Gender/Genre: The lack of gendered register in texts requiring genre knowledge

Brian N. Larson

Georgia Institute of Technology

DRAFT compiled August 26, 2016 at 15:45:42.

Appearing in 33 *Written Communication*, October 2016.

Author Note

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 discourse, and argumentation theory.

Abstract

Some studies have found characteristics of written texts that vary with author gender, echoing popular beliefs about essential gender differences that are reinforced in popular works of some scholarly authors. This article reports a study examining texts (n = 193) written in the same genre—a legal memorandum—by women and men with similar training in production of this type of discourse—the first year of U.S. law school—and finds no difference between them on the involved–information dimension of linguistic register developed by Biber (1988). These findings provide quantitative data opposing essentialist narratives of gender difference in communication. This essay considers relevance theory (Sperber & Wilson, 1995) as a framework for understanding the interaction, exhibited in this and previous studies, of genre knowledge and gendered communicative performances.

Keywords: legal texts, legal discourse, corpus, automated text analysis, relevance theory, cognitive environment

Gender/Genre: The lack of gendered register in texts requiring genre knowledge

Some studies have found stylistic characteristics of written texts that vary with author gender (Argamon, Koppel, Fine, & Shimoni, 2003, 2007; Koppel, Argamon, & Shimoni, 2002; Rao, Yarowsky, Shreevats, & Gupta, 2010; X. Yan & Yan, 2006). The media commonly report on gender-difference studies (for examples, see Castillo (2013), Robb (2014), Zimmer (2012)). The results of those studies probably echoed *doxa*, or popular beliefs, about essential gender differences and possibly some popular works by scholarly authors such as Deborah Tannen (Tannen, 1991/2001) that suggested that men and women are from different communicative cultures, and each gender may be challenged to adapt to the culture of the other (Maltz & Borker, 1982), or even that men and women come from different communicative worlds (Gray, 1992). One problem of course is that these discourses of gender difference support "deficit models," in which women's alleged communication techniques make "women unsuitable candidates for positions of public authority and responsibility" (Cameron, 2003, p. 454).

One approach to exploring differences in texts is to consider linguistic register, defined here after Matthews (2007), as a "set of features of speech or writing characteristic of a particular type of linguistic activity or a particular group when engaging in it." In particular, this study considers features of the *involved–informational* dimension in register described in Biber (1988, 1995). Previous studies of gender-differences in text have applied the same framework and found interactions between writer's purposes and audiences and text genre on the one hand, and gender on the other hand (Argamon et al., 2003; Biber, Conrad, & Reppen, 1998; Palander-Collin, 1999). Small-scale studies in technical and business communication classrooms, on the other hand, have tended not to find gender differences, though they have not applied the involved–informational framework (Smeltzer & Werbel, 1986; Sterkel, 1988; Tebeaux, 1990).

It may seem a commonsense observation that writers attempting to perform a professional or disciplinary genre will abandon gendered writing habits (if they had them

to begin with). But this process has not previously been theorized. I propose using the relevance theory of Sperber and Wilson (1995) to explain gender and genre as instances of the same phenomenon, the search by humans for relevance in communication. I offer a study of gendered novice writers attempting to conform to a professional genre (n = 193) and explain its findings, and the findings of previous studies of gender difference, according to relevance theory.

This essay presents background in the relevant literature, describes the methods and findings of the study, and suggests that relevance theory may be a way of understanding gender and genre. It also describes and makes available tools, including the corpus and all code used in this study, to permit other scholars to replicate, interrogate, and challenge these findings.

Gender, Genre, and Relevance Theory

A Relevance-Theoretic Account of Gender and Genre

Sperber and Wilson (1995) proposed an account of human communication they called "relevance theory," after the principles of relevance that guide its operation.[1] They offered their theory as an overt contradiction of "code models" (p. 11-12) of communication and as an alternative to the cooperative principle and maxims of Grice (1989). With regard to the latter, Sperber and Wilson (1995) showed that the cooperative principle was neither necessary nor sufficient to explain communication. Their theory has been employed widely in experimental pragmatics (see studies cited in Yus (2015)), and their observations may offer a hope for bridging rhetoric and pragmatics, a task that Dascal and Gross (1999) and Liu and Zhu (2011) have attempted without fully satisfactory results. This essay explores relevance theory and recasts genre knowledge and gender in terms of it.

---

[1]For a fuller exposition of relevance theory, see book-length treatments in Sperber and Wilson (1995) and Wilson and Sperber (2012). For a chapter-length summary, see Wilson and Sperber (2006). For an extensive, topically organized bibliography, see Yus (2015).

According to Sperber and Wilson (1995), each actor has a *cognitive environment* consisting of *manifest assumptions*—each assumption a propositional statement of a fact that is represented in the actor's mind or warranted by her perceptions of her immediate physical environment and each of which she believes to be true. Straßheim (2010) proposed extending the cognitive environment to include the goals and emotions of the actor, a proposal adopted here. When two actors are engaging in communication, components of their cognitive environments that are manifest to both of them are *mutually manifest*. It is perhaps easiest, though not precisely correct on Wilson and Sperber's account, to conceive of an actor's mutually manifest cognitive environment as those cognitive-environment components of which she believes she and her interlocutor share knowledge. In this sense, the mutually manifest cognitive environment for each actor is both a social construction—based on her understanding of her interlocutor—and an individual one—subject to the fact that she does not have experiences and capacities that are entirely congruent with her interlocutor's.

Relevance is a measure of the cognitive efficiency of a communicative performance.[2] The greater the cognitive effect an actor seeks, the more relevant a communicative performance or interpretation of it. The greater the cognitive effort required to produce or interpret a communicative performance, the less relevant it is. The effort to maximize relevance is a feature of human cognition, in that our minds are evolved not to expend cognitive effort without commensurate effects, what Sperber and Wilson (1995, p. 261) called the "First Principle of Relevance." A communicator composes her communicative performance to be *optimally relevant* to her addressee, what Sperber and Wilson (1995, p. 266) called the "Second Principle of Relevance." In other words, it will be "worth the addresse's effort to process it," and it "is the most relevant one compatible with the communicator's abilities and preferences" (p. 270). Straßheim (2010) offered the insight

---

[2]Sperber and Wilson (1995) referred throughout their work to *ostensive stimuli*. I have chosen "communicative performance" for reasons that become clear below.

that relevance is related to Schütz's concept of typification, which has played a role in genre theory (Miller, 1984) and writing theory (Bazerman, 2013). Thus, if a communicator believes that her audience expects a typified response to a typified communicative situation—that is, if these facts are mutually manifest in her cognitive environment—she will conclude that the typified response will be optimally relevant to her audience. In this sense, an actor's cognitive environment is influenced and shaped by her observations of others' *repeated* actions and by her own.

