH) and less formal than the other marketing sub-corpus (Strafford).

6.4 Qualifying DISCRIMINATION at the workplace

I will now turn to the genre analysis of the texts on the topic of DISCRIMINATION in the workplace. This analysis will help to contextualize the findings from the corpus linguistic analyses. These close readings will show us what types of rhetorical variations are at work in the texts from different sub-corpora. This will help us compare them on another level. Instead of looking at the lexico-grammatical variations as we did in the corpus linguistic analyses, we will be looking at the discourse strategies used in the texts to discuss (and sell) the topic of DISCRIMINATION to the same audience, L&E lawyers in the US. Examples of each of the webinar landing pages and blog posts, complete with their sidebars and banners which were not included in the corpus, can be seen in Appendix 2.

We have already seen that the blog texts are not similar in comparison to the marketing texts in terms of lexico-grammatical variations. The genre analyses here will therefore compare their rhetorical variations. It is important to remember, however, that they do not come from different domains so it will be interesting to see what variations pop up when comparing the sub-corpora.

In analyzing the marketing texts, I will be looking at the rhetorical move structure outlined in Bhatia (2004: 65), which is:
1. Headlines (for reader attention)
2. Target the market
3. Justifying the product or service
   • by indicating the importance or need of the product or service and/or
   • by establishing a niche
4. Detailing the product or service
   • by identifying the product or service
   • by describing the product or service
   • by indicating the value of the product or service
5. Establishing credentials
6. Celebrity or typical user endorsement
7. Offering incentives
8. Using pressure tactics
9. Soliciting response

The way in which these moves are applied to the marketing texts is best shown through an example. Figure 12 shows one of the marketing texts from the NCCC sub-corpus, with the rhetorical move labels attached to relevant sections. The reasoning for attaching the rhetorical moves will be discussed in further detail in the Analysis section below.
Figure 12. Example of a marketing text (NCCC7). Rhetorical moves are outlined on the right.

6.4.1 Register Now: Genre analysis of NCCC

Preventing Retaliation Claims: Cat’s Paw, Title VII & Third-Party Protection
Webinar Recording Price: 199.00 (U.S. Dollars)

SUMMARY
The recent SCOTUS decisions of Thompson v. N.A.S. and Staub v. Proctor Hospital have opened the door to increased retaliation claims. With decisions affecting Cat’s Paw liability, Title VII and third-party retaliation, how can you be sure your clients are truly protected against retaliation claims? Join us for this 60-minute webinar, where you and your colleagues will discover:
- Keys to preventing retaliation claims: Addressing Title VII and Cat’s Paw liability
- Strategies to protect against third-party retaliation in the wake of Thompson v. N.A.S.
- Strategies for drafting effective policies that mitigate retaliation claims
- How to protect against claims: Protected conduct, adverse action, and causation

PROGRAM BENEFITS
This practical, 60-minute webinar will provide you with the necessary information on how to effectively protect your clients from retaliation claims.

PROGRAM HIGHLIGHTS
- Keys to Navigating the Changing Landscape of Retaliation Claims
  - How the new Title VII definitions will affect your current policy
  - How to prepare for the expansion of “protected activity”
  - Recent SCOTUS rulings and how they impact your clients
- Keys to Preventing Third-Party Retaliation and Cat’s Paw Liability
  - Where and when employers are now liable for Cat’s Paw claims
  - Third-party retaliation claims: Protective steps your clients need to take now
  - Critical elements every “independent” investigation needs to have
  - How to properly train mid-management to prevent Cat’s Paw claims
- Drafting Disciplinary Policies that Protect and Prevent
  - Policies for properly documenting investigations and resolutions
  - How to prepare when taking disciplinary action against an employee
  - Effective procedures for internal complaint resolution
- Live question and answer session: Have your retaliation claims questions answered.

SPEAKER
Ralph R. Smith, III, Partner at Duane Morris LLP, practices in the area of employment litigation and prevention, employment practices, which includes counseling employers on the creation of employment policies and employer training to avoid employment-related litigation. Along with his labor and employment law practice, he also represents both companies and individuals in other kinds of related complex commercial litigation.

- Mr. Smith has represented clients before federal and state courts and administrative agencies in a broad array of labor and employment cases including, inter alia, race, gender, age, national origin, disability and workplace harassment and discrimination matters, wage-and-hour disputes, restrictive covenants, grievances, arbitrations, drug testing and other employment-related contract issues.
- Ralph is a frequent lecturer on labor and employment-related topics and formerly served as an adjunct professor at Widener University Graduate School of Business.

