First-year law students' court memoranda assignments: Description of data collection methods and files

Brian N. Larson
Georgia Institute of Technology

Abstract

This file provides a description of the corpus named above.

Keywords: Legal writing, legal memoranda, corpus

This README describes the following with regard to the above-named corpus:

- The circumstances and methods surrounding of data collection for this corpus. (See page 2.)

- A description of the writing samples, in the form of students' Microsoft Word word-processing files. (See page 10.)

- A description of the process for manual annotation in GATE (Cunningham et al., 2012). (See page 12.)

- A description of the participant survey and field values in the XML files after they were "married" to the survey responses. (See page 14.)

- A copy of the participant consent form for this study. (See page 16.)

- Copies of the coding guides for manual annotation. (See page 19.)

Brian N. Larson, School of Literature, Media, and Communication, Georgia Institute of Technology, 316 Skiles, 686 Cherry Street, Atlanta, GA 30332, USA. Email: blarson@gatech.edu.

Brian N. Larson is an assistant professor of rhetoric and technical and professional communication at the Georgia Institute of Technology. Among his research interests are quantitative and computational analyses of text, legal communication, and argumentation theory.

Data collection: Circumstances and methods

Introduction

This corpus was created for a descriptive, observational study applying the tools of statistics and natural language processing (NLP)—and particularly supervised machine learning—to examine texts produced in several sections of a law school legal writing class at two law schools and to assess whether the language of the texts the students produced exhibited differences that varied with their self-reported genders. The students in these classes prepared a year-end memorandum of law—also called a *brief*—with all the students writing a document in the same genre, and in many cases, on the same hypothetical legal case.

This empirical study took as a model Koppel, Argamon, and Shimoni (2002) and Argamon, Koppel, Fine, and Shimoni (2003).[1]

The study had the following specific research questions:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary in relation to their genders?

2. If so, do the differences appear in interpretable patterns?

3. Can machine-learning algorithms categorize the same texts by author gender based on the same features?

4. If so, do they provide interpretable models?

This section explains the circumstances and methods of data collection.

Law school context

In the American legal system, lawyers are trained in post-baccalaureate professional schools, usually for three years of full-time study. The pressures that students in these environments feel to conform to disciplinary conventions in general have been explored in popular fiction and memoir, including the novel and television series *The Paper Chase* (Osborn Jr., 2004) and Scott Turow's *One L* (Turow, 2010). Insiders in the legal education industry have sometimes criticized the legal academy for the stress and confusion it imposes on its students (Caulley & Dowdy, 1986). Educators and law students alike acknowledge the aptness of the old adage about law school education: "first year they scare you to death, second year they work you to death, and third year they bore you to death" (Kahlenberg, 1999, p. 159).

Some researchers have explored students' efforts in law school to function within and conform to the language of the law, both as it is spoken (Mertz, 2007) and written (Cauthen, 2010). These studies have emphasized the challenges that students face and the power dynamics enacted using language in the law school—which are usually presumed to be only a foretaste of the power dynamics of legal language in the courtroom and boardroom.

---

[1]Throughout this document, Koppel et al. (2002) and Argamon et al. (2003) and the underlying data set are occasionally referred to as as the *Argamon/Koppel 02/03 study.*

Based on my anecdotal experiences as a teacher in the legal-writing classroom for eight years, I claim that law students have an intense desire to conform to the disciplinary conventions of the profession into which they are training. These students are mostly very eager and often very bright. They passionately desire to succeed in law school in hopes that it will open doors for the kinds of jobs they imagine they want. In this context, when students are asked to write within recognized genres in their chosen profession, we can expect that they will direct all the effort they can to adhering to the genres' conventions, including linguistic register. Law school calls upon students, regardless of their gender, to leave behind old habits of thought and language and to embrace new ones; the students recognize that their responsiveness to this call may determine their future opportunities.

The data for this study were collected at two law schools in the U.S. Midwest during the 2011-12 academic year. One of these schools, referred to here with the pseudonym "Academy School of Law," is routinely ranked among the top 35 law schools by popular national assessments such as *U.S. News and World Report* and *Above the Law.* The other, referred to here as "Lyceum Law College," is not routinely ranked among the top 100 schools accredited by the American Bar Association. According to the administrations at these two schools, they enrolled a total of 545 new students in AY2011-12; of them, 263 were female and 282 male according to law school records. Each school required as part of its first-year curriculum several basic courses, including contracts and civil procedure. Importantly, each also required students to take a course or combination of courses in legal research, analysis, and writing.

It is in this context that I collected writing samples from 193 gendered authors and created the text corpus that was the object of analysis for the study. The research questions posed above call for texts written by single authors of different genders working in a context where the authors would be attending closely to, and attempting to adhere to, conventions of a single disciplinary genre.

Texts in a professional genre. I have proposed elsewhere (Larson, 2015) that *genre* as a research construct is the application of a category label to a set of texts exhibiting a loosely and culturally defined set of communicative behaviors, usually formal conventions, a Speaker or Writer expects to have a particular effect or effects on a Hearer or Reader, based on assumptions about a typified situation in the Speaker's imputed cognitive environment. In the present study, there is evidence that the participants, all students finishing their first year of training in law school, shared certain elements of their cognitive environments, including accessible, though possibly only weakly held, assumptions about the formal conventions of legal writing and of the hypothetical memoranda they were writing—the typified situation; intense and accessible goals to do well in this important assignment; and assumptions about the cognitive environments they imputed to their instructors. I argue that these elements, taken together, show the students in this study were all writing in the same *genre.*

This section describes how the first-year students at Academy School of Law and Lyceum Law College prepared such a set of texts, first describing the legal writing programs and then the year-end brief or memorandum assignment.

According to officials at at these law schools, the first-year legal writing classrooms at Academy School of Law and Lyceum Law College shared some characteristics and differed in others. At Academy School of Law, students were grouped in 25 sections, with each section

having between eight and ten students and each having an adjunct attorney instructor, usually a practicing attorney from the community, and a student instructor, a second- or third-year student acting as an "upper-level student teaching assistant." The syllabus and assignments for the year were controlled from a central legal writing administration. Thus, for the spring assignment that is the object of this study, all the students at Academy School of Law wrote about the same hypothetical problem. Required texts at Academy School of law were Clary and Lysaght (2010) and *The Bluebook: A Uniform System of Citation* (2011). I should note here that I have taught the course that is the locus of this study at Academy School of Law for eight years, though I did not teach it the year that I conducted this study. I have, from time to time, made observations in this document grounded in my intuitions or anecdotal experiences; where I have done so, I have tried to acknowledge the source of those observations and distinguish them from observations gathered by more systematic means.

Lyceum Law College also grouped students into small sections, in its case, 28 sections of nine to twelve students. There, however, each section was taught by a single adjunct professor, again usually a practicing attorney, but with no student teaching assistant. Furthermore, legal writing professors at Lyceum Law College were responsible for developing their own hypothetical problems for students to write about, within certain constraints established by the school's legal writing program. Required texts at Lyceum Law College included Schmedemann and Kunz (2007) and *The Bluebook: A Uniform System of Citation* (2011).

According to administrations at these law schools, each school required students to write a spring capstone assignment, typically an example of what lawyers call a "motion practice brief": The students wrote memoranda of law in support of or opposition to a hypothetical motion seeking dismissal of a claim or summary judgment on a claim. At Academy School of Law, each student wrote a memorandum supporting or opposing a motion to dismiss a hypothetical copyright claim. At Lyceum Law College, students' memoranda supported mostly motions for summary judgment and a few for dismissal; the legal subject matter of these hypothetical cases varied from contracts and negligence to civil rights and the First Amendment. Students were given page limits for their assignments, with none of them being permitted to write more than 20 double-spaced pages.

According to their responses to an email survey regarding teaching perspectives, legal writing instructors and professors at both schools shared many perspectives on teaching this year-end writing assignment. For example, many of these instructors/professors claimed that they had not discouraged students from using long quotations from cases (sometimes called "block quotes") and footnotes, but they also noted that most students had avoided frequent use of these rhetorical techniques. Generally speaking, citations in legal writing of this kind are in-line: all the relevant bibliographic information is included in a citation sentence or clause immediately after the name of the cited material or the assertion in the text that the cited material supports. The following is an example from paper 1019:

> When a statute's plain language is ambiguous, a court may use legislative history to help determine Congress's intent. *See Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47 (2007). It is unnecessary to analyze the legislative history in this case because the text of § 101(2) is unambiguous and does not require a signed writing prior to the creation of a commissioned work. The legislative history

> does, however, provide further support for this conclusion. Committee Reports are the most authoritative source of legislative history. *Tellabs, Inc. v. Makor Issues & Rights, Ltd.*, 551 U.S. 308 (2007).