Relevance can explain some features of genre. Berkenkotter and Huckin (1995) defined an individual's "genre knowledge" as her "repertoire of situationally appropriate responses to recurrent situations" (p. ix). They proposed a framework for genre knowledge with five elements: *Dynamism* considers the "contention [between] stability and change," that genres are instantiated differently in response to actors' knowledge and objectives (p. 6). *Situatedness* accounts for the fact that "genre knowledge is a form of situated cognition, inextricable from professional writers' procedural and social knowledge" (p. 13). Genre knowledge embraces both *form and content*, an acknowledgment that genre implicates both structure and epistemological commitments of the communities that use it. *Duality of structure* acknowledges that we simultaneously *constitute* and *reproduce* genres. *Community ownership* considers the fact that genres belong to professional and disciplinary communities: "actors are typically well aware of textual patterns and epistemological norms of their discourse community" (p. 25).

The focus by Berkenkotter and Huckin (1995) on *genre knowledge* avoids the dangers of reifying genre: Most genres are not fixed things subject to categorical descriptions, just as the meaning of most words is not fixed and subject to categorical description. In relevance-theoretic terms, the communicator's cognitive environment includes assumptions about the audience's expectations, based on the communicator's conception of the discourse community and its typified situations. Varying from the audience's expectations will cause the audience to engage in greater cognitive effort, which is warranted only if the

audience believes some cognitive effect—greater than the typical effect—will arise from interpreting the performance. Communicators naturally know this.

Relevance theory explains consequences of some circumstances where novice and experienced practitioners encounter genre: First, despite the central tendency among the participants in a discourse community or communicative exchange about appropriate formal conventions for that exchange, each participant has a unique, possibly idiosyncratic cognitive environment, and novice practitioners may not be aware of all the central-tendency elements or even who counts as being inside and who outside the discourse community. For the novice communicator, observing, identifying, and conforming to genre conventions is a highly relevant activity. A novice communicator's cognitive environment, and thus her genre knowledge, grows out of her assumptions about a recurring, typified situation and the expectations of other persons around her for the appropriate or fitting response to that situation. For an apprentice practitioner in a new discourse community, the intensity and accessibility of the goal of succeeding in it—the desired cognitive effect—likely makes the comparatively great cognitive effort required to conform to the genre worth it—in other words, genre conformity is highly relevant for the novice communicator.

Second, despite the fact that genres give typical results in typified situations, experienced communicators can and do depart from genre conventions all the time in order to achieve particular goals. For the experienced communicator, breaching a genre convention may be necessary to achieve some goal of the communicator that is not fulfilled by the genre as it exists. Consider a company employee who prepares a periodic report that has generic status within the company. Perhaps another communicative approach would cost her less effort, better highlight her own contributions, allow her to take credit for saving the company money, gratify her desire for efficiency, or advance some other goal. But any effort she makes to vary from the existing genre convention is likely to be opposed by other individuals to the extent that they individually believe that the increased

cognitive effort of understanding the new report exceeds the benefits attendant on it. Our hero can predict this resistance (thanks to the operations described by relevance theory) and revise her strategy to address the cognitive effort/effect balance. She can, for example, attempt to pitch the revised approach in such a way that it appears to offer greater cognitive effects for the audience, warranting their cognitive effort to adjust the expectations arising from their own genre knowledge. Relevance theory can thus explain how the genre dynamism of Berkenkotter and Huckin works

This essay views genre knowledge in a way consistent with relevance theory and with Berkenkotter and Huckin (1995). Genre knowledge comprises components of the communicator's cognitive environment: her assumptions about communicative behaviors she expects to have a particular effect or effects on a reader based on knowledge about a typified situation in the writer's cognitive environment.

Gender performances, too, are subject to the principles of relevance. Butler (1993) characterized gender as a form of performativity arising in "an unexamined framework of normative heterosexuality" (p. 97). Performativity is not merely *performance*:

> performance as a bounded "act" is distinguished from performativity insofar as the latter consists in a reiteration of norms which precede, constrain, and exceed the performer and in that sense cannot be taken as the fabrication of the performer's "will" or "choice" (p. 234)

The quoted text emphasizes half of a duality in Butler's treatment of gender: The communicative actor is not necessarily free or even possessed of agency, and yet the sex/gender dyad is "both produced and destabilized in the course of its reiteration" (p. 10).

> This "identification" is thus *repeatedly* produced, and in the demand that the identification be *reiterated* persists the possibility, the threat, that it will *fail* to repeat (p. 102; emphasis in original).

Relevance theory and genre theory as recast through relevance theory do not address the

many philosophical commitments that Butler makes in her work. They do, nevertheless, provide a means for understanding the operation of gender as Butler describes it. Gender knowledge shares many characteristics with genre knowledge: It takes place at a tense junction of stability and change (dynamism), gendered performances both constitute and reproduce gender (performativity/structuration), and gendered persons are typically aware, through observation and habit, of the gender patterns and norms of their societies (community ownership/citationality).

Relevance theory explains the operations of gender dynamism, performativity, and citationality. In an environment where people are slotted into a heteronormative gender binary in a culture that has differing social expectations of males and females, men and women will likely develop different habits of communicative performance because they have cognitive environments that differ in certain ways: They have knowledge of different typified situations and are trained to have different social goals for communication. Thus, the mutually manifest cognitive environment of males probably differs from that of females, and expectations of an audience when listening to a male are different than those when listening to a female. Conforming to these expectations for one's own gender is easier—and therefore more relevant—for the communicator than non-conforming, and non-conforming imposes a comprehension cost on the audience that may make the communicator's performance seem less relevant.

Under this view, gender knowledge can be seen as a generalization of the definition of genre knowledge: Gender knowledge comprises components of the communicator's cognitive environment: her assumptions about behaviors she expects to have a particular effect or effects on another based on knowledge about a typified situation in the actor's cognitive environment. This study begins to test this generalization.