- Pressure tactics
- Target the market
- Justify the product
- Detailing the product
- Offering incentives
- Establishing credentials
- Celebrity endorsement (?)
As can be seen from figure 12 above, not all of Bhatia’s (2004) rhetorical moves in advertisements can be found in the NCCC texts, and the moves that are in the texts are not in the same order that Bhatia lists them. They each, however, start with headlines that are meant to attract reader attention. These headlines are also the subject lines of the emails, so they are specifically designed to engage readers, with the goal of getting them to open the email. As mentioned above, webinar producers at NCCC are required to write headlines less than eighty characters long so that they will not get cut off in the potential customers’ email client.

The next sections of the NCCC marketing texts are titled “Summary” and they include pressure tactics in the form of clauses such as “How can you ensure your clients are properly represented…” (NCCC11) and “How can you be sure your clients are truly protected…” (NCCC2). These summaries of the webinar target the market as L&E lawyers who represent businesses. They also justify the webinars by establishing a niche and indicating their importance through a clause that leads into the next section, which is a set of bullet points. The clause is “Join us for this 60-minute webinar, where you and your colleagues will discover”. The set of bullet points that follow are four continuations of this lead-in clause and each of them are moves that act to simultaneously justify the webinar and detail its contents through phrases such as “How to best avoid age discrimination claims” (NCCC11) and “Keys to preventing retaliation claims: Addressing Title VII and Cat’s Paw Liability” (NCCC7).

The marketing then moves into a section called “Program Benefits”, a title which itself is an offer of incentives. The clauses in the sections are direct justifications of the webinar. Both of the texts have only one sentence in the section and they are very similar. They are

14 Where NCCC7 uses “colleagues” and “discover”, NCCC11 uses “clients” and “learn”.

---

52
• This practical, 60-minute webinar will provide you with the necessary information on how to effectively protect your clients from retaliation claims. (NCCC7)
• This 60-minute webinar will provide you with the knowledge and skills to effectively prepare, litigate, and defend any age discrimination cases you may encounter. (NCCC11)

These clauses indicate the importance of attending the webinar for lawyers who represent businesses, but it is obvious that they are both templates, with the relevant subject of the webinar filled in at the end of the sentence.

The next sections of the marketing texts are called “Program Highlights” and they feature three major bullet points, each with three related minor bullet points placed underneath them. As with the bullet points in the first section of the texts, these clauses serve to both describe and justify the webinar. They are not full grammatical sentences, but merely simple clauses of what the webinar will include, such as “Winning themes for judges and juries” (NCCC11) and “Drafting Disciplinary Policies that Protect and Prevent” (NCCC7). In each of the texts, the Program Highlights section ends with a clause that offers incentives to customers who attend the webinar, namely that there will be a question and answer session at the end.

The final sections of the marketing texts are called simply “Speaker” and they include the speakers’ names, job titles, the firms they practice at, and their speaker bios. This section serves the rhetorical move of establishing credentials by explaining the speaker’s qualifications for presenting the topic, sometimes by outright claiming that they are “a highly sought after speaker” (NCCC11). While it could perhaps be argued that this section also serves as a celebrity endorsement move, I prefer to take a more cautious approach since it is unclear whether the speaker or their law firm are really known by the L&E lawyers in the same way that, for example, a movie star is known to the general public.
There are two additional moves in the marketing texts. The first is that of soliciting response, which appears in the form of a link that takes potential customers to the page where they can register (and pay) for the webinar. On the landing pages, which are also the places that I drew the marketing texts from, the response solicitation appears in the form of a button which says “Register Now”. Clicking on the button takes potential customers to a form where they can fill in their personal information and pay for the webinar.

The other move in the marketing texts is that of offering incentives and it also appears on the landing pages. The incentives are that L&E lawyers can receive CLE credits for attending the webinar (if they pay an extra fee) and that the speaker’s PowerPoint presentation will be available for download to registered customers. These incentives are given their own section after the marketing text and before the “Register Now” button. I did not include them in the corpus because they are not in the emails. They are, however, on every single webinar’s landing page and the language on each is exactly the same.
6.4.2 Listen: Genre analysis of Strafford

For greater clarity, an overview of the rhetorical move structure in the Strafford texts is provided as a list below:

Strafford move structure
1. Headlines (for reader attention)
2. Target the market
3. Detailing the product or service
4. Offering incentives
5. Target the market
6. Justifying the product or service
7. Detailing the webinar
8. Using pressure tactics
9. Detailing the product or service
10. Offering incentives
11. Detailing the product or service
12. Establishing credentials

The list clearly shows that the move structure of the Strafford marketing texts is similar to the NCCC texts in that not all of Bhatia’s (2004) moves are present, nor are they in the same order. The Strafford texts start with a title and subtitle that serve as headlines to draw potential customers’ attention. These headlines sometimes include phrases which target the market, such as “Employer strategies” (Strafford2), but they essentially just explain the subject of the webinar.