Note that in this example, the student cited two cases—*Safeco* and *Tellabs*—in citation "sentences" following the textual sentences that rely on the cited cases. Note, too, that there are no attributive cues in the textual sentences; conventionally, the assertion preceding a citation is attributed to the majority opinion in the case cited, unless certain special markers are used.[2]

Instructors/professors generally did not provide models for the types of briefs the students were to write. The textbook prescribed by each law school included one or two model briefs of the appropriate kind. Of ten instructors/professors who responded to an email interview about their teaching, only two supplied other examples, and both said they did so not to provide models of good brief-writing but rather to show what such briefs look like in practice. The students could use online research tools to find examples of briefs actually filed by lawyers in real cases, but the legal writing instructors did nothing to mediate students' assessments of the quality of such models, so it would have been difficult for students to select models, other than the textbook examples, upon which to base their own briefs. Nevertheless, students had been steeped for the better part of an academic year in reading court opinions; such documents are not written for the same purpose as memoranda, but students could be expected to model some of their linguistic practices on the opinions they had read.

There is evidence the first-year law students at Academy School of Law and Lyceum Law College, though they no doubt varied a great deal in their personal characteristics and backgrounds, were all writing with very similar components of their cognitive environments accessible. Their training for the previous year prepared them with accessible assumptions about the typified situation of the memorandum and of legal writing in general. Their awareness of the importance of this assignment made their goal of success on it both accessible and strong. And their expectations of their instructors' expectations—the cognitive environments they imputed to their instructors—equipped them to adjust their writing styles to achieve their goals. Their year-end briefs are thus all of a single genre. This is true even across the law-school boundaries, owing to the similarities in the final assignments between the two schools and among the legal writing professors at Lyceum Law College. Of course, it would be ideal to supplement the data in this study with qualitative interviews with the students to support (or undermine) this speculative evidence.

At least some of the conceptions of genre might also call for the type of writing in question to be one that the writer engaged in repeatedly, the "conventional category of discourse based in *large-scale typification* of rhetorical action" described by Miller (1984, p. 163, emphasis mine). Or they may place the generic status of these texts in question because the classroom context makes the writing produced there "pseudotransactional" (Spinuzzi, 1996). Despite these concerns, students in these classes probably expected in the future to write texts in the genre or genres in which their assignments occurred. Their efforts to produce texts in a professional genre, even relating to hypothetical problems, likely constituted

---

[2]Law-school-trained readers may note that "pincite" page numbers are missing from the citations in this example.

efforts with intense and accessible goals to evoke a reader response (here, from practicing attorneys acting as legal writing instructors/professors), based on the students' accessible (though perhaps weakly held) assumptions about instructors' expectations. Thus, though these students' efforts may fall short of satisfying the technical definition of genre espoused by some scholars, the students' work certainly represents a more motivated response to a shared rhetorical situation than many previous studies.

Texts by single authors. Researchers now often contemplate authorship as a collective and distributive activity. Blog posts are ghost-written. Twitter accounts are ghost-written and shared. Even published fiction is subject to concerns that editorial involvement in texts makes them collaboratively authored; the works of an author like Iris Murduch, whose resistance to editing makes them truly single-authored, are a rare exception (Pakhomov, Chacon, Wicklund, & Gundel, 2011). It is my experience that in the professional context of law, court briefs often have many attorneys who claim authorship of them; a brief as filed might easily have four or five authors. Even the listed authors of a brief may not tell the story of authorship, given that associates in the law firm may be called on to draft segments of a brief edited, signed, and filed by a more senior lawyer.

Assessing gender differences in writing, however, demands that the texts studied be written by single authors, each of a gender recognized for purposes of the study. The writing assignments of first-year law students at Academy School of Law and Lyceum Law College address this concern because the schools limited students' ability to work together, and the structure of the assignments makes it unlikely that students will procure writing from outside.

In my view, the legal writing programs of both law schools in this study emphasize individual effort and assessment of the individual. Given the collaborative environment in which many professional legal briefs are written, this may seem strange. But law school is often an extraordinarily competitive environment; in the old days, it is reputed that students would intentionally misshelve books in the library to prevent their peers being able to use them for assignments (Turow, 2010). In fact, Academy School of Law's student honor code still expressly prohibited that practice in 2012. Legal employers are also acutely interested in students' class standing and individual level of achievement. So perhaps policies that prevent first-year students collaborating on writing and honor code provisions at both schools that assess harsh penalties for students working together are no surprise. They give rise to a much stronger presumption of single authorship than can be asserted with regard to most previous studies.

Law students are also unlikely to be able to procure writing assignments from online banks of papers sold by other students (Ariely, 2012; Hansen, 2004). The law school writing assignments relate to complicated hypothetical problems, often involving case files with excerpts of evidentiary exhibits and testimony. No stock paper purchased online could ever hope to address the issues the students must take up in their writing assignments. Even if an instructor used a very similar hypothetical case from year to year, she need only make a slight change in the supporting materials to require the next year's students to take a much different tack in their analyses. Of course, it is possible that one law student could pay another or some third party to write her brief based on the current year's case materials. The amount of time required to do so makes it unlikely most law students could afford such a service; and the consequences for another law student to take on the task if

she is caught make that unlikely, too. My own law students have occasionally told stories (always unsubstantiated, so far as I know) of other students who have parents or siblings who are lawyers who provide substantial editing services. Such circumstances would no doubt change the textual characteristics, but ghost-writing is a potential problem with any text not written before the researcher's eyes.

As this subsection has shown, the collection of samples from the first-year law students at Academy School of Law and Lyceum Law College resolves the single-author problem, at least to a reasonable degree of probability.

Authors who identify their own genders.   A study of gender differences in writing ought to be very sensitive to the way it identifies writer genders. Previous studies of gender differences in written communication suffered from limitations in this area. For example, some researchers relied unquestioningly on third-party assessments of author gender (Argamon et al., 2003; Koppel et al., 2002). Others used aspects of authors' computer-mediated communications to assess their genders and then used the resulting gender assessments to argue that aspects of the communicative performances varied with them (Rao, Yarowsky, Shreevats, & Gupta, 2010). Still other studies had authors take gender-role assessment tests that raise serious concerns about gender stereotyping and a failure to address diachronic change in gender roles in American culture (Janssen & Murachver, 2004). These approaches might be described as "black-box," "question-begging," and "stereotyping" assignments of gender; this study avoided them by asking authors to identify their own genders.

Elsewhere (Larson, 2015), I have described the gender construct in this study as a loosely and culturally defined set of social behaviors that are expected to make it possible to distinguish the two most common sexes from each other. I noted there that this study asked authors to identify their own genders.

But even that approach poses problems because people generally do not have a sophisticated understanding of what *gender* means. They fill out surveys, questionnaires, medical forms and the like that ask them to specify their genders. Such instruments typically offer two choices, "male" and "female." But from some theoretical standpoints, it may be inappropriate to refer to these labels as gender labels as opposed to sex labels, while other theorists would oppose a bright line dividing sex and gender labels. And the average person, probably even the average law student, is not aware of these debates. A further problem arises if one considers transgender persons. It is unclear where they are to check if given the option of two genders: male/female or masculine/feminine. For me, it is difficult to see how adding an "other" or "none of the above" option shows respect for research participants in my study.

The solution I chose for this study was to allow participants to identify their genders in a free-form questionnaire field in an online survey. In other words, students were asked their genders and allowed to write whatever they chose in response.[3] Of the 197 students who participated in this study, 193 responded to this question. Table 1 shows the results. (See page 14 for the survey instrument.)

As Table 1 shows, allowing for a free-form response creates a new problem: A proliferation of gender labels. Four different responses—*F, Fem, Female*, and *female*—came from participants who might describe themselves as being of a "female" or perhaps "feminine" gender. Four other responses—*Cis Male, M, Male*, and *Masculine*—came from participants

---

[3]I'm grateful to Dr. Christina Haas for suggesting this elegant solution.

Table 1

Self-reported genders of participants in present study (*n* = 197)

Gender	Number	Percent of total
Cis Male	1	1%
F	5	3%
Fem	1	1%
Female	95	48%
female	3	2%
M	3	2%
Male	84	43%
Masculine	1	1%
Not answered	4	2%
Total	197	

Percentages rounded to nearest whole number,
resulting in total tally of 103%.

who conceivably consider themselves of a "male" or "masculine" gender.[4] Of course, while a researcher might presume that "F" was meant as "female," that may not be what the participant intended.

Rather than impose the associations of traditional gender identities on these participants, this study takes the approach of establishing an *ad hoc* research construct, in which authors may be assigned to "Gender F" or "Gender M." Authors who gender-self-identify with any designation beginning with the letter "F" (not case sensitive) are classified as Gender F. Those who self-identify with any designation beginning with the letter "M" (not case sensitive) are classified as Gender M. The prefix "cis" is ignored. Had there been any participants who used "tran" or "trans," they could have been classified as "Gender T."

All gender classifications are problematic and suspect. They are also subject to changing gender landscapes and expectations. Given that the common understanding of gender is that there are two (with possible accommodation for those who are transgendered or prefer to be ungendered) it is not unreasonable to group gender self-identifications based on two categories with similar linguistic features (namely their initial phonemes or graphemes). Though it comes with some challenges, it warrants greater credit than the gender-category assignments in the studies mentioned above.

In this section I have made the argument that the texts collected from students at Academy School of Law and Lyceum Law College are of a single professional genre and written by individual authors; and that 193 of them can reliably be labeled as being written by either Gender F or Gender M. The next section describes how these data were collected and prepared for analysis.

Data collection

Students at two law schools in the U.S. Midwest, referred to here with the pseudonyms Academy School of Law and Lyceum Law College, prepared a major writing assignment

---

[4]The term "cismale" derives from gender studies, where it is used to refer to a person of the male sex who identifies with the masculine gender. Cisgendered persons thus contrast with transgendered persons in the congruity of their biological sex and the gender they feel or enact (DeFrancisco, Palczewski, & McGeough, 2014, p. 60).

at the end of their first year in law school. I approached the directors of the legal writing programs at these two schools in the fall of 2011 and obtained their support for this research; they cleared it with their administrations. After obtaining IRB approval for this study, I collected information regarding the structure of the course in which the writing samples were created by means of interviews with administrators of these programs and documents that they provided me. This included information from the legal writing programs and instructors regarding the texts, assignment prompts, and model documents; as well as supplemental materials the instructors provided, whether they encouraged students to find and review examples of briefs of the kinds they were drafting, and the extent to which they emphasized various mechanical issues (grammar, citation, argument structure) in their instruction. That information provided valuable context that was described above.

This subsection describes the collection of the student papers, including a summary of the process for collecting data from students via an online survey and a brief description of the samples of writing collected. The legal writing program administrators of the two law schools cooperated in transmitting the invitation to participate in this study to the eligible law students at their schools. They arranged for me to provide them with the text of the invitation message, along with the *Information Sheet for Research*, a copy of which appears on page 16, and a link to the survey instrument, which appears on page 14. In spring semester 2012, the legal writing programs then sent the initial invitations about the time the final brief was due to be completed and followed up at weekly intervals for less than a month. Participating students were offered a $15 Amazon.com gift card for completing the survey and uploading their writing samples.

The survey instrument was developed according to the procedures outlined in Murphy (2002), using a process similar to that used by Eaton, Brewer, Portewig, and Davidson (2008) for an online survey. The survey instrument was hosted on Wufoo (http://wufoo. com), which permitted students to upload their writing samples at the beginning of the survey. The survey instrument is reproduced at page 14. It asked questions regarding student age, gender, highest previous degree, most recent writing course, and how the student learned English. It also asked information about which section the student was in, so that this information might later relate the practices of particular teachers to peculiarities among their students's papers, if any.

In all, 197 students completed the survey. According to law school records, 545 students were eligible; there was thus a response rate of approximately 36%. Though all questions on the survey were optional, 193 students provided information about their genders that could be interpreted according to the approach described in above. Based on that approach and the responses, which are detailed in Table 1, the respondents were categorized into Gender F ($n = 104$) and Gender M ($n = 89$). The analyzed segments of student's briefs varied in length from 2,303 to 5,035 word tokens (including punctuation), with a mean length of 3,764 tokens. With one exception, all the papers were in Microsoft Word file formats; the exception, a PDF file, was converted into MS Word format using commercially available software. Before any other work on the briefs/memoranda, each file was reviewed to systematically remove all information that would identify the student author or the law school from the text of the file itself and from the file metadata.

The student participants and their papers are designated in the data and throughout this dissertation according to a four-digit number assigned during anonymization. Paper

numbers beginning with "1" originate with Academy School of Law and those beginning with "2" originate with Lyceum Law College.

Ethical and legal considerations

The consent form for this study was based upon contents suggested in Breuch, Olson, and Frantz (2002, p. 11), to meet the requirements of the University of Minnesota's Institutional Review Board, which were set out in its Protecting Human Subjects Guide (Board, 2004, p. 5) and on its web site. See page 16 for the complete consent form. The University of Minnesota IRB/Human Subjects Committee determined that this study is exempt from review under federal guidelines 45 CFR Part 46.101(b), Category #2 "surveys/interviews; standardized educational tests; observation of public behavior" (Study Number: 1202E10685).

Because this study involves the copying and transformation of texts by students that they have fixed in a tangible medium of expression, the sample texts are subject to U.S. Copyright law, Title 17 of the United States Code. My intention was to publish the texts so that other researchers can reproduce this study's efforts using the same texts. The consent form included a grant of a license from each student for those purposes for this study and for any similar studies conducted by researchers using the same texts.

Writing samples described

A preliminary assessment and review of the writing samples showed that the students had followed a largely formulaic approach to high-level structure similar to that suggested by the samples in the course textbooks. The memoranda were double-spaced, and each began with a caption of the kind shown in Figure 1 and concluded with a signature block like that shown in Figure 2. Some papers, including 1007, 1025, 1044, and 1098, had front or back matter that was not part of the memorandum itself. These elements included formal pleading documents like the motion, notice of motion, and certificate of service. In the brief or memorandum itself, the structure was highly consistent:

- Caption: Every brief exhibited this.

- Introduction or summary: Not all briefs had this section (see papers 2084, 2091, 2093). In those that did, this section consisted of a brief introduction to the substance of the memorandum and the relief that student-attorney's client was seeking from the court. It may sometimes have been styled by the author as an "Issues" section (see papers 2026, 2095).

- Facts: Every brief exhibited this section, though it may have gone by other names, such as "factual background," "undisputed facts," and the like. In each memo, this section provided the facts of the instant case. This section was always the second longest section in the brief, after the argument section.

- Legal standard or standard of review: This section was not always present. If present, it articulated the standard for summary judgment or a motion to dismiss, the basis upon which the court would have to decide the motion. It was sometimes styled as

"Procedure" (see papers 2057, 2086). Sometimes the content typical of this section appeared at the beginning of the argument section instead.

- Argument: This section appeared in every memo, and it was always the longest section. In this section, the student-attorney argued how the law, applied to the facts earlier discussed, should result in her client obtaining the relief requested of the court.

- Conclusion: The great majority of briefs included a section set off by a "Conclusion" header. Usually a paragraph or two at most, the conclusion reiterated the relief that the student-attorney was seeking from the court for her client and sometimes summarized the main points from the argument section.

Figure 1. **Caption from student brief**

Figure 2. **Signature block from student brief**

Though students were discouraged by their textbooks from using footnotes in their briefs, some still chose to do so at least a few times (see papers 1035, 1043, 1070, all from Academy School of Law; footnotes were hardly present at all in papers from Lyceum Law College). Many students, however, used "block quotes," quotes of 50 words or more that conventionally must be indented on both sides and appear single spaced (*The Bluebook: A Uniform System of Citation*, 2011). Some used quite a lot of block quotes (see papers

1014, 1113, 2024, 2041). Figure 3 shows an example of such a quotation. All students used at least two levels of headings, one for the major sections identified above and one for key segments of their arguments. Some students used more levels of headings.
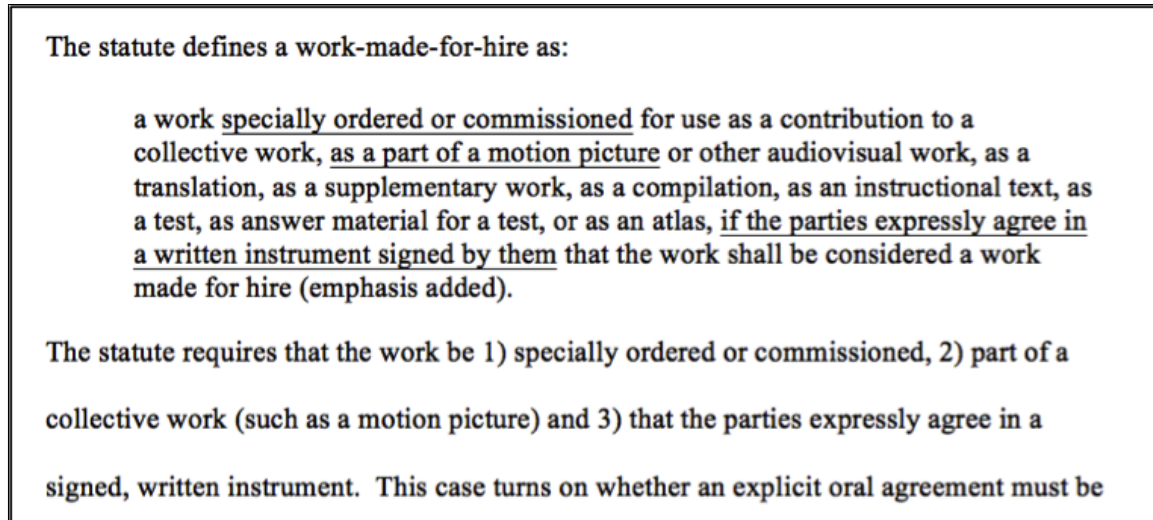
> The statute defines a work-made-for-hire as:
>
> > a work <u>specially ordered or commissioned</u> for use as a contribution to a collective work, <u>as a part of a motion picture</u> or other audiovisual work, as a translation, as a supplementary work, as a compilation, as an instructional text, as a test, as answer material for a test, or as an atlas, <u>if the parties expressly agree in a written instrument signed by them</u> that the work shall be considered a work made for hire (emphasis added).
>
> The statute requires that the work be 1) specially ordered or commissioned, 2) part of a
>
> collective work (such as a motion picture) and 3) that the parties expressly agree in a
>
> signed, written instrument. This case turns on whether an explicit oral agreement must be

Figure 3. **Block quote from student brief**

As a result of the processes described in this section, there were 193 texts written by law students at the end of their first year of law school, with each text classified by its author's self-identified gender.

Data preparation: Annotation

The study for which this corpus was created was performed using statistics and machine learning algorithms and the lexical and quasi-syntactic text features used in Koppel et al. (2002) and Argamon et al. (2003). Before these analyses were performed, I wished for some text segments to be excluded from analysis. In order to exclude the undesired segments for my study, I chose to mark the writing samples up in GATE (Cunningham et al., 2012), which allowed me to create extracts of the samples without the undesired segments while leaving the original texts intact.

Manual annotation of the texts to permitted me to identify segments of text that either would not be analyzed or those that would be held out from analysis at least temporarily. Working with a research assistant,[5]I developed a coding guide for manually annotating the papers in hard-copy using legal briefs other than those submitted by participants in this project. We then transferred our annotations from the paper copies to electronic copies using the General Architecture for Text Engineering or GATE (Cunningham et al., 2012). GATE is open-source software and available free of charge from the University of Sheffield.

Manual coding of paper copies involved two levels of the text's structure. At the large-segment level, we marked each of the following portions of each text, including any heading at the beginning of it:

- Caption: This is the formulaic block shown in Figure 1 at page 11.

- Tables: Though none of our student papers included tables of contents or tables of authorities, some courts require them in filed briefs and some attorneys provide them whether they are required or not. Because we developed the coding guide based on "published" attorney briefs, we had this segment type, but never used it in the context of this study.

- IntroSum: This included any introduction or summary immediately after the caption.

- Fact section, described above.

- Argument section, described above.

- Conclusion section, described above.

- OtherText: This is any material between the caption and the conclusion that does not fit any of the other large segments.

- OtherFormal: This is material before the caption or after the conclusion, usually consisting of pleading documents, such as motion and notice of motion, and the student's signature block after the conclusion.

See page 19 for the complete coding guide. I determined at once that I would not analyze materials (such as formal pleading documents) incidental to the brief and that I would not analyze the caption or signature sections because of their highly formulaic nature.

Within the large segments, we coded many other segments of text:

- Heading: The heading at the beginning of a section or subsection of the memo.

- Cite: This is any legal citation. These were coded depending on whether they were sentence citations (standing outside a textual sentence) or clause citations (appearing within a textual sentence). They were also coded by how many authorities were cited in a given citation. This measures (at least in part) the tendency of lawyers to employ "string cites," long citations of multiple authorities with little text to explain their purposes.

- BlockQuote: This is any quotation of 50 words or more, indented as required by legal writing conventions.

- Footnote: Any footnote reference or footnote text in the memorandum.

See page 19 for the complete coding guide. I decided to exclude section heads from my analysis, as I was uncertain of their linguistic status. I also excluded block quotations from my analysis, as they represent long stretches of text not composed by the students. I did not attempt to remove smaller quotations embedded within a student's text. So, for example, the following sentence appears in paper 1102:

> The general rule under the Copyright Act is that a "work protected under this title vests initially in the author or authors of the work."

My view is that such a sentence presents a hybrid of the student's language and the language of the quoted text because the student integrates her original composition with that of the quoted text. I did not attempt to identify the instances where students use such quotations frequently or where they constitute a large percentage of the student's paper.

We transferred this coding from the paper copies to electronic copies of the briefs in GATE. The process for doing so is described at length in the second coding guide. I performed a check of inter-rater reliability on 10 of the papers (a little over 5% of them) to see whether the research assistant and I were consistently coding text spans the same way. I assessed inter-rater reliability using an F-measure with a β of 1. Using the IRR capability embedded in GATE, I assessed my codes as the *key set* and the research assistant's as the *response set*. Recall thus measured the percentage of spans that I annotated that were annotated in the research assistant's; precision, the percentage of spans the research assistant annotated that were annotated in my work. This effort is somewhat complicated by the fact that coders were not just assigning codes—which might be different depending on coder—to text spans, but they were also identifying the beginning and end of each span—which might not overlap exactly depending on coder. GATE provides for calculation of strict, lenient, or average agreement: For strict agreement, the text spans must overlap exactly and the codes assigned must be the same. For lenient agreement, if any part of the spans overlaps and the codes assigned are the same, the code is counted as a match. Average agreement counts codes where the spans do not overlap perfectly as half a match.

For this project, I was not worried about spans overlapping perfectly. If one coder included a space that the other did not, it was unlikely to affect the outcome of the project. On the other hand, it seemed very important that we were identifying the same spans on spans that would be excluded from analysis. When I did the analysis, I needed to be able to exclude citations and "OtherFormal" text from it completely, and the only way to do that is if we have carefully annotated them. So, I set these targets for IRR F-measures: strict > .80, lenient > .95, and average > .90. There is a variety of ways to run these tests with GATE, but our test instances met these thresholds in each case. When I examined the specific bases for disagreement, almost all were slight differences in span length, usually the inclusion or exclusion of a single space. I also noted that papers we annotated later had higher agreement than those annotated earlier.

Survey instrument for gathering data/key to interpret XML

Participants in the empirical study responded to a survey, and the results are encoded in the XML file for each paper.

Student survey [XML element: "Questionnaire"]
Please upload your year-end legal writing assignment by clicking on the link at the [right/left/above]. Please answer the following questions. You are not required to answer any of these questions, but your answers may assist in making the research results more useful.

[XML element: "Age"] Age: [multiple choice consisting of following options]

A Under 18

B 18-24

C 25-33

D 34-45

E 46+

[XML element: "SI_Gender"] Gender: [blank box permitting a response in which participant self-identifies for gender]

[XML element: "Education"] Highest level of education you have completed: [multiple choice consisting of following options]

A Bachelors degree (U.S. institution)

B Bachelors degree or equivalent (Institution outside U.S.)

C Law degree (Institution outside U.S.)

D Master's Degree, post-baccalaureate professional degree, or equivalent

E PhD or equivalent

F Other [blank box permitting response]

[XML element: "LastWCourse"] Before your current legal writing/research course, when is the last time you took a course that you would describe as a "writing course": [multiple choice consisting of following options]

A I have never taken any other writing course.

B I took a writing course in a post-baccalaureate degree-granting program.

C I took a writing course as an upper-level undergraduate in the U.S.

D I took a writing course as a lower-level or freshman undergraduate in the U.S.

E I took a writing course as a student at a university outside the U.S.

F I took a writing course in secondary school (high school, for U.S. students)

G Other [blank box permitting response]

[XML element: "LearnedEng"] Describe how you learned English:

A I learned English in the U.S. as my first language

B I learned English in the U.S. as my second (or subsequent) language

C I learned English outside the U.S. as my first language

D I learned English outside the U.S. as my second (or subsequent language)

[XML element: "Section"] What is the section number of your legal writing class in law school? [blank space permitting response]. (This information is likely of little value to other researchers. I collected it to see if there were patterns across the papers submitted by individual students.)

What is the last name of your legal writing professor or instructor? [blank space permitting response] [This information is not provided in the XML files.]

Email address: (You must answer this question in order to receive your $15 Amazon gift card) [blank box permitting student to enter email address] [This information is not provided in the XML files.]

The XML also includes the following elements:

- XML element: "PlorDef". This indicates whether the student brief was written on behalf of a hypothetical plaintiff or defendant.

- XML element: "Anonymize". This indicates whether the paper has had personally identifiable information removed. (All papers should be "Yes+".)

- XML element: "Analysis Gender". This indicates the gender assigned by the researcher based on the SI_Gender element: 0 for Gender M and 1 for Gender F. See discussion above.

- XML element: "Topic". This is the researcher-assigned value indicating the legal topic addressed by the participant's brief.

- XML element: "Genre". This is the researcher-assigned value indicating the type of legal document required for the assignment.

Participant consent form

The *Information Sheet for Research* distributed to students with the invitation to participate in this study appears on the following two pages.

INFORMATION SHEET FOR RESEARCH
Analysis of Law Student Writing Assignments

You are invited to be in a research study of law student writing assignments. You were selected as a possible participant because you are a first-year law student enrolled in a legal writing course. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Brian N. Larson, J.D., Writing Studies Department, University of Minnesota ([email]).

Procedures:

If you agree to be in this study, we would ask you to do the following things:
* Indicate your consent below to proceed to a brief online survey.
* When the survey begins, you will be asked to upload a copy of the final version of your major spring writing assignment in law school.
* The survey itself will ask you some demographic questions and will take no longer than 12 minutes to complete.
* When you have completed the survey, you will be able to provide your email address, which is where the researcher will send your code for a $15 Amazon gift certificate, provided in gratitude for your willingness to participate in the study.
* After any identifying marks (your name, address, phone number, email address, etc.) are removed from your writing sample, it may be published as part of a database of student papers that other researchers may use for other projects. By consenting below, you are consenting to the ongoing use of your writing sample by other researchers.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. After any identifying marks (your name, address, phone number, email address, etc.) are removed from your writing sample, it may be published as part of a database of student papers that other researchers may use for other projects. By consenting below, you are consenting to the ongoing use of your writing sample by other researchers.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota or [Name of law school]. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher(s) conducting this study is (are): Brian N. Larson and his supervisor, Mary Lay Schuster. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact them at Department of Writing Studies, University of Minnesota, [phone], [email]. Larson's advisor, Mary Lay Schuster, is available at the Department of Writing Studies, University of Minnesota, [email and phone].

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Coding guides

The coding guides used in manual annotation of the data in GATE appear on the the following pages.

Paper coding guide

Researcher: Brian N. Larson
Revised July 9, 2013

Overview

The purpose of this coding project is to create annotated versions of memos written by law students and professional attorneys; the annotations identify parts of the memos like large sections, text headings, and citations to legal authorities. This annotation process will help to create a *corpus* (pl. *corpora*) of texts that the researcher will use for various projects.

The process will consist of two steps:
1. You will read and mark the memos in paper form.
2. You will record the annotations on a computer using software called "GATE: General Architecture for Text Engineering."

This document describes the first step, marking of memos in paper form.

Marking document segments

In this first phase, you will read and mark text segments in paper memos. For those segments that have types LargeSegment and Cite, you will indicate which type each instance is. For Cite segments, you will identify the number of authorities for each cite.

Use whatever hand annotations are convenient for you to mark the paper copies. You will sit down with the researcher and compare notes after you complete some samples.

You will find it helpful to mark each kind of segment separately. For example, mark beginning and ends of all LargeSegments before moving on to marking Cites.

Keep a separate journal about your experiences. Note any challenging coding in your journal, making reference to the paper number and page number when you have a problem. Assign codes in any case, using your best judgment and making a note. (It's better to assign codes even if you are unsure whether you should.)

Document segments defined

LargeSegment: A LargeSegment is a large 'chunk' of the text of the memo. Every memo is divided into several LargeSegments. Every portion of each memo is included in one LargeSegment or another. The following are the LargeSegements possible in these memos:

Caption: This is the 'top' of the memo as it would be filed with a court. It includes the name of the jurisdiction and court, usually in block capitals, the names of the parties to the litigation, and the title of the memo itself. This section is often (though not always) separated from the rest of the document by horizontal lines or a "box" around it. The title of the memo is included in this segment, even if it appears immediately following the horizontal line or box. In student memos, the Caption may be preceded by a title page that indicates a word count or other information; that front matter should be marked as LargeSegment:OtherFormal. This section appears in every memo.

TOCTOA: This type of segment includes any table of contents or table of authorities appearing in the document. A table of contents outlines the contents of the document, usually showing page numbers where headings appear. A table of authorities is a list of legal (and possibly other) authorities cited in the document.

IntroSum: This introduction or summary appears immediately after the Caption. It is usually two or three paragraphs at most. It usually has a header titled "Introduction," "Summary," and less commonly, "Procedural Background," but it may follow the caption directly without any header at all. It may include a sub-section titled "Issues," identifying the

issues before the court. This section appears in almost every memo.

Facts:
This section almost always follows the IntroSum section and is almost always titled "Facts" or "Factual Background," but may have names like "Stipulated Facts" or "Undisputed Facts." The key component here is the term "Fact" in the header and a recounting of the facts associated with the case. This section appears in every memo.

Argument:
This section follows the Facts section. It usually, but not always, begins with a heading titled "Argument" or something similar. It may occasionally not be marked by such a heading but instead begin with a heading marking the beginning of the memo's argument. For example:

> **UNDER 17 U.S.C. § 101(2), THE WRITTEN INSTRUMENT NEED NOT BE SIGNED PRECEEDING CREATION OF THE WORK AS LONG AS THERE IS A PRIOR EXPRESS ORAL ARGEEMENT.**

This heading signals a shift from the fact section, which generally does not make reference to the law or to legal conclusions. If such a heading appears after the beginning of the Fact section, it signals the beginning of the Argument. This section appears in every memo.

Conclusion:
This section appears after the Argument Section. (Some might consider it part of the Argument Section, but you should treat it as the beginning of a new section.) It usually begins with a heading titled "Conclusion" or words to that effect. Not all memos will have a separate conclusion section.

OtherFormal and
OtherText: Any material that does not fit into the other LargeSegments identified here should be marked as "OtherFormal" or "OtherText."

OtherFormal covers any front matter before the Caption or after the Conclusion and other pages that may be styled as "Motion," "Notice of Motion and Motion," "Certificate of Service," "Proposed Order," or the like. It includes any signature block or formulaic closing where the attorney says "Respectfully submitted" (or words to that effect), indicates the date of the filing, and provides her name, firm, and contact information, where applicable. (Almost all memos will have such a signature block.)

OtherText is for segments of text between the caption and conclusion that do not appear to fit into other large segment categories, such as "Standard of Review," "Summary of legal principles," etc.

Heading: This indicates a heading at the beginning of a section or subsection of the memo. Headings are sometimes complete sentences. Headings are almost always set off from the surrounding text by one or more of the following typographical conventions: bold, Italics, underlining, centered, numbered or lettered sequentially. Headings may be marked by different typographical conventions within a single document; for example, one level of heading might be bold and another Italics. Do not identify headings within the Caption section. Headings may appear at different "levels," usually distinguished by the use of different typographical conventions or different numbering lettering. The heading is included in the LargeSegment it precedes. It is possible for to headings to appear consecutively, in which they should be coded as two consecutive headers; for example:

ARGUMENT

I.	OGS PLED SUFFICIENT FACTS TO ESTABLISH A VALID CLAIM FOR COPYRIGHT INFRINGEMENT ON WHICH RELIEF CAN BE GRANTED

Cite: This is a reference to a text or authority outside of the memo. It may be a reference to a case, statute, or other authority. It is always set off from the grammatical portion of a sentence by commas or other punctuation. It is sometimes bracketed by parentheses. For example, the cites in the following sentences are highlighted:

> As children reach adolescence, courts recognize that the process of graining independence is an important consideration in determining duty and reasonable care. ==Restatement (Third) of Torts: Affirmative Duty § 42 (Tentative Draft No. 4, 2004).==
>
> Lime is a well known screenwriter with fifteen years of experience in television writing==, (Compl. ¶ 11.)==, and OGS sought to commission Lime to write an episode of *Lawless Love*==, (Id. ¶ 8.)==.

The two preceding examples also illustrate two kinds of Cites:

Sentence: A Sentence Cite is punctuated as a complete sentence set off from the rest of the author's text. (Like the *Restatement* cite in the previous examples.) A sentence cite can be very short. ==(Id.¶24.)== or ==(Id.).== It is also possible for two citation sentences to appear in a row.

Clause: A Clause Cite appears within one of the author's prose sentences but it set off from it by commas (or sometimes a comma and a semi-colon). (Like the Compl. and Id. cites in the previous examples.)

A citation may include explanatory material in parentheses or an explanatory clause. Parenthetical and explanatory information is included in the cite and should be marked as

part of the cite. Citations that appear in footnotes should be treated just like citations in the text.

Each citation refers to one or more authorities. When you mark a cite, you will also indicate the number of authorities identified in the cite (1, 2, 4, or 5 or more). When several authorities are identified in a cite, it is referred to as a "string cite," though you won't annotate it as such. See these examples (citations highlighted and number of sources identified in [brackets] at the end of the sample):

> In keeping with Plaintiff's contractual obligations, Vendor grants access to Plaintiff's content only to third parties that either subscribe to the Plaintiff Database or have obtained express written permission from a customer. (*See* Mem. Op. & Order 5; *cf. also* Countercl. ¶ 15 (alleging that "[i]nformation contained within databases is shared with other members"); *id.* ¶ 18 (alleging that vendors have the technical capability to grant access to data to third parties).) **[2 authorities]**

> Accepting this allegation as true for purposes of this motion, Plaintiff still does not overcome *Noerr-Pennington* immunity. "[I]t is clear that a defendant's invocation of adjudicative process to press legitimate claims is protected even though its purpose in doing so is to eliminate competition." *Razorback Ready Mix Concrete Co. v. Weaver*, 761 F.2d 484, 487 (8th Cir. 1985) (citing *Noerr*, 365 U.S. at 140; *Pennington*, 381 U.S. at 669); *see also MCI Commc'ns Corp. v. Am. Tel. & Tel. Co.*, 708 F.2d 1081, 1156 (7th Cir. 1983), *cert. denied*, 464 U.S. 891 (1983) ("Without a doubt, the intention to harm a competitor is *not* sufficient to make litigation or administrative proceedings a sham. That anticompetitive motive is the very matter protected under *Noerr-Pennington*."). [**2 authorities**, but note how the cited authorities also refer to authorities. So, the "cert. denied" is part of the citation to the *MCI* case because it describes the history of that case; similar explanatory marks include *aff'd, rev'd, overruled by,* etc.]

If the author of the memo refers to the name of an authority (or author) in an actual sentence of prose, that name is not a cite,

but citation information following it within a sentence is a citation clause. In other words, if the name has a grammatical role in a sentence, it is not a citation. For example:

> This case differs from *Bjerke.* In *Bjerke,* the defendant provided a home away from the plaintiff's family, adopted many rules, and had extensive authority over the plaintiff's welfare. 742 N.W.2d at 665.

> In *Northwest Wholesale Stationers, Inc. v. Pacific Stationery & Printing Co.*, 472 U.S. 290 (1992), the Supreme Court described the essential attributes of a *per se* illegal boycott, including (1) joint efforts by a firm or firms to disadvantage competitors, and (2) the conspirators' possession of "market power or exclusive access to an element essential to effective competition." *See id.* 294-96.

In this example, the references to "*Bjerke*" and "*Northwest Wholesale...*" are not Cites, but the citation sentences at the ends of the sentences are, as is the clause after the first instance of the name of the *Northwest Wholesale* case.

BlockQuote: When a memo includes a long quote from another authority, the author sets it off from the surrounding text by indenting it on the left and perhaps by single-spacing it. (Most of the memo text will be double-spaced.) Do not annotate citations in block quotes.

Footnote: The author may have place a footnote reference in the text of her memo and display a footnote at the bottom of the page. The footnote reference number in the text and the reference number and footnote text at the bottom of the page should both be marked as "Footnote." Any citations in a footnote should be annotated as citations.

GATE annotation guide

Researcher: Brian N. Larson (BNL)
Revised July 14, 2013

Overview

The purpose of this coding project is to create annotated versions of memos written by law students and professional attorneys; the annotations identify parts of the memos like large sections, text headings, and citations to legal authorities. This annotation process will help to create a *corpus* (pl. *corpora*) of texts that the researcher will use for various projects.

The process will consist of two steps:
1. You will read and mark the memos in paper form.
2. You will record the annotations on a computer using software called "GATE: General Architecture for Text Engineering."
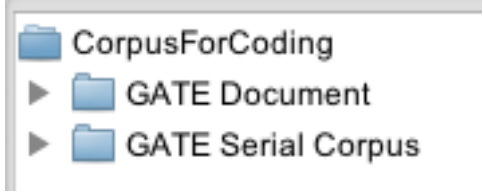
This document describes the second step, putting the annotations on the document in GATE. Follow these instructions whenever you are annotating documents for BNL using GATE. ***These instructions are designed with the assumption that you'll have at least half an hour to work or so; that's because the instructions for initiating an closing a coding session are a little cumbersome, and if you do many short coding sessions, you'll end up spending too much of your time starting up and shutting down.***

Beginning a session

In each of your work sessions, follow this process:
1. Make sure that BNL is not coding at the same time. Do this by checking your email and seeing whether he as sent you an email saying he is coding or an email indicating that he has finished.
2. Send an email to BNL indicating that you are beginning to code (so he knows not to).
3. Make sure that your Dropbox sync is completed (the files you work on are local on your computer, but they must be sync'ed to Dropbox to ensure you have the most recent copy). Look for green check at top of screen and green checks on the relevant folders:
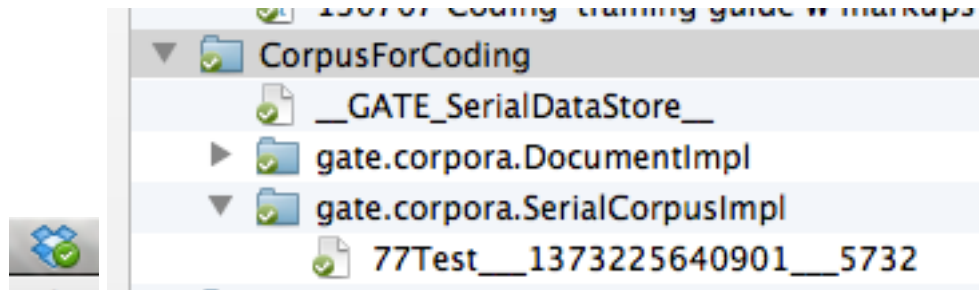
4. Open GATE.
5. On the "Messages" screen at startup, be sure that the plugins "ANNIE" and "Scheme_Annotation_Editor" are loaded.
6. Load the five annotation schemas. For each, go to "Language Resources," right click, choose "New," and then "Annotation Schema." The schemas are in the folder titled "GATE materials." Load each of the following files with the names given.
   a. LargeSegmentSchema.xml, "LargeSegment"
   b. HeadingSchema.xml, "Heading"
   c. FootnoteSchema.xml, "Footnote"
   d. CitationSchema.xml, "Citation"
   e. BlockQuoteSchema.xml, "BlockQuote"
7. Right click on "Datastores" and choose "Open Datastore." Follow these instructions:
   a. Choose "SerialDataStore…"
   b. Navigate to the Gender-Genre Team Annotation folder, click on the "CorpusForCoding" folder once, and choose "Choose."
   c. Click on the little triangle to the right of "Datastores" ▶ if necessary so that it points downward. (This is called "expanding" Datastores.) You should see the little file cabinet labeled "CorpusForCoding": . Double click on it, which should cause the main panel to display this:

   d. Expand "GATE Serial Corpus" so that it displays "77Test." Double-click on that.
   e. "77Test" should now show up in the Language Resources.
   f. Double-click on the "77Test" appearing under in the Language Resources. That should reveal a list of documents in the main panel, numbered according to the file numbers, with a code added. Thus, a file called "1001.docx" might appear as "1001.docx_0008."
   g. You are ready to work on coding a document.

GATE Annotation Guide: Analysis of law student writing assignments—Page 2
© 2013 Brian N. Larson

Coding a document

1. Click or double click on "77Test" under Language Resources so that the list of documents in the corpus appears on the main panel, something like this:
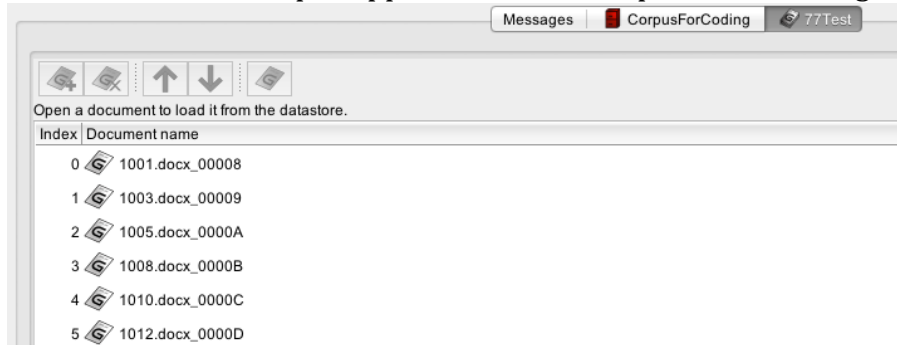

2. Double click on the document you want to edit. That will open it in GATE.
3. In the upper left-hand of the main panel, click on "Annotation Sets," which will display the annotation set list on the right, like this.
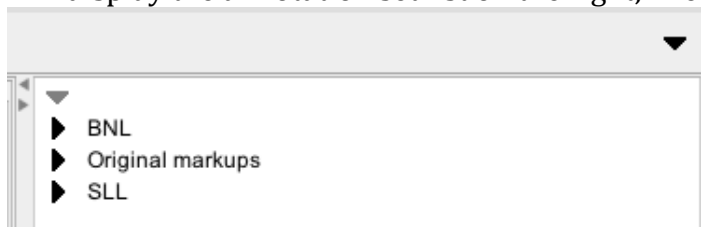
   In this example, there are three markup sets. You may find that there is only one ("Original markups"). If there is not an SLL markup set, create one by typing "SLL" at the bottom of this pane and hitting "New."

4. Using the "down arrow" in the upper right of the main display, select "Read only" from the menu, here:
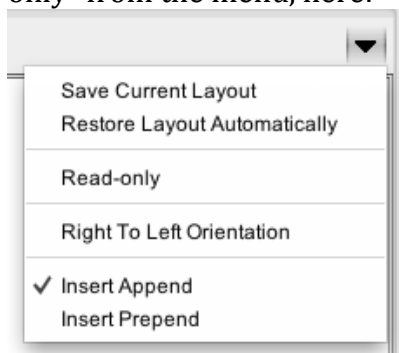
   This prevents you editing the underlying text, but permits you annotate it as required here.

5. It's probably easiest to annotate all segments of a given kind (LargeSegment, Heading, Cite, Footnote, etc.) at once, since GATE assumes that you want a new annotation to be the same as the last you gave.
6. Before adding any annotation, be sure that the SLL annotation set is selected on the right (that prevents you entering annotations as if they are mine).

7. Select the first segment you want to annotate, and then hover the mouse over it. You should see something like this:
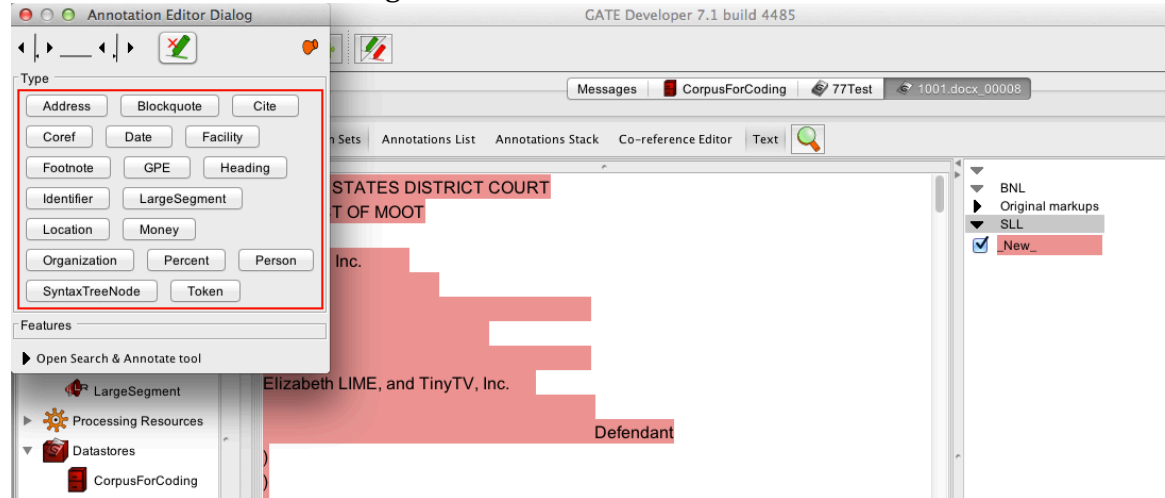

You must chose an annotation type from the Editor dialog. Some of these are defaults of GATE; don't use them. Use only the ones defined in the coding guide. I would start with the LargeSegment type to get it out of the way.

8. When you select an annotation type, you will generally have to fill out one or more "Features." Here are the possible features for LargeSegment:

The red box around the feature means that it is required that you choose one.

9. Once there are no red boxes remaining in the Annotation Editor Dialog, close it by clicking the red circle in the upper left of it.

10. WARNING: Sometimes the Annotation Editor Dialog (AED) pops up before you make a new selection or finish making your selection. That's because whenever you "mouse over" an existing annotation that is visible on screen (i.e., it is colored), the AED opens so you can "inspect" the annotation. If that happens, be sure to close the AED and make sure you are hovering over where your new annotation will be when the AED pops up. (It's easy to accidentally replace an annotation you've already made.)

11. When you've added annotations of a particular type, that type shows up with a color code and check box under your annotation set:
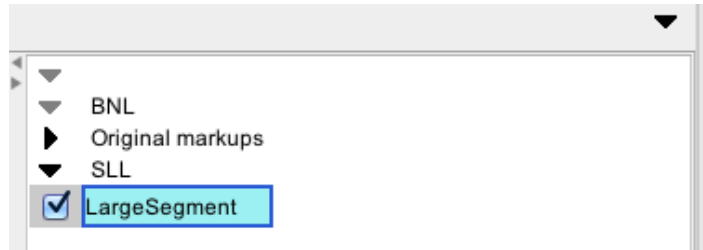
Hide annotations you are not currently using by unchecking their boxes. TIP: If you finish adding your LargeSegments, then click on this box, and then try to add your first Heading, GATE defaults the new annotation to LargeSegment, which causes the LargeSegment check to reappear in the box and makes the LargeSegment codes visible. You can just click on it again, and the next time, GATE will default to Heading, or whatever was the next code you added.

12. Add further annotations.

13. TIP: If you have trouble finding something you need to annotate, click the magnifying glass and search for a word in or near the span of text you are looking for. (Unusual words work better, of course!)

14. When annotating headings, note that GATE will not import the numbers on the automatically numbered heading paragraphs from Word. What looks like this in Word…

> **6. *Federal Case Law <u>Indicates</u> that the Issue of Writings Precedent is Strongly Contested, Suggesting Ambiguity.***

…looks like this in GATE…

> favored.
> Federal Case Law Indicates that the Issue of Writings Precedent is Strongly Contested, Suggesting Ambiguity.
>   In relatively recent history, there has been an important circuit split on the

15. When annotating a footnote, be sure to grab the square brackets that open and close it:

> heeding above all others in issues of ambiguity, it is surely Barbara Ringer.[1: United States Copyright Office. "Barbara Ringer, 1925-2009". Copyright Notices, April 2009. http://www.copyright.gov/history/bios/barabara-ringer-special-edition-2009-04.pdf]

16. When doing citation annotations, follow these conventions for covering the surrounding punctuation:

   a. For well-formed sentence citations, start your span with the first number, letter or parenthesis of the citation and end your span with the space before the next sentence begins:

   > . (Compl. ¶ 9). Tl

   b. For well-formed clause citations, start your span with the comma that begins the citation. Include the comma that ends it only if that comma

is not necessary for the sentence to be properly punctuated absent the citation. Do not include a sentence-ending period in your citation span. Examples:

> Indeed, deference to the urgings of Congressional agencies has been a long-standing tradition recently reaffirmed by the Supreme Court in U.S. v. Mead Corp ., 533 U.S. 218 (2001). Quoting from the historic Skidmore v. Swift & Co ., 323 U.S. 134 (1944), the court iterated that "an agency's interpretation may merit some deference whatever its form, given the "specialized experience and broader investigations and information" available to the agency, id . at 139, and given the value of uniformity in its administrative and judicial understandings of what a national law requires, id., at 140 . "

c. Generally, with citations, whether well formed or ill-formed, sentence or clause, try to select the citation span so that what would be left if the citation were deleted would be a grammatical sentence, properly punctuated. Examples:
    <<still coming>>

17. When you are done doing your annotations on this document, you should find that you have a set of annotation types and their colors displaying in the annotation sets pane under SLL, like this:

> ▼
> ▼ BNL
> ▶ Original markups
> ▼ SLL
> ☑ Cite
> ☑ Footnote
> ☑ Heading
> ☐ LargeSegment
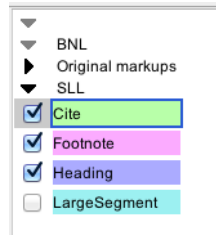
Unless there were annotations under "BNL" before you started annotating this document, there should be none now. All your annotations should appear under the SLL set. Unfortunately, if they appear elsewhere, the only way to get rid of them is to right click on the offending entry and delete it (along with all annotations associated with it). Say you accidentally put one of your citation annotations under Original markups. Your annotation sets pane might look like this:

> ▼
> ▼ BNL
> ▼ Original markups
> ☑ Cite
> ☐ a
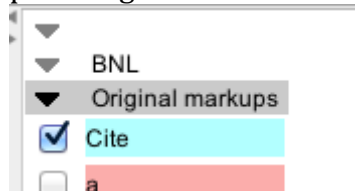
Right-click on the "Cite" label here and choose Delete. You'll need to go back through the document and reapply any citation annotations you have just deleted.

18. Take a few moments to browse back over your annotations to see that they look correct.

19. TIP: You can see whether you have missed italicized "Id." anywhere in the document by showing the original markup for Italics and underlining.

Expand "Original markups" in the Annotation sets panel, then click on "i" and "u" if they are visible, like this:

Then when you look through the document, you can see that your annotations of "Id." citations will generally overlap an "i" or "u" annotation, like this:


20. **REALLY IMPORTANT BIT:** When you have finished with this document, go to its name in the left pane in Language Resources, right-click and choose "Save to its Datastore." Nothing is saved properly until you do this!

21. After saving, go to the document's name in the left pane in Language Resources, right-click and choose "Close."

22. On the paper copy of the document, note the date and time you finished putting annotations on the computer. (We made need this for disaster recovery. If we lose a day's work, we need to know which documents we worked on that day.) Retain the paper copy—give it to BNL at your convenience.

23. Repeat this process with the next document.

For an example of a document where BNL has already layered annotations into it in GATE, open paper 1001, 1003, 1005, 1008, or 1010 and view the BNL annotation set.

Ending a session

1. Save the last document you worked on and close it.

2. In the left pane in Language Resources, right-click "77Test" and choose "Save to its Datastore."
3. Then right-click "77Test" and choose "Close."
4. Right-click "CorpusForCoding" under Datastores and close it, too.
5. Exit GATE; make sure the program has completely quit (it no longer appears in the list when you Command-TAB).
6. Wait for your Dropbox sync to be completed; green check at top of screen and green checks on the relevant folders:
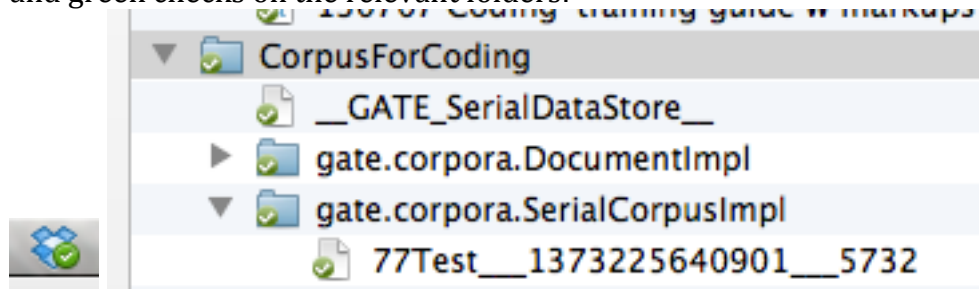
7. Send an email to BNL indicating that you are done coding (so he knows that he can work on the corpus).

References

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text*, *23*(3), 321–346.

Ariely, D. (2012, June). Essay mills – a coarse lesson on cheating. *Los Angeles Times.* Retrieved October 9, 2014, from http://articles.latimes.com/2012/jun/17/opinion/la-oe-ariely-cheating-20120617

Board, I. R. (2004, October). Protecting human subjects guide. University of Minnesota. Retrieved from http://www.research.umn.edu/irb/download/Protecting%20Human%20Subjects%20Guide.pdf

Breuch, L.-A. K., Olson, A. M., & Frantz, A. (2002). Considering ethical issues in technical communication research. In L. J. Gurak & M. M. Lay (Eds.), *Research in technical communication* (pp. 1–22). Westport, CT: Praeger Publishers.

Caulley, D. N. & Dowdy, I. (1986). Legal education as a model for the education of evaluators. *Educational Evaluation and Policy Analysis*, *8*(1), 63–75.

Cauthen, R. (2010). *Black letters: An ethnography of a beginning legal writing course.* Cresskill, NJ: Hampton Press.

Clary, B. G. & Lysaght, P. (2010). *Successful legal analysis and writing: The fundamentals* (3rd ed.). St. Paul, MN: Thomson Reuters.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., . . . Peters, W. (2012, December). Developing language processing components with GATE version 7 (a user guide). Retrieved January 1, 2013, from http://gate.ac.uk/sale/tao/split.html

DeFrancisco, V. P., Palczewski, C. H., & McGeough, D. D. (2014). *Gender in communication: A critical introduction* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Eaton, A., Brewer, P. E., Portewig, T. C., & Davidson, C. R. (2008). Examining editing in the workplace from the author's point of view. *Technical Communication*, *55*(2), 111–139.

Hansen, S. (2004, August). Dear plagiarists: You get what you pay for. *The New York Times.* Retrieved October 9, 2014, from http://www.nytimes.com/2004/08/22/books/review/22HANSENL.html

Janssen, A. & Murachver, T. (2004). The relationship between gender and topic in gender-preferential language use. *Written Communication*, *21*(4), 344–367.

Kahlenberg, R. D. (1999). *Broken contract: A memoir of Harvard Law School.* Amherst: University of Massachusetts Press.

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, *17*(4), 401–412.

Larson, B. (2015, May). *Gender/Genre: Gender difference in disciplinary communication* (Ph.D. dissertation, University of Minnesota, Minneapolis).

Mertz, E. (2007). *The language of law school: Learning to "think like a lawyer".* New York: Oxford University Press.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*(2), 151–167.

Murphy, D. J. (2002). Surveys and questionnaires. In L. J. Gurak & M. M. Lay (Eds.), *Research in technical communication* (pp. 93–110). Westport, CT: Praeger Publishers.

Osborn Jr., J. J. (2004, September). *The paper chase* (Special Anniversary Ed). Albany, N.Y: Whitston Publishing.

Pakhomov, S. V., Chacon, D., Wicklund, M., & Gundel, J. (2011). Computerized assessment of syntactic complexity in Alzheimer's disease: A case study of Iris Murdoch's writing. *Behavior Research Methods*, *43*(1), 136–144.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010, October). Classifying latent user attributes in Twitter. In *Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents* (pp. 37–44). Toronto, ON, Canada: ACM.

Schmedemann, D. A. & Kunz, C. L. (2007). *Synthesis: Legal reading, reasoning, and writing* (3rd ed.). New York: Aspen Publishers.

Spinuzzi, C. (1996). Pseudotransactionality, activity theory, and professional writing instruction. *Technical Communication Quarterly*, *5*(3), 295–308.

The Bluebook: A uniform system of citation (19th ed.). (2011). Cambridge, MA: The Harvard Law Review Association.

Turow, S. (2010). *One L*. New York: Farrar, Straus and Giroux.