Exploring Gender and Genre through Linguistic Register

This study focuses on a subset of communicative behaviors subject to relevance theory and an actor's genre and gender knowledge: the conventions of linguistic register. In particular, this study considers features of the *involved–informational* dimension in text register described in Biber (1988, 1995). (See the detailed discussion of these features in *Method*.) According to Biber (1988), the informational features "are used to integrate high amounts of information into a text," and "are associated with communicative situations that require a high informational focus" (p. 105). By contrast, the involved features are "associated in one way or another with an involved, non-informational focus, due to a primarily interactive or affective purpose" (p. 105). Biber (1988) explored the register dimensions in part to analyze differences among *text types*, including "official documents," "academic prose," "personal letters," and "professional letters," and found that they tended to exhibit differences along the involved–informational dimension.

Several studies have also suggested that men and women communicate differently along the involved–informational dimension *without regard* to text genre. Biber et al. (1998) examined a 123,000-word corpus of letters written by men (n = 187) and women (n = 51) between the seventeenth and twentieth centuries. Using the involved–informational dimension, they concluded that letters women wrote showed higher prevalence of involved characteristics than those men wrote and that the difference was especially pronounced in the 20th century. They also concluded that letters women wrote to other women were more involved than those they wrote to men, and letters men wrote to men were more informational than those they wrote to women (p. 219-20).

Palander-Collin (1999) examined a corpus of letters—792,000 words written by men (n = 136) and women (n = 77) in English in the 17th century. She studied the frequency of the expression "I think" and other features associated with the involved–informational dimension. She concluded that women used the involved features "I think," first-person pronouns, second-person pronouns, private verbs (which express intellectual states and acts

that "are 'private' in the sense that they are not observable" (Quirk, Greenbaum, Leech, &
Svartvik, 1985, p. 1181)), and *that*-deletion (see explanation below) significantly more
frequently than men. Palander-Collin also tallied possibility modals (instances of the verbs
can, could, may, and *might*), an involved feature according to Biber, but concluded there
were no statistically significant differences.

One study of contemporary prose began with a collection of texts (n = 604) from the
British National Corpus (BNC), including newspaper and magazine essays, novels, and
other works of fiction and non-fiction, in which the researchers assessed some
characteristics of linguistic register—1,143 features, including the relative frequency of
function words, parts of speech, and part-of-speech bigrams and trigrams (Argamon et al.,
2003). In this corpus of 25,000,000 words, Argamon et al. (2003) showed that there were
statistically significant differences between male and female writers both in the fiction and
non-fiction "genres." They found that females used the involved features first-person
pronouns, second-person pronouns, and contractions significantly more often than males in
both fiction and non-fiction, and present tense verbs significantly more often in non-fiction.
Males used the informational features attributive adjectives and prepositions significantly
more often and had significantly higher mean word lengths in non-fiction texts. In fiction
texts, men used significantly more nouns and prepositions and had significantly higher
mean word lengths and type-token ratios. The other features did not vary significantly
with gender in fiction or non-fiction.

Perhaps from the theoretical perspective described above, these studies cannot be
taken to have explored texts in the same *genre* at all, because they consider genre "pitched
at such a broad level of generality that [it] can describe only superficial parameters of form
or content" (Berkenkotter & Huckin, 1995, p. 14). Authors write letters, official documents,
and academic prose and fiction and non-fiction for a wide variety of purposes, and those
text types may reflect a wide variety of expectations on the parts of writers and readers,
confirming it is difficult to refer to them as a single genre. Nevertheless, the fact that

researchers recognize these texts as types—and "gender" and "genre" share a common etymology in "genus," which means *kind* or *type*—suggests communicators have assumptions about them that may not be as definite as assumptions they hold about texts that are generic in the genre-theoretic sense but are still definite enough to guide text production in a relevance-theoretic sense.

By contrast, some earlier studies that considered written professional or technical communication, did *not* find differences that correlated with gender, though they did not necessarily look for features on the involved–informational dimension. For example, Sterkel (1988) performed statistical analyses of 20 stylistic characteristics of students (n = 108) in an undergraduate business writing class. Sterkel found no statistically significant differences between the male and female students' writing. Smeltzer and Werbel (1986) also found no significant differences between the writing of male and female students in a business class (n = 79, 16 variables), and Tebeaux (1990) presented the results of a qualitative examination of the quality of student responses to a case-study problem (n = 99), concluding that work experience was a more useful predictor of work quality than gender. We may speculate that the students in each of these studies probably had many similar components in their cognitive environments with regard to this writing: they likely had similar goals with regard to grades, similar assumptions about appropriate communicative performance and instructor expectations due to the training previously received in the class, and similar assumptions arising from having the same assignment prompt. These studies may thus have considered texts of the same genre, but they did not explore the features of the involved–informational dimension.

These studies and the theoretical discussion above suggest two research questions: When female and male writers attempt to write in a disciplinary genre in which they are being trained, do they use involved–informational features with relative frequencies that vary in relation to their genders? Further, can relevance theory explain the findings?

Should Scholars Research Gender Differences?

Before we consider the method for this study, we should address the argument that some scholars of gender and communication have made that variationist studies of language using the gender binary as a variable are inadvisable or unproductive (DeFrancisco, Palczewski, & McGeough, 2014; Eckert & McConnell-Ginet, 1992/2011) or that they perpetuate the gender binary even as they contest it (Cameron, 2003, p. 465). For example, DeFrancisco et al. (2014) warned against the risks of doing studies that focus on "sex difference only" in communication research, contending that researchers asking the sex-difference question are "likely asking the wrong question and will only reinforce essentialist views of gender/sex in communication" (p. 59), that researchers often conflate *sex* and *gender* as variables, that studies looking for statistical differences between communicative practices of gendered persons may overlook the very substantial similarities in their communications, and finally that studies focused on sex-difference are missing the possibility of an "intersectional analysis," which would "explore whether interdependent ingredients serve to influence the gender performance in unique ways yielding unique privileges [and] inequalities" (p. 60).

Their critique is thoughtful, but it overlooks at least four important counterarguments. First, to contend that sex-difference (or gender-difference) studies will "reinforce essentialist views" is to assume that such studies will find patterns of difference. If a study shows that there are few or no differences between the communications of females and males when they are attempting to conform to a professional genre, it can help to counter essentialist views in a society that appears to embrace them as a matter of *doxa* (whatever scholars in the academy may believe). Second, researchers in communication are almost never studying *sex differences* but rather *gender differences.* Communication researchers are unlikely to engage in the kind of physical examination of study participants that would allow them to make an ascription of a sex category.