The main bodies of the marketing texts start with a section that details the webinar, but unlike NCCC this section also includes language that offers incentives (“This CLE webinar”) and phrases that target the market (“for employment counsel”). On a whole, however, the sections are short one- or two-sentence paragraphs that act to quickly detail the contents of the webinar and lead into the next sections, which are more complex in terms of their move structure.

The next sections of the marketing texts, called “Description”, consist of four short paragraphs. These paragraphs are made up of either one or two sentences, but they are
set off from each other. This section is interesting in that it mixes the move of justifying the webinar with the moves of detailing the webinar and using pressure tactics. The first two paragraphs of this section justify the webinar by establishing the niche (like in the NCCC texts), but they also include language which apply pressure tactics, sometimes with worrying notions such as “Employees are filing workplace discrimination lawsuits in record numbers” (Strafford17) and sometimes by stating what companies lose each year in monetary values such as “Litigation arising from EEOC claims costs employers more than $44.3 million annually” (Strafford8). The third paragraph of the Description section is decidedly a pressure tactic move and it features clauses which explicitly state what the businesses and the L&E lawyers representing them should or must do. The use of modals of obligation is common in the pressure tactics in the Strafford texts. Interestingly, in Strafford17, this pressure tactic move actually comes before the two justification paragraphs instead of after them. This means that Strafford17 has the move structure flipped compared to Strafford2 and Strafford8, which shows that while there is definitely a move structure template at work in the Strafford texts, its order is not as rigid as what we saw in the NCCC texts, where the moves in the texts were in the exact same order.

The final paragraph of the Description section is both a continuation of the justification of the webinar and a lead-in to the description of the webinar. The interesting part of this paragraph, in comparison to NCCC, is how each of the Strafford texts invites potential customers to “Listen” to the webinar. This word, which starts the paragraphs in each of the texts, makes it seem that attending the webinar is as simple as just paying attention, never mind the couple of hundred dollars it costs to “listen”.

The next sections of the Strafford texts are called “Outline” and they are exactly that, bullet point lists which are very similar to the Program Highlights section of the NCCC
texts in that they feature three to four major bullet points with several minor bullet points nested beneath them. The bullet points in this section serve the move of detailing the webinar. The structure of the clauses in these bullet points seems to not be as rigid as the Program Highlights section of the NCCC texts because the phrasing of the bullet points is much simpler in comparison to NCCC. For example, some of the bullet points are simply two word noun phrases, such as “Severance agreements” (Strafford 17) or the names of specific court cases that will presumably be discussed in the webinar.

The following sections are called “Benefits”, which is an obvious offering of incentives, just as the title “Program Benefits” was in NCCC. Unlike the NCCC Program Benefits sections, however, these sections consist of four questions which serve to detail the contents of the webinar, such as “What should companies consider in evaluating settlements of EEOC claims?” (Strafford 8). These questions, which are also descriptive of the webinar’s contents, are prefaced in each of the Strafford texts with the clause “The panel will review these and other key questions”.

The final sections of the Strafford marketing texts are the speaker bios. Interestingly, these sections are titled “Faculty”, which lends and air of scholarly qualification to the speakers and the webinars. In comparison, NCCC simply titles their speaker bio sections “Speaker”. In every other way, however, the Strafford speaker bio sections work just like the NCCC section. They establish credentials by explaining how the speakers have “extensive experience” (Strafford 8) or how they have “successfully resolved thousands of workplace disputes” (Strafford 2). Each speaker bio is prefaced with the speaker’s name and the firm in which they practice, just as they were in NCCC. Again, although these speakers are not “celebrities” in the common definition of the

---

15 With sub-titles like “Program Highlights” and “Benefits”, some of the moves in the marketing texts are very easy to identify.
term, the law firms that they work at could serve as a proxy celebrity endorsement, but arguing this claim is too problematic without knowing more about the L&E lawyer community’s knowledge of the speaker or their firm. The other interesting fact about the speaker section in the Strafford texts is that the city in which each speaker works is also listed, even though it is hard to see how this serves to help market the webinar better. Perhaps lawyers working in larger cities are held in higher regard than others, which could be why the cities listed are Chicago, Los Angeles, New York, etc., but this is just conjecture.

6.4.3 Cat’s paws and going fishing: Genre analysis of E&theL

As will be shown in the analyses below, the blog texts feature rhetorical moves and tactics mostly associated with reporting genres and they will analyzed as such. Bhatia (2004: 82) lists different types of reports, such as news reports, business reports, and police reports, and says that they are likely to lexico-grammatical and rhetorical variation even though they share the communicative purpose of reporting on. The rhetorical moves that I have formulated to analyze the blog texts are:
Move structure in blogs

1. Headlines (for reader attention)
2. Target the market
3. Description (reporting) of events
4. Explanation
5. Commentary
6. Advice
7. Soliciting response

None of these moves are necessary and not all of them are present in every blog text. Instead, they are used and combined when appropriate for the bloggers’ communication goals. I believe that the moves are self-explanatory based on their names. Their usage will be discussed in further detail below. Figure 13 shows a blog text and an example of how these moves are applied to it.
Figure 13. Example of a blog text (E&theL5). Rhetorical moves are outlined on the right.