Third, the absence of well designed studies that look for gender differences has two

possible consequences: If the studies would have shown popular beliefs about men's and women's communicative practices are wrong, at least in professional contexts, they would have proved useful for the very project for which DeFrancisco et al. (2014) advocated. If, on the other hand, such studies would discover differences, an intersectional theory of gender and communication should either account for them based on other "interdependent ingredients" or acknowledge that there are deeper differences between the genders (or sexes) than the theory previously admitted. Finally, if a study can show that participants need to have very little common training in genre knowledge before their writing exhibits none of the patterns of difference seen in previous studies, researchers and educators are in a position to help students understand how they can alter their communicative performances (if they seek to do so) to achieve their goals. This is one way for teachers of writing to help students address the privileges and inequalities about which DeFrancisco et al. (2014) were concerned.

Method

Participants and genre

One hundred ninety-seven students at two law schools in the midwestern United States took part in a study, filling out a brief demographic survey and submitting as a writing sample the legal memorandum they had prepared for their capstone first-year legal writing assignment. The full text of the questionnaire is available in the online appendix to this article at **WEB ADDRESS TBD**. I also conducted email interviews with several of the individual instructors.[3] According to the administrations and records at these two schools, they enrolled a total of 545 new students in the applicable year; of them, 263 were female and 282 male. Each school required as part of its first-year curriculum several basic

---

[3]Each participant gave informed consent, and the students agreed to license the writing sample for distribution as described later in this essay in return for a $15 gift card. The University of Minnesota Institutional Review Board determined this study was exempt: Study number 1202E10685.

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.

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 one school, all students wrote memoranda supporting or opposing a motion to dismiss a single hypothetical copyright claim. At the other school, the final problem varied from class section to section: 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.

Of the 197 students who participated, 193 identified their genders (see discussion below). Of participants, 180 (91.4%) learned English as their first language, nine (4.6%) learned English in the U.S. as a second language, five (2.5%) learned English outside the U.S. as a second language, and three participants (1.5%) did not respond. A great majority of them were under age 34: 109 (55.3%) of them aged 18-24; 78 (39.6%) aged 25-33; and the remaining 10 (5.1%) older than 33.

Ascription of gender category

Participants were asked 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. Though the researcher anticipated that student responses (which are themselves gender performances) would conform to the heteronormative gender binary, this approach prevented the researcher imposing it on them, and it prevented the researcher insulting transgender persons by proposing that they should chose "Other" or

"None of the above" in a pick-list. Of the 197 students who participated in this study, 193 responded to this question.

As anticipated, the responses fell along a gender binary: Two percent of respondents did not answer this question and are excluded from the analyses below. The responses of 52.8% of participants suggested they performed a "female" or perhaps "feminine" gender. Their responses were *F, Fem, Female*, and *female*, and they are referred to in this study as "Gender F." The responses of the remaining 45.2% of participants suggested they performed a "male" or "masculine" gender. Their responses were *Cis Male, M, Male*, and *Masculine*,[4] and they are referred to in this study as "Gender M."

Data preparation and analysis

The participants' writing samples were prepared to allow a comparison of involved–informational features of linguistic register. The texts were manually annotated to identify segments that would not be included in the statistical analyses below, including "block quotations" (text not written by the participants), legal citations (which tended to pose problems for automated processing), headings, and formulaic text components, like case caption and signature block, which were nearly identical from sample to sample. The corpus consisted of 650,000 words; the texts exhibited a mean length of 3366 words (stDev = 502). The corpus of texts is (or shortly will be) available in XML format and the original word processing format from the Linguistic Data Consortium (Larson, In preparation).

Python code including the Natural Language Toolkit (Bird, Klein, & Loper, 2009) was written to calculate the following features. With the exceptions noted, the features are those identified by Biber (1988) as components of the involved–informational dimension and were implemented with code constructed according to the rationales provided by Reymann (2002). Biber's model for the involved–informational dimension included 23

---

[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 et al., 2014, p. 60).

features co-occurring in involved texts and five co-occurring in informational texts (Biber, 1995, p. 142).

Involved components.   According to Biber (1988), the involved features are "associated in one way or another with an involved, non-informational focus, due to a primarily interactive or affective purpose" (p. 105). Except for the *adverbs* feature, I have excluded features that Biber classified as having "lower weights" and that he did not use in computing dimension scores. Also excluded is the *Non-phrasal coordination* category because of the difficulty of representing it in an algorithm. Except as otherwise indicated, all values are relative frequencies calculated on the basis of instances per 1,000 tokens.

Private verbs: According to Quirk et al. (1985, p. 1181), private verbs express intellectual states and acts that "are 'private' in the sense that they are not observable." They include, for example, forms of *anticipate*, *feel*, and *sense*. The code here counted them based on a list drawn from Quirk et al. (1985, p. 1181). E.g.: *Since it is not included, the plain language allows us to **realize** that there is no requirement for a written agreement to precede the work.* (Sentence 1037-29.)[5]

Subordinator-*that* deletion: The code counted instances of that-deletion, which according to Biber (1988, p. 244) is a form of "syntactic reduction" where *that* functioning as a subordinator may be omitted. E.g.: *Second, the Register indicates that she believes* **[that]** *the writing must be executed before the work is to be created.* (Sentence 1003-162.)

Contractions. The code counted instances of contractions, such as *I'm, we're, won't, can't, they'll*, etc. E.g.: *. . . authors will simply sign these printed form contracts and thereby assign away their rights without realizing that **they're** doing so.* (Sentence 1114-54.)

Present-tense verbs. The code counted instances of present-tense verbs. E.g.: *The phrase "contribution to a collective work" **implies** that a "work made for hire" **represents** a slice of a larger whole.*

---

[5]Sentence-number references are to the four-digit paper number and the number of the sentence based on the Python code's sentence tokenizer, which starts its count at 0.

Pronouns: The code separately tallied instances of the first-person pronouns *I, me, we, us, my, our, myself,* and *ourselves*, the second-person pronouns *you, your, yourself,* and *yourselves*, the pronoun *it*, the indefinite pronouns *anybody, anyone, anything, everybody, everyone, everything, nobody, none, nothing, somebody, someone,* and *something*, and the demonstrative pronouns *this, that, these,* and *those.*

Pro-verb or prof-form *do*: The code counted instances where a form of *do* stands in for "an entire clause, reducing the informational density of a text" (Biber, 1988, p. 226). E.g.: *Defendant had a duty to reasonably verify the record of an employee working directly with the public as Francis **did**.* (Sentence 2081-72.) In this example, *did* substitutes for *worked directly with the public.*

Analytic negation: The code counted instances of *not* and *n't* as instances of analytic negation. This excludes instances of synthetic negation, negation with *no* or *neither/nor* (Biber, 1988, p, 245). For an example of synthetic negation: *Here, Peterson's speech is **neither** being prominently brandished at a school-sanctioned event, **nor** published within a school publication.* (Sentence 2082-88.)

General emphatics: The code tallied instances of this category described by Biber (1988, p. 241) and including the following: *for sure, a lot, such a, real* (when followed by an adjective), *so* (when followed by an adjective), *just, really, most, more,* and forms of the verb *do* followed by another verb. E.g.: *These agreements. . . would save each party **a lot** of money. . . .* (Sentence 1082-121.) *Wager **did know** that Atticus was trying out for the second play and had made a similar statement after the first time he did not receive a role.* (Sentence 2048-88.)

Be as main verb: The code counted instances of the forms of the verb *be* when it was used as a main verb, that is, not as an auxiliary verb. According to the pattern defined by Biber (1988, p. 229), this would include any form of the verb *be* followed by a determiner, possessive pronoun, title of address, preposition, or adjective. The code here did not count instances of *be* in front of titles of address, as the part-of-speech tagger used here did not

identify them. E.g.: *Mr. Cosgriff **was unaware** of the nature of the "premium fee", and was surprised to learn that others could access his research and drafts through this service.* (Sentence 2049-27.) In this example, the second instance of *was* was not counted, as it is was auxiliary to the passive form *was surprised.*

Causative subordination with *because*: The code counted all instances of *because*, as it "is the only subordinator to function unambiguously as a causative adverbial" (Biber, 1988, p. 236).

Discourse particles: The code counted instances of the *well, now, anyway, anyhow,* and *anyways* at the beginning of sentences or clauses, as they "signal a transition in the evolving progress of the conversation. . . [or] an interactive relationship between speaker, hearer, and message" (Reymann, 2002, p. 43). *Now, Golfman is asking the court for a remedy of breach of contract to one that was fully performed.* (Sentence 2066-3.)

General hedges: A hedge "can mark a proposition as uncertain" (Biber, 1988, p. 240). Biber included instances of the following: *at about, something like, more or less, almost, maybe,* and instances of *sort of* and *kind of* where they are not functioning as true nouns. The code here did not count the instances of *sort of* and *kind of.*

Amplifiers: Amplifiers have the effect of "boosting the force of the verb" (Biber, 1988, p. 240). They are *absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly,* and *very.*

Sentence relatives: These forms "function as a type of comment clause, and they are not used for identificatory functions in the way that other relative clauses are" (Biber, 1988, p. 235). Biber defined this pattern as *which* following a comma. E.g.: *Plaintiff's injury**, which** caused permanent loss of sight in his left eye, resulted from an incident that took place inside Defendant's stadium**, which** is located in St. Paul, Minnesota.* (Sentence 2068-10.)

WH-questions: The code counted instances of *WH*-questions, which are defined as any *wh*-word at the beginning of a clause followed by an auxiliary verb. See Biber (1988,

p. 227) and Reymann (2002, p. 50-51) for fuller discussion. The *wh*-words include *how.*
E.g.: *This leads to a second question: if there is a clear mechanism by which an author can
transfer a copyright to a buyer in § 204(a),* **why would the work-for-hire provision
add a redundant method for an author to transfer ownership in § 101(2)?**
(Sentence 1096-39.)

Possibility modals: This feature consists of instances of the verbs *can, could, may,*
and *might.*

WH-clauses: Biber defined these clauses as those where a *wh*-word follows a public,
private, or suasive verb and is not succeeded by an auxiliary verb. The public, private, and
suasive verbs used here and in Biber are defined and listed in Quirk et al. (1985,
p. 1180-1183). E.g.: *The Court may find it instructive to* **consider how courts** *have
construed the writings requirement of the § 204(a) transfer provision.* (Sentence 1096-127.)

Final or 'stranded' prepositions: The code counted instances of prepositions followed
immediately by any punctuation (Biber, 1988, p. 244). *Even if there was no legislative
history or case precedent to rely* **on,** *well founded principles of contract law ought to. . . .*
(Sentence 1100-144.)

Adverbs: The code counted instances of adverbs.

***Informational* components.**   According to Biber (1988), the informational
features "are used to integrate high amounts of information into a text," and "are
associated with communicative situations that require a high informational focus" (p. 105).
This study included all of Biber's informational features.

Nouns: The code tallied all the nouns, common and proper, in the texts. E.g.: **Dave
Nelson ("Nelson") is the executive producer of Lawless Love,** *and an* **employee** *and
an* **agent** *of* **Plaintiff***.* (Sentence 1101-6.) The code counted nine nouns in this sentence.
Note that it treated each instance of a proper-noun sequence like *Dave Nelson* and *Lawless
Love* as two nouns.

Mean word length: This feature was *not* expressed as a frequency per 1,000 tokens.

Instead, it is the mean length of all tokens not tagged as punctuation in each paper.

Prepositions: The code tallied all instances of prepositions (less those counted as final prepositions, see above). Note that the part-of-speech tagger used here classed subordinating conjunctions with prepositions, and they were counted here. E.g.: ***During the meeting, Dave Nelson and Defendant Lime agreed orally that she would prepare a screenplay for Lawless Love by April 1, 2010.*** (Sentence 1101-9.)

Type-token ratio: Following the model of Biber (1988) and Reymann (2002), the code calculated the number of unique types (words or punctuation marks) in the first 400 tokens of each text, dividing it by 4 to generate a fractional percentage. A higher value indicates a larger vocabulary being used in the text and a higher information content for the text.

Attributive adjectives: The code counted all instances of attribute adjectives, those modifying a following noun. E.g.: *The **plain** language is the guide to **legislative** intent underlying a statute, and only if the **plain** meaning is regarded as ambiguous or unclear in indicating the legislative intent, then courts look beyond the statute itself.* (Sentence 1101-47.) Note that *ambiguous* and *unclear*, though tagged as adjectives here, do not immediately precede their noun (the preceding *meaning*) and are thus not counted as attributive adjectives.

All the Python code used for this study is available in a Github repository (Larson, 2016). Lists with all the *wh*-words and the private, public, and suasive verbs used in this study are available in the online appendix to this article at ***WEB ADDRESS TBD***. A table with the values calculated for each participant's paper is also available there.

Once the feature values where calculated for each text in the Gender F and Gender M categories, the values were tested for Gaussian distribution.[6] The samples had Gaussian distributions for only six of the 28 features. Consequently, the Mann-Whitney U score (which is equivalent to the Wilcoxon rank-sum test) was calculated for all features so that

---

[6]The Gaussian distribution is often called a "normal" distribution, but I worry about the implications of the term "normal" in other contexts and try to avoid it in this one.

comparable tests of significance and effect size could be run for all features (Utts & Heckard, 2006, p. 614). Statistical significance was assessed based on the U score, with α = 0.05, which is a common and conventional choice in the social sciences (Utts & Heckard, 2006, p. 448).[7] Effect size was calculated from U score and sample sizes as described by Fritz, Morris, and Richler (2012).

Findings

The year-end motion practice memos of 104 Gender F and 89 Gender M law students were analyzed to determine whether they differed on the features described by Biber (1988) and others as constituting the involved–informational dimension. Previous studies suggested that texts by Gender F and female writers would exhibit more of the involved characteristics. Table 1 presents findings regarding involved features and Table 2 regarding informational features.

As Table 1 shows, only three of 23 involved characteristics studied—present-tense verbs, pronoun *it*, and *be* as a main verb—varied by gender statistically significantly. But contrary to the expectation that women use involved features more than men do, all three of these features were used more frequently by Gender M authors. Note that the effect size of the difference for these features did not reach even the minimum threshold of $r > 0.10$ that Cohen recommended for calling an effect "small" (Fritz et al., 2012). In fact, among these 23 characteristics, only five exhibited effect sizes that reached the small threshold and none exhibited a medium ($r > 0.30$) or large ($r > 0.50$) effect size (Fritz et al., 2012).

Table 2 presents the smaller number of informational features, none of which varied

---

[7]No adjustment to α was made here to address the multiple-comparison problem: When one compares two samples using multiple variables, the probability that one variable will vary between the samples merely by chance increases as a statistical matter (Utts & Heckard, 2006). The Bonferroni adjustment provides a means for addressing this concern, dividing α by the number of comparisons being made or the number of variables. But the Bonferroni adjustment makes the assumption that the variables are not related, and Biber's dimensional model is built expressly on the observation that these features tend to co-occur.

significantly with gender and none of which exhibited an effect size reaching even the "small" threshold.

Discussion

Relevance theory predicts a tension between gendered aspects of linguistic register and genre knowledge. Those elements of linguistic register associated with gendered communicative performance are easy for an actor to access and use because they are typified: We are trained early and often to engage in gendered performances. Gendered communicative performances are therefore relevant to the communicator because she expects (probably unconsciously) that they will be relevant to the audience. To the extent that a professional genre requires a departure from gendered register, performing it requires much greater cognitive effort from the novice. But abandoning gendered register probably has commensurate cognitive effects: In the classroom, success means a grade that may prove important in finding a job and for other material reasons; in the workplace, success in using professional genres may be essential for success as a professional.

In this study, novice legal writers, after receiving a year of professional training and while writing in a form they understood to be convention-bound, wrote without register features that varied with their genders. Unfortunately, we do not have a pre-test showing whether study participants *entered* these law schools exhibiting differences in their writing that varied with their genders. This study is nevertheless useful in showing that *even if the participants* had *entered law school exhibiting gender-related differences in their writing*, they did not exhibit any such differences in this year-end project.

This study does not support the view that communicative differences of men and women (if they exist to begin with) are grounded in biology (a possible implication of studies like Skuse et al. (1997)) or in deep acculturation that would be difficult to overcome (as suggested by Maltz and Borker (1982), Tannen (1991/2001)). This study suggests that men and women can adapt to communicative genres readily without regard to their genders

(again, assuming they came to the task with gender-linked differences in the first place).

Relevance theory can help us see the findings of this study as part of a pattern consistent with earlier studies. In the Biber et al. (1998) study, we noted that the involved–informational characteristics of a text depended on the gender both of the author and the recipient, so that letters from women to other women showed more involved characteristics than did letters from women to men, and letters from men to women showed more involved characteristics than did letters from men to men but less than letters from women to men. In relevance theoretic terms, the authors prepared communication that reflect an intersection between their typified, gendered communicative performances and the practices that they believed would be most relevant to their audiences. But gendered practices won out here—male-to-female letters were still more informational than female-to-male letters. We can attribute this both to the strength of typified, gendered practices and to the flexibility of the letter genre, the style of which is not as convention-bound as that of many professional genres.

The Argamon et al. (2003) study represents a middle case. There, authors were writing in genres that are perhaps more definite in terms of linguistic register than private correspondence. The texts they examined were published texts in a variety of genres (which they grouped together under the categories *fiction* and *non-fiction*). These pieces, written for a broader audience than private letters, likely reflected greater author concern for conventions of the text-type or genre. In relevance-theoretic terms, authors sought to meet expectations of the readers in authors' mutually manifest cognitive environments. Nevertheless, the broad categories of "fiction" and "non-fiction" do not raise such definite expectations that writers have a goal or incentive to overcome all typified, gendered practices. In other words, writers may not have believed the cognitive effort of abandoning gendered practices would have been accompanied by a commensurate increase in desired effects. Consequently, not all of the involved–informational features varied with gender.

Contrast that study with the studies in the business-writing and

technical-communication classrooms (Smeltzer & Werbel, 1986; Sterkel, 1988; Tebeaux, 1990). In each of those cases, a student was being trained in professional genres to write a graded assignment for her instructor. These students' understanding of audience expectations—the students' mutually manifest cognitive environments—were likely more definite than in the letter and fiction/non-fiction studies. Similarly, the students in the present study had been exposed for several months to the register of legal writing as they read court opinions in their law-school classes. Their textbooks provided examples of the type of writing they were being asked to do (Clary & Lysaght, 2010; Schmedemann & Kunz, 2007), and some individual instructors provided additional examples. Their textbooks also addressed some particular issues of register: avoiding excessive use of adjectives, adverbs and emphatics, the use of particular verb tenses (particularly emphasizing the past tense), and preferring verbs over nominalizations. In each student's case, the brief was one of only two or three graded assignments in the semester and one weighted most heavily in assigning final grades.

The students' assumptions about the genre combined with very definite mutually manifest cognitive environments for their instructors resulted in students' cognitive environments where the cognitive effect—achieving the goal of a good grade on an important assignment—warranted the effort to conform in every way possible with audience expectations, including abandoning typified, gendered communicative practices entirely.

Conclusion

Relevance theory explains the findings in this study and earlier ones, but it may also be useful for theorizing other phenomena in communication and gender. The *citationality* of Butler (1993) and the *dynamism* of Berkenkotter and Huckin (1995) acknowledge the tension between stability and change in gender and genre theory. Relevance theory provides a framework for understanding how individual performances fit into this model of performativity (Butler, 1993) or structuration (Berkenkotter & Huckin, 1995). Relevance

theory emphasizes the situated nature of genre knowledge (Berkenkotter & Huckin, 1995) and the intersection of identity characteristics like race and SES with gender to which Butler (1993, p. 116) counsels us to attend.

Given these characteristics, I will suggest that relevance theory may prove a fruitful model for understanding human communication, a model that is more *rhetorical* in its commitments than other theories of linguistic pragmatics—particularly those of Austin (1975), Grice (1989), and Searle (1970, 1979).

Several questions remain regarding the register associated with this genre and these gendered persons. First, a pre-test of participants in the present study writing in an ungenred text would have shown whether the law students entered law school with a tendency to use gendered register. If they did not, then the application of relevance theory here really explains nothing. Second, the participants in this study skewed to persons under age 30, and it is possible that they were still socializing to their genders as well as their professions. A study of older participants might have found more 'stubborn' gender differences, or perhaps gender differences appear (or reappear) in the writing of lawyers a number of years after they have left law school. This question could be the object of future study, though it is often difficult to identify texts written by individual women and men in law firms where many attorneys may have a hand in drafting memoranda such as those studied here.

The current study, like many of these cited here, may be criticized on grounds that it lacks statistical power. In other words, the sample may be too small to show differences that are statistically significant. Note that I have made no effort here to provide a post-hoc power analysis, as they are disfavored in the literature and in any event are isomorphic with *p*-values (Walters, 2009). The small or negligible effect sizes seen in these samples, however, suggest that further research with larger samples will not be successful in revealing statistically significant differences that are of practical significance.

Another question is commonly raised in the literature: Whether women are forced to

assimilate to male communicative practices when they enter the professions. Earlier studies (Smeltzer & Werbel, 1986; Sterkel, 1988) took it for granted without empirical basis that professional discourse is masculine discourse; this is probably an a priori conclusion grounded in the fact that the professions had been dominated for centuries by men. If this were true, relevance theory predicts that women would have greater difficulty assimilating to legal writing than men, because they would be required to abandon more of their typified, gendered communicative practices. Future study may answer this question.

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.

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2007). Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday*, *12*(9). Retrieved September 27, 2010, from http://firstmonday.org/issues/issue12_9/argamon/index.html

Austin, J. L. (1975). *How to do things with words* (2nd ed.). Oxford; New York: Oxford University Press.

Bazerman, C. (2013). *A theory of literate action: Literate action Volume 2.* Fort Collins, CO: The WAC Clearinghouse.

Berkenkotter, C. & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power.* Hillsdale, NJ: Lawrence Erlbaum.

Biber, D. (1988). *Variation across speech and writing.* Cambridge U.K.: Cambridge University Press.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison.* New York: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge U.K.: Cambridge University Press.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python* (1st ed.). Cambridge, MA: O'Reilly Media.

Butler, J. (1993). *Bodies that matter: on the discursive limits of "sex".* New York: Routledge.

Cameron, D. (2003). Gender and language ideologies. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (pp. 447–467). Malden, MA: Blackwell Publishing.

Castillo, M. (2013, December 3). Different brain wiring in men, women could explain
    gender differences [CBS news]. Retrieved November 27, 2015, from
    http://www.cbsnews.com/news/different-brain-wiring-in-men-women-could-explain-
    gender-differences/

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

Dascal, M. & Gross, A. G. (1999). The marriage of pragmatics and rhetoric. *Philosophy &
    Rhetoric, 32*(2), 107–130.

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

Eckert, P. & McConnell-Ginet, S. (1992/2011, March). Communities of practice: where
    language, gender, and power all live. In J. Coates & P. Pichler (Eds.), *Language and
    gender: a reader* (2nd ed., pp. 573–582). Chichester, West Sussex, U.K. ; Malden,
    MA: Wiley-Blackwell.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect Size Estimates: Current Use,
    Calculations, and Interpretation. *Journal of Experimental Psychology: General,
    141*(1), 2–18.

Gray, J., Ph D. (1992). *Men are from mars, women are from venus: a practical guide for
    improving communication & getting what you want in your relationships.* New York,
    NY: Harper Collins.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University 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. N. (2016). *Gender/genre: Python code Github repository.* Github code
    repository. Retrieved from
    https://github.com/rhetoricked/WrittenCommunication2016

Larson, B. N. (In preparation). *Corpus of first-year law students' court memoranda assignments*. Linguistic Data Consortium.

Liu, Y. & Zhu, C. (2011). Rhetoric as the antistrophos of pragmatics: Toward a "competition of cooperation" in the study of language use. *Journal of Pragmatics*, *43*(14), 3403–3415.

Maltz, D. N. & Borker, R. (1982). A cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and social identity* (pp. 196–216). Cambridge U.K.: Cambridge University Press.

Matthews, P. (2007). Register. Oxford: Oxford University Press. Retrieved from http://www.oxfordreference.com/10.1093/acref/9780199202720.001.0001/acref-9780199202720-e-2841

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

Palander-Collin, M. (1999). Male and female styles in 17th century correspondence: I THINK. *Language Variation and Change*, *11*(02), 123–141.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow, Essex, UK: Pearson Education.

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.

Reymann, M. (2002, June). *The textual dimension involved-informational: A corpus-based study* (Magisterarbeit (Master's), Universität Regensburg, Regensburg).

Robb, A. (2014, May 14). Women get interrupted more—even by other women. *The New Republic*. Retrieved November 27, 2015, from https://newrepublic.com/article/117757/gender-language-differences-women-get-interrupted-more

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

Searle, J. R. (1970). *Speech acts: An essay in the philosophy of language.* London: Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts.* Cambridge U.K.: Cambridge University Press.

Skuse, D. H., James, R. S., Bishop, D. V. M., Coppin, B., Dalton, P., Aamodt-Leeper, G., ... Jacobs, P. A. (1997, June 12). Evidence from turner's syndrome of an imprinted x-linked locus affecting cognitive function. *Nature, 387*(6634), 705–708.

Smeltzer, L. R. & Werbel, J. D. (1986). Gender differences in managerial communication: Fact or folk-linguistics? *Journal of Business Communication, 23*(2), 41–50.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Wiley-Blackwell.

Sterkel, K. S. (1988). The relationship between gender and writing style in business communications. *Journal of Business Communication, 25*(4), 17–38.

Straßheim, J. (2010). Relevance theories of communication: Alfred Schutz in dialogue with Sperber and Wilson. *Journal of Pragmatics, 42*(5), 1412–1441.

Tannen, D. (1991/2001). *You just don't understand: Women and men in conversation* (1st ed.). William Morrow Paperbacks.

Tebeaux, E. (1990). Toward an understanding of gender differences in written business communications: A suggested perspective for future research. *Journal of Business and Technical Communication, 4*(1), 25–43.

Utts, J. M. & Heckard, R. F. (2006). *Statistical ideas and methods.* Belmont, CA: Thomson Brooks/Cole.

Walters, S. J. (2009, June). Consultant's forum: Should post hoc sample size calculations be done? *Pharmaceutical Statistics, 8*(2), 163–169.

Wilson, D. & Sperber, D. (2006). Relevance theory. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Hoboken, NJ: Wiley-Blackwell.

Wilson, D. & Sperber, D. (Eds.). (2012). *Meaning and relevance.* Cambridge U.K.:
    Cambridge University Press.

Yan, X. & Yan, L. (2006). Gender classification of weblog authors. In *AAAI Spring
    Symposium: Computational Approaches to Analyzing Weblogs* (pp. 228–230).

Yus, F. (2015, January). Relevance theory online bibliographic service. Retrieved January
    31, 2015, from http://personal.ua.es/francisco.yus/rt.html

Zimmer, B. (2012, November 4). How twitter language reveals your gender—or your
    friends: social media is giving linguists new insight into how speech varies. *The
    Boston Globe.* Retrieved November 27, 2015, from
    https://www.bostonglobe.com/ideas/2012/11/03/how-twitter-language-reveals-your-
    gender-your-friends/e68H6Z0Z2GAfnJ6UjU3IxO/story.html

Table 1

Involved register features by author gender (relative frequency per 1,000 tokens)

Feature	Gender M (n = 89)		Gender F (n=104)		Gender Prevalence	Mann-Whitney U-value	p-value		Effect size (r)
	Mean	StDev	Mean	StDev					
Private verbs	12.82	3.45	13.04	3.25	F	4444.0	0.318		0.03
That-deletion	8.62	2.20	8.67	2.46	F	4517.0	0.388		0.02
Contraction	0.20	0.36	0.27	0.75	F	4561.5	0.422		0.01
Present-tense verbs	36.43	6.02	33.74	5.99	M	3546.5	0.003	*	0.20
First-person pronouns	0.53	0.58	0.46	0.72	M	4022.5	0.054		0.11
Second-person pronouns	0.24	0.33	0.28	0.63	F	4270.0	0.143		0.07
Pronoun *IT*	5.66	2.44	4.94	2.02	M	3869.5	0.025	*	0.14
Indefinite pronouns	0.73	0.89	0.64	0.81	M	4172.5	0.118		0.08
Demonstrative pronouns	2.58	1.42	2.43	1.25	M	4434.5	0.309		0.04
Pro-verb *DO*	3.68	1.73	4.22	2.62	F	4215.5	0.143		0.08
Analytic negation	9.17	2.88	9.68	4.13	F	4606.0	0.478		0.00
General emphatics	2.08	1.19	1.93	1.22	M	4262.0	0.172		0.07
BE as main verb	13.38	3.61	12.63	3.30	M	3939.0	0.038	*	0.13
Causitive subordination	2.71	1.77	2.63	1.50	M	4626.5	0.499		0.00
Discourse particles	0.02	0.07	0.01	0.05	M	4398.0	0.061		0.04
General hedges	0.07	0.13	0.09	0.23	F	4614.0	0.481		0.00
Amplifiers	0.67	0.53	0.62	0.64	M	4081.0	0.078		0.10
Sentence relatives	0.74	0.65	0.69	0.66	M	4241.5	0.158		0.07
WH-questions	0.02	0.07	0.01	0.05	M	4509.0	0.154		0.02
Possibility modals	4.70	1.85	4.31	1.59	M	4173.0	0.120		0.08
WH-clauses	0.60	0.57	0.51	0.43	M	4329.0	0.219		0.06
Final prepositions	0.77	0.61	0.70	0.57	M	4244.5	0.161		0.07
Adverbs	37.94	7.07	38.37	7.38	F	4581.0	0.452		0.01

* Designates difference between Gender F and Gender M significant at $p < 0.05$.

Table 2

Informational register features by author gender (relative frequency per 1,000 tokens, except †)

Feature	Gender M (n = 89)		Gender F (n=104)		Gender Prevalence	Mann-Whitney U-value	p-value	Effect size (r)
	Mean	StDev	Mean	StDev				
Nouns	284.02	19.28	286.61	18.70	F	4191.0	0.130	0.08
Mean word length†	5.02	0.17	5.01	0.17	M	4411.0	0.288	0.04
Prepositions	128.81	8.76	128.37	11.11	M	4597.0	0.469	0.01
Type-token ratio†	48.07	3.38	47.95	3.80	M	4459.0	0.331	0.03
Attributive adjectives	46.34	7.74	45.99	8.09	M	4426.0	0.301	0.04

* Designates difference between Gender F and Gender M significant at $p < 0.05$.