As can be seen from figure 13, the structure of the E&theL blog texts is decidedly different than that of the marketing texts. Nevertheless, there is a distinct formula or move structure shared across each post that can be defined. The first part is obviously the title of each text. In the case of the E&theL texts, the titles are mostly generic descriptions that help to set up the subject of the post or blog text, such as “Failure to Exhaust Remedies” (E&theL23) and “Employment Law – Class Action Update”
There are two titles, however, which use colorful language, the first being the idiom used in E&theL5’s title, “In the Zone: Retaliation under Title VII”. This term, which is derived from sports and describes a peak of consciousness or performance, is cleverly contrasted in the text with the legal term “zone of interests”. The second creative title, “Gone Fishing – Pennsylvania District Court Reels in the EEOC” (E&theL19) similarly contrasts a legal term, “fishing expedition”, with a common phrase, “gone fishing”. The meaning of the legal term is actually defined in the post, which adds context to the subject of the text.

The second part of the structure of the E&theL blog texts that can be defined is description or reporting of events move. E&theL5 is the simplest example of this. The first three paragraphs of E&theL5 report the events of a particular case, contextualizing the information with quotes from the specific case in the first paragraph. The use of quotes, I should mention, is common in my blog texts (as it is in journalism). The next paragraph then adds quotes from the legal act examined by the case and includes more quotes from the case. The third paragraph is simply a long block quote statement from the court that heard the case. The fourth and final paragraph of E&theL5 contains the moves of explanation and commentary on events. This commentary is formal and technical in style.

The move structure is followed to varying degrees by the other E&theL blog texts which talk about DISCRIMINATION. Sometimes the description of events is more detailed and placed in context with the history and definition of a legal term, as in E&theL8, which discusses a Cat’s Paw case. In other texts, however, the reporting of events is followed by an explanation and/or commentary and hyperlinks to places where readers can learn more about the subject of the post, such as in E&theL10, which actually links
to another blog in my corpus, FMLAI.\textsuperscript{16} E&theL21, which offers a quick description of events and commentary or explanation of them, is more like the marketing texts in that it promotes two podcast interviews that the author participated in and links to them. This text also links readers to another blog for further information on the legal case which is the subject of the text and the podcasts. Despite looking similar to promotions, I would hesitate to label them as such and instead classify them as examples of the hyperlinking aspect that is so common to the blog genre.

Another aspect that is consistent across the E&theL texts is the inclusion of a specific legal case in each text. Each text therefore discusses a legal case and names it up front, usually in the first or second sentence. This is indicative of the restrained structure of E&theL. There is a very definite focus on specific events in the legal world, such as the outcomes and meanings of specific court decisions. The texts do not stray from this focus. They introduce a court case, describe the events that occurred in it, and offer an explanation and/or commentary on the court’s decision.

As shown in the corpus linguistics analysis, the language in the E&theL texts is less formal than the marketing texts. This can be easily explained by their different rhetorical move structures and purposes. Whereas the promotional goal of marketing texts calls for noun heavy and verb free phrases, such as bullet points, the expository goal of E&theL (and the blogs in general) would naturally allow for a relatively more informal style.

\textsuperscript{16} The post linked to was not included in the DISCRIMINATION topic because it was one of those written by the author whose posts were removed.
6.4.4 Get a grip: Genre analysis of SYGIGH

Since the rhetorical moves in SYGIGH do not exactly follow the move structure in E&theL, they are listed below:

1. Headline (for reader attention)
2. Target the market
3. Explanation
4. Commentary
5. Advice
6. Soliciting response

Although the SYGIGH blog is more varied in its move structure than E&theL, each text still has a title or headline. These titles are not as formal or informational as they were in E&theL. Instead, there are a few different styles of titles in the discrimination topic texts in SYGIGH. The first is what I would call an informal and attention-grabbing style. It is evident in SYGIGH2 and SYGIGH12, where the titles are respectively “The Genetic Information Nondiscrimination Act and The Mentalist” and “Your Employer Can’t Discriminate Because of Your Bankruptcy (But Your Potential Employer Can)”. The first of these titles is informal in that it relates a legal act with a popular American television show, thus shifting the reader’s attention from the formal name of the act that starts the title to the catchy name of the show (especially if readers happen to be fans of the show). The second title captures the reader’s attention by directly addressing them as “you” and warning that they can legally be discriminated against in the future.

The other types of titles in the SYGIGH texts are interesting when related to the contents of their respective texts. The first type is what I will call the numbered list article. This style of titling articles is common across internet sites and an example of it is “Six Ways You Might Be Discriminated Against If You Put Family First” (SYGIGH16). Interestingly, each of the three texts with this style of title ends up linking out to a post by the author on AOL Jobs, a division of the America Online news website. This is not the type of linking common to blogs that was discussed above. Instead, these
texts are enticingly truncated, forcing readers to click through to AOL Jobs if they want to get the point of the text. This indicates that this style is probably not the choice of the author, who is directly in charge of what appears in the blog, but of AOL Jobs, where the author presumably works with an editor.

The third type of title in SYGIGH takes the form of a simple question which sets the subject matter of the text. An example is SYGIGH60’s “Can My Employer Make Me Speak English, Even On Breaks?”. This question style of titles appears in six of the texts which discuss DISCRIMINATION and in all but one of them they also serve to target the audience as employees instead of employers or L&E lawyers representing employers, who were the main target audience for the marketing texts and the E&theL blog texts. The texts take the form of problem-solution, which is interesting because, according to Kim (2007: 97), this is a discourse feature “commonly found in journalism texts and expository writing”. This reinforces my claim that the goal of the blogs is expository. The one question title that does not directly target the audience is SYGIGH43. This title, “Does Discrimination Still Exist? Of Course It Does”, also features an answer, while the posts links out to an article by the author of SYGIGH on the Huffington Post, an online newspaper.

The structures of the texts are, like their titles, varied in comparison to the structure of E&theL. Not every post in the collection of SYGIGH texts on DISCRIMINATION mentions a specific court case or legal act, although many of them do. However, each post has the explanation rhetorical move in the form of explaining the law, even if a specific case or act is not mentioned by name. In the cases where the explanations are not of court case decisions or legal acts, they may be of legal terminology, such as in SYGIGH24 and SYGIGH53, or of the general state of affairs in discrimination law, such as in SYGIGH18.
Another rhetorical move shared by all of the texts is a commentary. This move functions in the same way as it did in the E&theL texts, but unlike the commentary in E&theL, the commentary on the law, legal acts, or courts cases in SYGIGH is much more informal, direct, and often personal. In fact, the commentary move in SYGIGH is identifiable by the use of first-person pronouns. Examples of this include statements like “In my opinion, employers who refuse to hire people just because of bankruptcies are idiots” (SYGIGH12), “I find the issue of whether any kind of discrimination still exists to be a continuing uphill battle when I represent employees in discrimination cases” (SYGIGH43) and perhaps most interestingly

Some people tell me they were so traumatized by one remark that they can never go back to work there. Okay. I mean, really? I guess if you have severe depression going into the situation I might buy that. But c’mon. Get a grip. People say awful things sometimes. You can’t let one remark ruin your life. If you turn tail and run, you’re giving the bigot exactly what they want. (SYGIGH24)

This last quote of informal commentary is also an example of the third move common among the SYGIGH texts, that of the way that each text offers simple and practical advice to employees. The most obvious examples of advice come under the heading “Donna’s tips”, which appears in six of the texts and presents the advice in a list format. But other examples of advice come as simply explaining the responsibilities under the law of employees wishing to file a claim (and be successful in court). The use of advice, similar to Bhatia’s (2004) offering incentives move that was found in both the marketing sub-corpora, also targets the audience of the texts as employees.

There is however another interesting aspect to the way that the SYGIGH texts offer advice. Although I already stressed in the corpus linguistic analyses and in the discussion of how titles and commentaries are written that the SYGIGH texts are more informal, this does not mean that the language of the texts and the ideas behind them are
somehow less complex than those in either the marketing or E&theL texts. On the contrary, by presenting the advice in an if...then conditional conjunctive adjunct, the author of the SYGIGH texts is using an analytical and logical tone. So SYGIGH also represents a complex and analytical approach to discussing the law (not exactly surprising since the texts are written by a lawyer), but it presents its discussion in a more informal way.

One final interesting aspect of the SYGIGH texts comes at the end of SYGIGH60, where the author makes a direct appeal to readers for discussion on the topic. The final paragraph of this text starts with “I’d love to hear from you on this” and then lists five questions related to the topic that was discussed in the text. This is a direct soliciting response rhetorical move. The ability to comment on the texts is one of the defining aspects of a blog (Myers 2010, Blood 2003), and here it shows an overlap of the L&E law blog sub-genre with the larger blog genre. Commenting also shows the interesting difference between the soliciting response moves in the blogs and the marketing texts. While customers of the webinars are permitted to ask questions of the speakers, they must first be present during the live webinar to do so. And that requires that they purchase the webinar. Commenting on the blogs, on the other hand, is free. Although this move is evident here, it is the only example of it in the texts in the DISCRIMINATION topic, and it is unclear whether it appears in any of the other blog texts.

6.4.5 Short and sweet: Genre analysis of WM

In the WM blog, the rhetorical move structure is again different from the other two blogs, so a list of the specific moves is in order. They are:

1. Headlines (for reader attention)
2. Description
3. Explanation
4. Commentary
5. Advice
Although it cannot easily be seen from the list of moves, the structure of the WM blogs is actually very different than either of the other blogs or the marketing texts. Unlike the other blogs and marketing texts, there is no real commonality among the titles in the five WM texts about DISCRIMINATION. One title, “Back to basic on the defense” (WM18), is very generic and does not indicate the subject matter well, although the last two sentences in the body text tie back to the title nicely (“It talks about the basics […]. Never forget the basics.”). The title for WM23 is simply “Say ‘cheese’”, an idiomatic title that does not indicate the subject matter of the text at all. The final three titles can possibly be grouped together in that they are all descriptive of the subject matter, but it is a stretch since the level of description varies.

The next structural aspect to notice is that unlike the other blog texts or marketing texts, the WM texts are each only one paragraph long. The paragraphs still feature descriptions of events, explanations of the law, and commentaries, but these are not set off from one another as they were in E&theL and SYGIGH. In fact, instead of going from one to the next (but without any line breaks), these moves are mixed in WM38 and WM39. This does not mean that the language of the WM texts is confusing, just that they take a much more “short and sweet” approach to describing recent events in L&E law. In comparison to especially SYGIGH, which may in one text have offered commentary on many issues related to a specific court case, the texts in the WM blog discuss one particular aspect of a case and offer commentary or advice on that case alone, such as “I like this case. It talks about the basics – the blocking and the tackling that it takes to mount a successful defense. Never forget the basics” (WM18). Despite their short nature, each of the texts on DISCRIMINATION in the WM blog does reference a case or hearing by name, just like the E&theL texts did.
The variation within the five DISCRIMINATION texts in WM is also not entirely unexpected. Bhatia (2004: 82) says that texts which share a genre, such as news reports, often show variation both in their lexico-grammatical properties and in the rhetorical move structuring. The variations in text length and titles in WM, however, are indicative of the subtle variations in rhetorical structuring. The moves of reporting on events, commenting on them, and advising the audience are still in the blog texts, but their ordering is rather different in WM compared to either E&theL or SYGIGH. The analysis of WM shows that the E&theL and SYGIGH sub-corpora are more similar to each other in terms of rhetorical move structuring.

Finally, there is one very interesting text to look at. WM26, while being similar in structure to the other texts, sets up a hypothetical situation for the reader to consider. This is similar to what was done in SYGIGH46 except that here the hypothetical situation is prefaced with a subtitle, “Let’s play claim or no claim”. The author does this to introduce an aspect of the law that he wants to offer advice on. The title of the text alludes to this advice although it does not explain why employers should act a certain way (“Don’t use unemployment as a proxy when it’s a poor predictor”). After the hypothetical situation is set up, the author shifts into describing a real world event that bears on it. Then comes the advice, “My take: Employers must always be wary of absolute rules” and the reasoning, “Why? Civil rights laws are designed for individual assessments, not barroom generalities. Employers should tread carefully in this area. The EEOC is watching”. I find it interesting that both the advice and the reasoning for it are introduced with explicit statements (“My take” and “Why?”) because doing so adds an informal nature to the writing, when according to the corpus linguistic analyses, WM is the most formal of the blogs. It shows that the means in the corpus linguistic analyses
do not necessarily predict the linguistic structure of each of the texts in the sub-corpora, hence the importance of the range and the whiskers in the figures.

7. Conclusion and Discussion

The first important result of this thesis is that the corpus linguistic analysis showed that dividing the texts on formality is not as straightforward or clear cut as it would seem to be. We saw that the overall comparison of the genres showed that the marketing texts displayed lexico-grammatical features that pointed to a more formal style than the blogs. The analysis of the DISCRIMINATION topic, however, showed conflicting results. The division of the sub-corpora based on formality markers in the topic analysis was not between the blog and marketing genres, but rather between NCCC, E&theL, and WM on the one side and Strafford and SYGIGH on the other. I believe this shows the need to perform more precise analyses when using corpus linguistics and making claims.

The second matter to discuss is how the genre analysis of E&theL showed that the community (L&E lawyers) values a direct approach and possibly even a repetitive (or at least familiar) structuring. Each of the blogs had a definite focus on L&E law, but the texts in E&theL stayed committed to specific events in the L&E world and discussing the outcomes and meanings of them. The popularity of this blog, based on the two “best-of” lists, shows that such a one-track focus is not a bad thing in terms of how the community responds to the texts. I think this result has meaning for genre studies in that it shows there are reasons certain blogs are one-topic. If the texts are shaped by the community, then in some way they fulfill the needs of the community. The analysis of E&theL shows that those needs may be very specific, which in this case were direct explanations of particular court cases.
The analysis of SYGIGH showed that the community also values texts not aimed directly at them. If we assume that L&E lawyers were the ones voting on the “best-of” lists, then SYGIGH’s placement on them is curious in that the texts directly targeted employees, rather than employers or lawyers representing employers, which was the case with the other blogs and the marketing texts. It is possible of course that a portion of the community is made up of lawyers who represent employees in L&E cases, but only the DISCRIMINATION texts in SYGIGH were directed at that side of the community.

Why weren’t more texts from the marketing sub-corpora aimed at L&E lawyers who represent employees? Do the marketers feel that this section of the community does not have enough to offer them; that they cannot turn a profit from webinars targeted at them? Or was it simply the informal (and one could argue humorous) style that got SYGIGH placed on the “best-of” lists?

The final aspect of this thesis that should be discussed is how the comparison of the corpus linguistic analysis to the genre analysis of the WM blog showed what the corpus annotation software missed. While the WM texts were the most formal based on their lexico-grammatical properties analyzed, they still included very informal language, which was found in the close readings using genre analysis. This is precisely the problem that Archer (2012) warned researchers about when she reiterated Sinclair’s concerns about corpus annotation schemes which are not sensitive to certain linguistic properties. Corpus linguists would be wise to remember Archer’s warning and always analyze their texts more closely to see if texts on specific topics follow the general properties of the corpus. In the case of the DISCRIMINATION texts in the WM sub-corpus, claiming that they are more formal would be misleading.

I think my analysis showed that comparing marketing to blogging really is like comparing apples to oranges. Despite the many similarities of the subgenres in my
corpus (community, topic, etc.), their differing goals or purposes is enough to result in two relatively different sets of texts. This fact is emphasized when we remember that genres are unstable and influential on each other. Whereas the L&E law blogs directly influence the L&E law marketing, the genre of advertising influences (or invades) the genre of blogging.

But these cross purposes and influences need not be a bad thing. Instead, marketers and bloggers who wish to create texts which will be valued by the community would do well to remember what the community expects from them and especially what the purposes of their texts are. For their part, linguists wishing to compare genres should prepare to see variation in and among those they analyze. And they should not expect the implications of former research to hold for every new analysis, especially if that analysis is somehow more specific as it was when I examined the topic of DISCRIMINATION in my texts.

The other benefit to marketers is that further research in this area can provide them with the information they need to write successful marketing in the future. Email marketers have recourse to certain ways of analyzing their copy, such as A|B testing, where a few small groups of people on a large email list are sent different messages selling the same product. The message which does the best, which is the one that sells the most products to these smaller groups, will then be used to market the product to the entire email list. For example, President Barack Obama’s election team successfully used A|B testing with their email list to raise money for his reelection campaign in 2012 (Green 2012). This type of analysis is insufficient, however, in that it can only tell marketers which text performed better. What they should be asking is why that text performed better and what its specific linguistic properties are so that they can mimic the successful style in their future copy writing.
On the linguistics side, there are some important areas that were not discussed in this thesis, but which are open for future research. The first of these is obviously the comments and hyperlinking aspects of the blogs, two aspects of the genre that previous researchers have deemed to be essential to its definition (Blood 2003, Myers 2010). Since the blogs here represent a sub-genre, research into these aspects of them could show whether they follow the larger blog genre, or whether there is an important and interesting variation between these two.

Another area that could be developed further is the software used in email marketing. These days it is easy to automatically append a recipient’s name to the email, or any other personal information that is known. This changes the emails slightly in that they can include personal details about the recipients, and yet still be produced for the mass community. So it would be interesting to see how this technology affects the interpretation of the marketing texts. Do the community members feel a more personal connection with the emails when their name is automatically appended, therefore making the emails more informal? Or are these viewed as technological gimmicks, quickly picked up and brushed aside? If the community sees this automatic appending of their names to the emails, it may serve to place the two at a greater distance from each other, which is exactly the opposite of what the technology is supposed to do.

Finally, there are many more directions to go in both the corpus linguistic and genre analyses. This thesis looked at the relationship between nouns and verbs, and what that says about the formality of the text, but the corpus could be divided in many other ways. For example, the specific pronouns used in the different sub-corpora could be analyzed, a comparison of tense could be made, or the lexical bundles could be researched. On the other hand, a genre analysis could focus on whether the texts are prototypical of email marketing and single-topic blogging. Another possibility for future research, which was
performed but unfortunately not included here due to space constraints, is to use a program called Text Variation Explorer to analyze the type-token ratio, average word length, and clustering of pronouns in the texts (Nevalainen et al. 2011). Changing the window size and looking at these measures in Text Variation Explorer showed some very interesting results. These areas of research would contribute to the analysis by helping to define the corpus better and situating it in terms of the (sub)genres that it includes.
References


**Corpus sources:**

**Webinars:**


**Blogs:**


Appendix 1

Corpus figures and frequencies

Each of the figures from chapters 5 and 6 are listed along with their corresponding tables of frequencies. Figure 14 includes the text counts for the entire corpus, while figure 19 includes the text counts for the DISCRIMINATION topic.

Figure 14. Normalized noun frequencies per thousand words (ptw) in the corpus.
### Adjectives

<table>
<thead>
<tr>
<th>Adjectives</th>
<th>NCCC</th>
<th>Straford</th>
<th>E&amp;theL</th>
<th>FMLAI</th>
<th>SYGIGH</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>94.70</td>
<td>92.47</td>
<td>76.94</td>
<td>64.64</td>
<td>62.44</td>
<td>60.11</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>18.79</td>
<td>15.77</td>
<td>15.02</td>
<td>21.02</td>
<td>18.93</td>
<td>22.68</td>
</tr>
<tr>
<td><strong>Min</strong></td>
<td>66.01</td>
<td>70.56</td>
<td>44.48</td>
<td>25.64</td>
<td>23.81</td>
<td>18.18</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>128.15</td>
<td>131.05</td>
<td>104.60</td>
<td>142.86</td>
<td>105.84</td>
<td>131.15</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>62.14</td>
<td>60.49</td>
<td>60.12</td>
<td>117.22</td>
<td>82.03</td>
<td>112.97</td>
</tr>
</tbody>
</table>

**Figure 15.** Normalized adjective frequencies (ptw) in the corpus.
Figure 16. Normalized verb frequencies (ptw) in the corpus.
Figure 17. Normalized adverb frequencies (ptw) in the corpus.
Figure 18. Normalized pronoun frequencies (ptw) in the corpus.
Figure 19. Normalized noun frequencies (ptw) of the texts that discuss DISCRIMINATION.
Figure 20. Normalized adjective frequencies (ptw) of the texts that discuss DISCRIMINATION.
Figure 21. Normalized verb frequencies (ptw) of the texts that discuss DISCRIMINATION.
<table>
<thead>
<tr>
<th>Adverbs</th>
<th>NCCC</th>
<th>Strafford</th>
<th>E&amp;theL</th>
<th>SYGIGH</th>
<th>WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>39.73</td>
<td>19.99</td>
<td>37.35</td>
<td>53.65</td>
<td>44.52</td>
</tr>
<tr>
<td>SD</td>
<td>10.03</td>
<td>9.89</td>
<td>8.65</td>
<td>14.54</td>
<td>13.92</td>
</tr>
<tr>
<td>Min</td>
<td>29.70</td>
<td>9.01</td>
<td>21.59</td>
<td>31.09</td>
<td>22.22</td>
</tr>
<tr>
<td>Max</td>
<td>49.76</td>
<td>32.99</td>
<td>49.79</td>
<td>87.21</td>
<td>65.16</td>
</tr>
<tr>
<td>Range</td>
<td>20.06</td>
<td>23.98</td>
<td>28.20</td>
<td>56.12</td>
<td>42.94</td>
</tr>
</tbody>
</table>

Figure 22. Normalized adverb frequencies (ptw) of the texts that discuss DISCRIMINATION.
Figure 23. Normalized pronoun frequencies (ptw) of the texts that discuss DISCRIMINATION.
Appendix 2

Examples of the webinar landing pages and blog posts

Figure 24. The NCCC7 landing page. The language outside of the red box was not included in the corpus. URL for the landing page:

http://www.constitutionconferences.com/1R4/0
Figure 25. The Strafford8 landing page. The language outside of the red box was not included in the corpus. URL for the landing page:
Figure 26. E&theL5 as it appears on the internet. The language outside of the red box was not included in the corpus. The URL for the page is: http://employmentandthelaw.com/2011/01/24/in-the-zone-retaliation-under-title-vii/.
Figure 27. SYGIGH12 as it appears on the internet (the image is cut off at the end of the comments section, but the page continues down because the sidebar is so long). The language outside of the red box was not included in the corpus. The URL for the page is: http://employeeatty.blogspot.fi/2011/05/your-employer-cant-discriminate-because.html.
Figure 28. WM26 as it appears on the internet. The language outside of the red box was not included in the corpus. The URL for the page is: