So Much for Plain Language: An Analysis of the Accessibility of U.S. Federal Laws Over Time

Eric Martínez[1], Francis Mollica[2], and Edward Gibson[1]
[1] Brain and Cognitive Sciences, Massachusetts Institute of Technology
[2] School of Psychological Sciences, University of Melbourne

Over the last 50 years, there have been efforts on behalf of the U.S. government to simplify legal documents for society at large. However, there has been no systematic evaluation of how effective these efforts—collectively referred to as the "plain-language movement"—have been. Here we report the results of a large-scale longitudinal corpus analysis ($n \approx 225$ million words), in which we compared every law passed by congress with a comparably sized sample of English texts from four different baseline genres published during approximately the same time period. We also compared the entirety of the U.S. Code (the official compilation of all federal legislation currently in force) with a large sample of recently published texts from six baseline genres of English. We found that laws remain laden with features associated with psycholinguistic complexity—including center-embedding, passive voice, low-frequency jargon, capitalization, and sentence length—relative to the baseline genres of English, and that the prevalence of most of these features has not meaningfully declined since the initial onset of the plain-language efforts. These findings suggest top-down efforts to simplify legal texts have thus far remained largely ineffectual, despite the apparent tractability of these changes, and call into question the coherence and legitimacy of legal doctrines whose validity rests on the notion of laws being easily interpretable by laypeople.

Public Significance Statement
Over the last 50 years, there have been efforts on behalf of the U.S. government to simplify legal documents for society at large. How effective have these efforts been? Here we analyzed every law passed by Congress between the years 1951 to 2022 along with a large sample of baseline texts published over roughly the same time period and found that laws remain laden with features associated with processing difficulty relative to other genres, and that the prevalence of most of these features has not meaningfully declined. We found similar results when comparing laws with academic texts. These findings suggest top-down efforts to simplify legal texts have thus far remained largely ineffectual, despite the apparent tractability of these changes, raising and informing questions of law and public policy.

Ignorantia juris nonexcusat is an ancient maxim of the law which holds that "ignorance of the law is no excuse" (Garner & Black, 2004). This ancient maxim remains at the heart of modern legal systems, which typically presume that the public understands the entirety of the legal doctrine and, consequently, do not typically allow ignorance or mistakes of the law as a defense to a crime (American Law Institute, 1984; Arsanjani, 1999). Of course, the presumption that a nation's citizenry is aware of the content of its laws does not appear to be well-grounded in fact. While part of the public's ignorance of the law may be attributed to a mere lack of exposure, it seems intuitively obvious that when the public does attempt to understand legal documents they have difficulty doing so. Indeed, the difficulty of reading legal texts has long been acknowledged not just by those tasked with reading these documents

---

Eric Martínez ⬡ https://orcid.org/0000-0003-2180-6268

but by those creating these documents as well. Sporadic attempts to draw up laws in "simple language, using words that everyone could understand" date back as far back as the eighteenth century in Europe (Mattila, 2016), but have mostly been ignored (Adler, 2012).

In the United States, top-down efforts to simplify government documents for the benefit of the public began as early as the 1970s, when Richard Nixon mandated that the Federal Registry be drafted in "layman's terms" and Jimmy Carter issued Executive Orders intended to make government regulations "easy-to-understand by those who were required to comply with them" (Exec. Order No. 13648, 1979; Plain Language Action Information Network, 2011). These and subsequent attempts to make government language more accessible have been collectively referred to as the "plain-language movement." The most recent call-to-arms, the Plain Writing Act of 2010, established formal guidelines regarding how to write government documents clearly for a lay audience (Plain Writing Act of 2010, n.d.).

The plain-language movement spurred research exploring how best to write "plain-English" layperson summaries of official legal documents, such as jury instructions (Charrow & Charrow, 1979; Diamond et al., 2012; Elwork et al., 1982; Heuer & Penrod, 1989) and Miranda warnings (Goldstein et al., 2003; Rogers et al., 2007). Many of the insights from this literature, as well as the general psycholinguistic literature, are now reflected in the Federal Plain Language Guidelines. While these studies have successfully demonstrated the feasibility and importance of using "plain-English" layperson summaries of legal documents to improve comprehension of legal content among laypeople, these examples apply only to a small portion of the total corpus of legal language and appear less relevant to people's experience with the legal system than actual laws.[1]

With regard to official legal documents, recent work has found that private contracts, such as online terms of service agreements, remain laden with complex psycholinguistic features, including center-embedding and low-frequency jargon (Martínez et al., 2022). Recent experimental work has also found that people are less able to understand and recall legal documents drafted with these features relative to legal documents of equivalent meaning drafted without these features (Martínez et al., 2022, 2023a).

With respect to public legal documents, however, there remains no systematic analysis of to what extent the plain-language movement impacted the accessibility of federal laws.

Moreover, on a more general level, there also remains no systematic evaluation of the accessibility of federal laws over time relative to baseline forms of English. In addition to comparing legal texts to forms of "standard" or "plain" English, such as newspaper articles or popular press books, comparing the accessibility of laws relative to more conceptually complex forms of writing, such as academic texts, might reveal the extent to which the inaccessibility of legal texts can be attributed to inherently complex concepts as opposed to needlessly complex psycholinguistic structures. Given that academics are also tasked with establishing and communicating complex ideas that are relevant to the general public, such a comparison could also provide useful insight regarding how well the academic community is successfully achieving that aim relative to lawmakers.

As alluded to above, the potential inaccessibility of official legal documents poses problems not just for those tasked with reading legal documents but for the validity of the documents themselves, as well as the coherence and legitimacy of legal doctrines that either expressly assert or implicitly assume that legal documents are or should be easily interpretable by laypeople.

For example, in U.S. constitutional law, the Fair Notice Doctrine requires "that laws give the person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly" (Grayned v. City of Rockford, 1972; Love, 2011; McBoyle v. United States, 1931; Robinson, 2005). Insofar as laws are incomprehensible to the typical layperson, this would arguably imply that laws are not giving laypeople fair notice, which would in turn undermine both the constitutionality of those laws and the legitimacy of the fair notice doctrine.

Meanwhile, the Ordinary Meaning Doctrine, which has been referred to as "the most fundamental principle of legal interpretation," not only of U.S. law but of jurisdictions across the world, requires that words in legal documents typically be interpreted according to how they are ordinarily understood by laypeople (Moskal v. United States, 1990; Richards v. United States, 1962; Slocum, 2019; Slocum & Wong, 2021; United States v. Turkette, 1981). However, insofar as legal documents are not ordinarily understood by laypeople, the coherence and legitimacy of this doctrine would also be undermined.

To address the above questions, we first conducted a corpus analysis of (a) every law passed by congress between January 1951 and May 2022 (as well as concurrent resolutions not signed into law and proclamations issued by the president), and (b) a large sample of magazine articles, newspaper articles, nonfiction books, and fiction books published over roughly the same time span. We analyzed a variety of linguistic and stylistic features, whose use is (a) discouraged by the Federal Plain Language Guidelines (Plain Language Action Information Network, 2011), (b) associated with language processing difficulty in psycholinguistics research (Martínez et al., 2022, 2023a, 2023b; Masson & Waldron, 1994), and (c) purportedly common in legal documents (Tiersma, 1999). We found that most of these features had not meaningfully decreased in prevalence since the start of the plain-language movement, although we find an increase in the variability of some features post-2010. Nonetheless, compared to time-matched baseline texts, each of the features remains strikingly more prevalent in public legal documents.

We additionally conducted a comparison between (a) the entirety of the U.S. Code (an official compilation of every federal law currently in force) and (b) a broad sample of six baseline texts from the Corpus of Contemporary English. We found that even compared to academic articles, laws contained higher rates of nearly every complex psycholinguistic feature we looked at, suggesting that the inaccessibility of laws may be the result of needlessly complex linguistic structures as opposed to inherently complex concepts.

Materials and Method

Corpus Materials

For our primary analysis, we constructed an exhaustive corpus of every public law, private law, concurrent resolution, and proclamation issued by the American federal government between the 1951

---

[1] For example, although jury instructions can be an important part of cases that go to trial, a small and diminishing percentage of civil and criminal cases actually go to trial (as low as 3% for the former and 5% for the latter: Rakoff et al., 2014; Refo, 2004). Moreover, while Miranda warnings provide crucial information to criminal suspects in police custody, the majority of individuals' contact with legal language takes place outside the context of criminal or civil suits.

and May 2022 using publicly available online resources from the United States Library of Congress (Library of Congress, 2021). As a baseline, we extracted a comparably sized sample of English texts drawn from the Corpus of Historical American English (Davies, 2012), which consisted of a broad sample of fiction books, nonfiction books, magazine articles, and newspaper articles also published between 1951 and 2009.

Because the Corpus of Historical American English only extends to 2009, we did not directly compare laws from 2010 to 2022 with this baseline corpus.

In addition to our primary materials, we also collected two additional corpora to compare the linguistic complexity of current laws relative to a baseline of contemporary texts, including those of comparable conceptual complexity. These two additional corpora consisted of (a) the 2021 edition of the U.S. Code, the official compendium of all federal laws that are in effect in the United States and (b) a comparably sized sample of academic texts, fiction books, magazine articles, newspaper articles, and spoken English drawn from the Corpus of Contemporary American English.[2]

Preprocessing Tools and Indices

To process and analyze our corpora, we used a number of natural language processing (NLP) tools. One of the primary tools we used was the Stanford Stanza natural language package (Qi et al., 2020), a state-of-the-art NLP toolkit which we used to tokenize each document into sentences, lemmatize and tag each word by part of speech, and syntactically parse each tokenized sentence. Stanza has been shown to achieve over 90% accuracy on a variety of NLP tasks (Qi et al., 2020). To verify its accuracy on our specific corpora and for our specific metrics, we spot-checked a random sample of 1,000 sentences across our corpora by (a) hand-coding whether a given sentence had a passive-voice structure or a center-embedded clause, and (b) for each sentence comparing whether the parser's judgments aligned with the hand-coded judgments. Using this method, we found that the parser was 97.93% accurate at detecting by-passive structures (95% CI [97.04, 98.82]) and 88.95% accurate at detecting center-embedding structures (95% CI [86.98, 90.73]).

In addition to Stanza, we also used the SUBTLEX word frequency dictionary (Brysbaert & New, 2009), which we used to get a word frequency estimate as a proxy for how common a given word in each corpus appears in everyday speech. The SUBTLEX frequency values themselves are derived from a large-scale corpus of American film subtitles have been shown to correlate with reading-time behavior (Brysbaert & New, 2009). Finally, we also used WordNet (Miller, 1995), which, in tandem with SUBTLEX, was used to estimate whether a given word could have been replaced by a higher frequency word with the same meaning.

Preprocessing for all corpora was identical. Sentences were first tokenized and dependency-parsed using the Stanford Stanza NLP package. We then removed sentences without punctuation, as well as those with fewer than 10 words so as to remove headings, which are not really sentences but would otherwise be counted as such. We also removed sentences with 3+ consecutive punctuation marks or related symbols (such as "@") so as to get rid of more nonsentences in both corpora. The total number of words after filtering was 150,393,499 (47,769,955 words for the legal corpus and 102,623,544 for the nonlegal corpus). After filtering out nonsentences, we then dependency-parsed each corpus, lemmatized

and tagged each word by part of speech and computed our indices of processing difficulty, which we further clarify and motivate below.

Word Frequency

For each of our corpora, we sought to determine, on average, how frequently the words in said corpora occur in everyday speech. Words that are infrequently used in everyday speech cause comprehension difficulties for readers relative to higher-frequency synonyms (Marks et al., 1974). Legal language is reportedly laden with low-frequency jargon, such as aforesaid, hereinafter, and to wit (Rayner et al., 2004), and recent work has shown the language in contracts to be lower-frequency than that of other genres of English (Martínez et al., 2022). According to the official plain-language guidelines, government writing should avoid the use of such low-frequency "dry legalisms" and "jargon" (Plain Language Action Information Network, 2011).

Frequency values were extracted from the SUBTLEX corpus of American film subtitles (Brysbaert & New, 2009), commonly used as a proxy for standard-English word frequency and which has been shown to correlate with reading-time behavior. Because the impact of word frequency on reading times is logarithmic as opposed to linear (Shain et al., 2024), we used the Zipf values (which are both logarithmic and standardized) as opposed to raw counts (Van Heuven et al., 2014).[3]

To avoid including noncontent words, we limited our analysis of frequency to the words in our corpora marked as a verb, noun, adjective, or adverb according to Stanza. Proper nouns and other words that did not appear in the SUBTLEX corpus received a score of not applicable, and not applicable values were excluded from the analysis.

Word Choice

Although many argue that the processing difficulty of unfamiliar language is a necessary consequence of the specialized concepts and corresponding terminology used to refer to those concepts by lawyers (cf., Tobia, 2020), recent work suggests that private legal documents contain a high proportion of overly complicated language that can be replaced with simpler terms that have the same meaning (Martínez et al., 2022). The official plain-language guidelines encourage the use of "familiar or commonly used" words over such "unusual," "obscure" or "unnecessarily complicated language" (Plain Language Action Information Network, 2011). Here we sought to quantify the amount of unnecessarily complicated language in federal laws by calculating the percentage of words in each corpus that could have been replaced with a higher-frequency synonym.

We operationalize word choice difficulty as the proportion of content words in each corpus that had a higher-frequency synonym. We conducted three versions of this analysis using three separate assumptions. Under the first version, we make the conservative assumption

---

[2] According to Corpus of Contemporary American English documentation, the academic texts were drawn from more than 200 different peer-reviewed journals and cover the full range of academic disciplines, with a good balance among education, social sciences, history, humanities, law, medicine, philosophy/religion, science/technology, and business. The texts were published between 1990 and 2012.

[3] The Zipf values we used can be found at: https://osf.io/djpqz/files/osfstorage.

that the authors intended the least common sense of each word used in a given corpus because while legal terms may resemble common words in form, they may have a more specialized meaning, such as the concept of "consideration" in contract law (American Law Institute and National Conference of Commissioners on Uniform State Laws, 2002).

Under the second version, we make an anticonservative assumption that the authors intended the most common sense of each word in a given corpus. Under the third version, we assume that the authors intended neither the most common nor least common sense of each word but rather a random sense of a word in a given corpus. Again, we limit our analysis to verbs, common nouns, adjectives, and adverbs.

For all three methods, we determined the least common (conservative), most common (Anticonservative 1), or random (Anticonservative 2) meaning/sense of that word according to WordNet (Miller, 1995). For all words sharing that meaning/sense (i.e., synonyms), we looked up the SUBTLEX frequency value and coded whether the SUBTLEX frequency value of any synonym was higher than that of the actual word used in the text (1 = yes; 0 = no). Results of the conservative method are reported in the main text. The results of the two anticonservative versions are visualized in the online supplementary materials, along with three additional versions that extend the analysis to all words as opposed to content words. All six versions yielded converging results.

Figure 1

Comparison of Indices of Linguistic Processing Difficulty in Federal Laws Versus Four Genres of English, Including Fiction Books, Magazine Articles, Newspaper Articles, and Nonfiction Books (1951–2009)
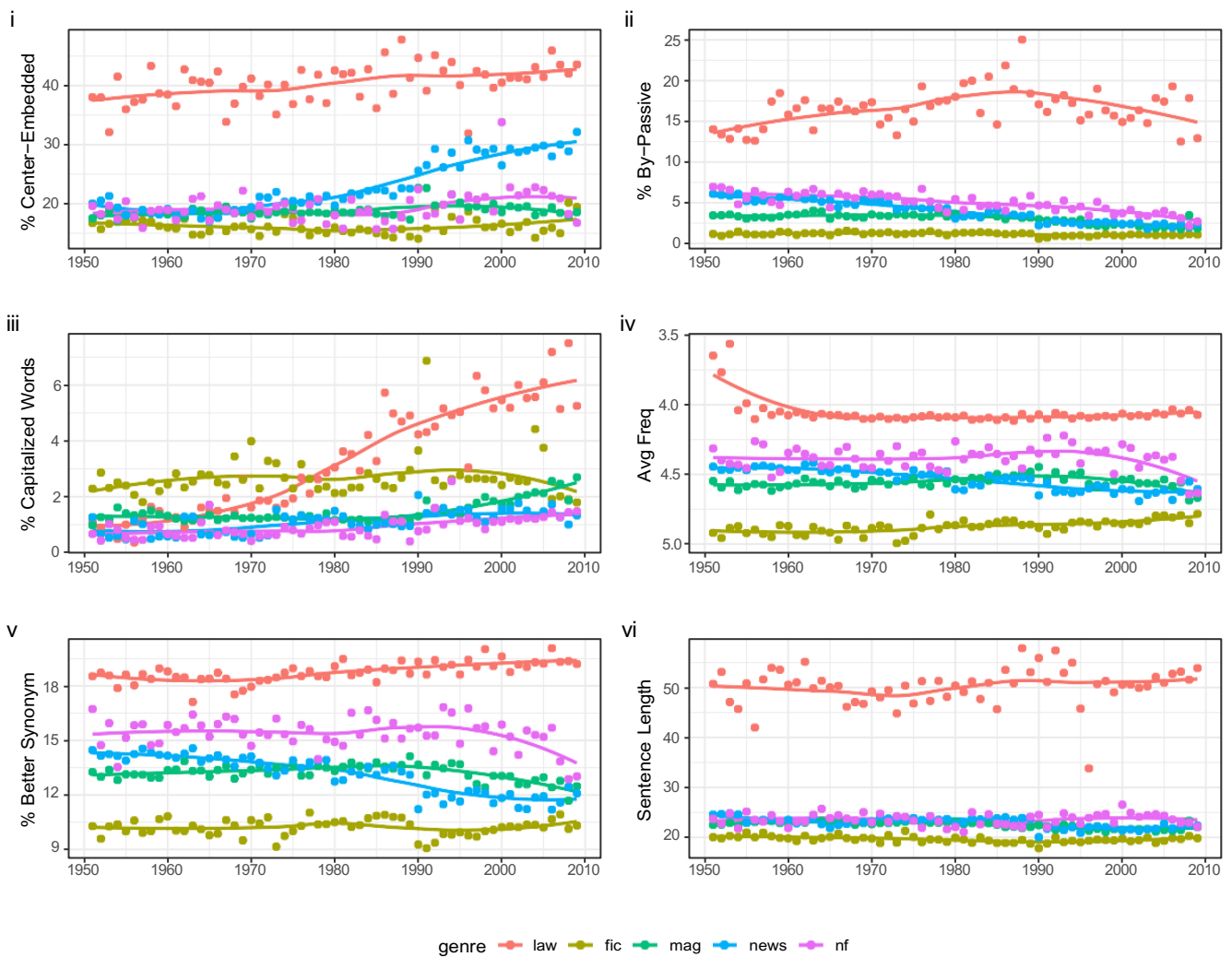
Note. For any given year, most, if not all texts, indices were vastly more prevalent in laws than any of the baseline genres. Individual points reflect mean values of an index within a genre. Lines reflect locally estimated scatterplot smoothing (LOESS) regression lines capturing the year-by-year trend of the prevalence of an index within each genre. Baseline texts were taken from the Corpus of Historical American English. Avg Freq = average word frequency; fic = fiction; mag = magazine; news = newspaper; nf = nonfiction. See the online article for the color version of this figure.

Capitalization

In each corpus, we computed the percentage of words that contained nonstandard capitalization (specifically, those that were in all caps). Although the plain-language guidelines do not discourage the use of all-capitalization in government writing, evidence suggests that nonstandard capitalization ("ALL WARRANTIES ARE HEREBY DISCLAIMED") is common in certain types of private legal documents (Martínez et al., 2022) and has shown to inhibit comprehension in older readers (Arbel & Toler, 2020), relative to standard capitalization.

We coded a word as being in "all caps" by calculating the proportion of alphabetic word tokens that were marked by Stanza as being entirely in uppercase letters.

Sentence Length

Plain-language guidelines encourage the use of shorter sentences so as to "break the information up into smaller, easier-to-process units" (Plain Language Action Information Network, 2011). Legal texts, especially laws and other public documents, are reportedly filled with long sentences (Hiltunen, 2012; Kurzon, 1997). Although some evidence suggests sentence length is less of a predictor of processing difficulty than center-embedding and other types of syntactic complexity (Marton & Schwartz, 2003), words per sentence remains a consistent measure of processing difficulty in the reading literature (Flesch, 2007; Solnyshkina et al., 2017).

Here we computed sentence length by calculating the number of alphabetic words in each sentence as determined by Stanza.

Center-Embedded Clauses

Plain-language guidelines discourage the use of "convoluted" sentences, particularly those that are "loaded with dependent clauses" and which separate the "essential parts" of a sentence from each other (i.e. the subject, verb, and object). The most notorious examples of such sentences contain center-embedded structures, in which a sentence or clause is embedded within the center of another sentence or clause ("all such payments and benefits, including the payments and benefits under Section 3(a) hereof, being hereinafter referred to as the 'total payments'"). Center-embedded structures cause processing difficulty for readers (Gibson, 1998; Miller & Chomsky, 1963) and have been shown to inhibit recall of legal content relative to clauses of equivalent meaning that have been unembedded into separate sentences (Martínez et al., 2022, 2023a, 2023b). Here we calculated the percentage of sentences in each corpus containing a center-embedded clause.

We coded a sentence as containing a center-embedded clause if a predicate-dependent clause as parsed by Stanza (i.e. clausal subjects, clausal complements, open clausal complements, adjectival clauses, and adverbial clauses) was followed by a word as opposed to an end-of-sentence punctuation mark. As noted above, this method was 88.95% accurate in identifying these structures in a given sentence (95% CI [86.98, 90.73]).

Passive-Voice Structures

For each corpus, we calculated the percentage of sentences containing a reversible passive-voice structure. Federal Plain Language Guidelines advocate for using the active voice instead of the passive voice. Passive-voice structures are acquired later than active-voice structures and have been shown to pose comprehension difficulties for adults in certain circumstances, particularly in the context of implausible sentences, for example, "the girl was kicked by the ball" (Ferreira, 2003). Although Martínez et al. (2022) recently found evidence that passive-voice structures did not inhibit recall of legal content relative to active-voice structures in contracts, it may be that the stimuli used in Martínez et al. (2022) did not span the circumstances shown to induce the comprehension errors seen in adult experiments. To err with caution, we include passive-voice structures in our analysis, particularly reversible passives or by-passives (e.g. "the information shall be maintained by the Federal Government" as opposed to "the information shall be maintained"), which can be more easily replaced by active-voice structures without a loss or distortion in meaning.

We coded a sentence as containing a reversible passive-voice structure if a word was marked with the passive-voice features by Stanza and had the word by in the same head according to the Stanza parse. As noted above, this method was 97.93% accurate in identifying by-passive structures in a given sentence (95% CI [97.04, 98.82]).

For robustness purposes, we also computed the percentage of sentences containing nonreversible passives. These results are reported in the online supplementary materials.

Analysis Plan

To evaluate the influence of the plain-language movement, for each of our six indices of processing difficulty, we conducted both a classical and a break-point Bayesian regression limited to the legal-corpus data. Whereas a classical regression estimates a single slope for predictors across their range, a break-point regression assumes there is a fixed "break" in the range of a predictor (in this case time) and estimates two sets of slopes: one set for before the break point and one set for after the break point. For word frequency and sentence length, we used a linear regression to predict the mean

Table 1

Estimates and 95% Confidence Intervals for the Intercept and Slopes of the Break-Point Regression Models in 1972

Metric	Intercept	Before 1972	After 1972	Dispersion	BF
Word frequency	4.614 [4.613, 4.615]	−0.002 [−0.002, −0.002]	−0.002 [−0.002, −0.002]	0.582 [0.582, 0.582]	Inf
Word choice	−1.58 [−1.63, −1.54]	−0.00 [−0.01, 0.00]	0.00 [−0.00, 0.00]	200.55 [165.56, 240.29]	5440
Capitalization	−3.60 [−3.74, −3.46]	0.04 [0.02, 0.05]	0.02 [0.01, 0.02]	83.20 [68.13, 99.59]	311
Center-embedding	−0.75 [−0.89, −0.62]	0.01 [−0.01, 0.02]	0.00 [−0.00, 0.01]	14.46 [12.03, 17.15]	2576
Sentence length	50.92 [50.71, 51.13]	0.7 [0.04, 0.09]	−0.01 [−0.02, −0.00]	49.63 [49.56, 49.70]	Inf
Passive voice	−1.93 [−2.09, −1.78]	0.02 [0.00, 0.03]	−0.00 [−0.01, 0.00]	21.05 [17.26, 25.46]	403

Note. Bayes factors reflect the evidence for a linear trend across years over a nonlinear (break-point) model. BF = Bayes factor; Inf = infinite.

Table 2
Estimates and 95% Confidence Intervals for the Intercept and Slopes of the Break-Point Regression Models in 2010

Metric	Intercept	Before 2010	After 2010	Dispersion	BF
Word frequency	4.562 [4.561, 4.563]	−0.001 [−0.002, −0.001]	−0.058 [−0.058, −0.057]	0.580 [0.580, 0.580]	Inf
Word choice	−1.59 [−1.64, −1.54]	−0.00 [−0.00, 0.00]	0.02 [0.01, 0.03]	207.06 [170.81, 247.55]	86
Capitalization	−2.86 [−2.99, −2.73]	0.02 [0.02, 0.03]	−0.04 [−0.10, 0.00]	84.19 [68.97, 101.28]	22
Center-embedding	−0.61 [−0.75, −0.47]	0.00 [−0.00, 0.01]	0.02 [−0.03, 0.06]	14.56 [12.04, 17.32]	847
Sentence length	51.49 [51.30, 51.68]	0.03 [0.02, 0.04]	−0.68 [−0.75, −0.61]	49.62 [49.55, 49.69]	≈0
Passive voice	−1.96 [−2.12, −1.80]	0.00 [−0.00, 0.01]	0.02 [−0.04, 0.06]	20.80 [17.12, 24.89]	712

Note. BFs reflect the evidence for a linear trend across years over a nonlinear (break-point) model. BF = Bayes factor; Inf = infinite.

value of these metrics per sentence and report standard deviation as a measure of variance. For all other indices, we used a beta-binomial logistic regression, which estimates an overdispersion parameter, as a measure of variance, in addition to fixed effects. In the case of our sentence-level metrics (center-embedding and passive voice), the regression estimated the influence of our predictors on whether a sentence had a given metric. In the case of our word-level metrics (capitalization and word choice), the regression estimated the influence of our predictors on whether a word had a given metric. For all break-point regression models, we used normal priors, with a mean of 0 and a standard deviation of three (primarily for increased computational speed). Break-point models were compared with a baseline model containing a fixed effect of time—that is, one set of slopes, using Bayes Factors. All analyses were conducted using the brms package in R (Bürkner, 2017).

To more generally evaluate the accessibility of federal laws over time and as compared to plain English, we conducted separate Bayesian regression models (poisson for frequency and sentence length; logistic for remaining indices) that included both of our corpora. For each index, we first considered two models: one with a main effect of Corpus (Legal vs. baseline) and Year, and one with an additional interaction term between Corpus and Year. We used the default priors in brms and stan, which are flat priors. A Bayes-factor comparison for each index except average word frequency suggested the model with an interaction is a better explanation of the data (Bayes factor [BF] > 10) than the model without an interaction. For word frequency, both models perform equally well. We therefore only report the results of the model with an interaction term in Table 3.

Transparency and Openness

The methods of this paper comply with the TOP guidelines of the *Journal of Experimental Psychology: General*. In particular, all original data and code for this project can be cited as Martínez et al. (2023b) and are available at https://osf.io/ambp4/?view_only=b4ab367e4cfb4f83acd2c51a000cfa68.

Results

Efficacy of the Plain-Language Movement

Were the plain-language movement to have been effective, one would expect (a) the prevalence of difficult-to-process features to have meaningfully decreased over time, and (b) the decrease to coincide with the onset of the plain-language movement. To evaluate this prediction, for each of our six indices of processing difficulty, we conducted classical and break-point Bayesian

regressions limited to the legal corpus. We used two break points: 1972, a plausible year for the plain-language movement's call-to-arms (left panels of Figure 1), and 2010, the year of the passage of the Plain Language Act (right panels of Figure 1). If the plain-language movement overall had a simplifying effect, one would expect the slope of the regression line after the 1972 break point (i.e., 1972–2022) to be both negative (positive for word frequency) and less (greater for word frequency) than the slope of the regression line before the break point (i.e., 1951–1972). If the Plain Writing Act of 2010 led to a decrease in features, one would expect the slope of regression line after the 2010 break point (i.e., 2010–2022) to be both negative and less (positive/greater for word frequency) than the slope of the regression line before the break point (i.e., 1951–2010). Regression coefficients for all indices can be found in Table 1 for 1972 and Table 2 for 2010.

For both time points, a classical regression model—that is, a single linear effect of time, explains the data better than a break-point effect (BF > 10) for five out of six features: center-embedding, by-passives, capitalization, word choice, and word frequency, suggesting no significant impact of plain-language movements on decreasing the prevalence of these features. For the remaining feature (sentence length), a classical regression model also explains the data better than a break-point effect for 1972 but not for 2010. That said, it should be noted that the effect size, while significant, is negligibly small and likely driven by the large amount of data and higher variability in values on the sparser side of the break point (i.e., after 2010).[4]

Visually, the single slope linear effect of time is clear even from the posterior draws of the break-point model (the red shading in the left panels of Figure 1). Due to the increased variance in the 2010 models, the posterior draws appear as if there might be differences in slope across the break; however, this is less clear looking at the data points themselves. Further, for frequency, by-passive, and word choice, the slopes post the 2010 break point are trending toward worse language processing outcomes contra the plain-language movement. For capitalization, the slope post the 2010 break point is trending toward better language processing outcomes, which while not statistically significant, suggests improvement.

---

[4] As a follow-up, we conducted heteroskedastic break-point regressions—that is, we fit the variance before and after the breakpoint separately, for 2010. This analysis provides a little evidence (1 < BFs < 5) to suggest that the data are more variable post-2010. This extra variance is partly due to the fewer time points sampled post-2010.

Regardless, models with a linear effect of time are still better explanations of the data than the heteroskedastic break-point regressions.

Figure 2

Comparison of Indices of Linguistic Processing Difficulty in Contemporary Federal Laws Versus Five Genres of English, Including Academic Articles, Fiction Books, Magazine Articles, Newspaper Articles, and Spoken Language Transcripts
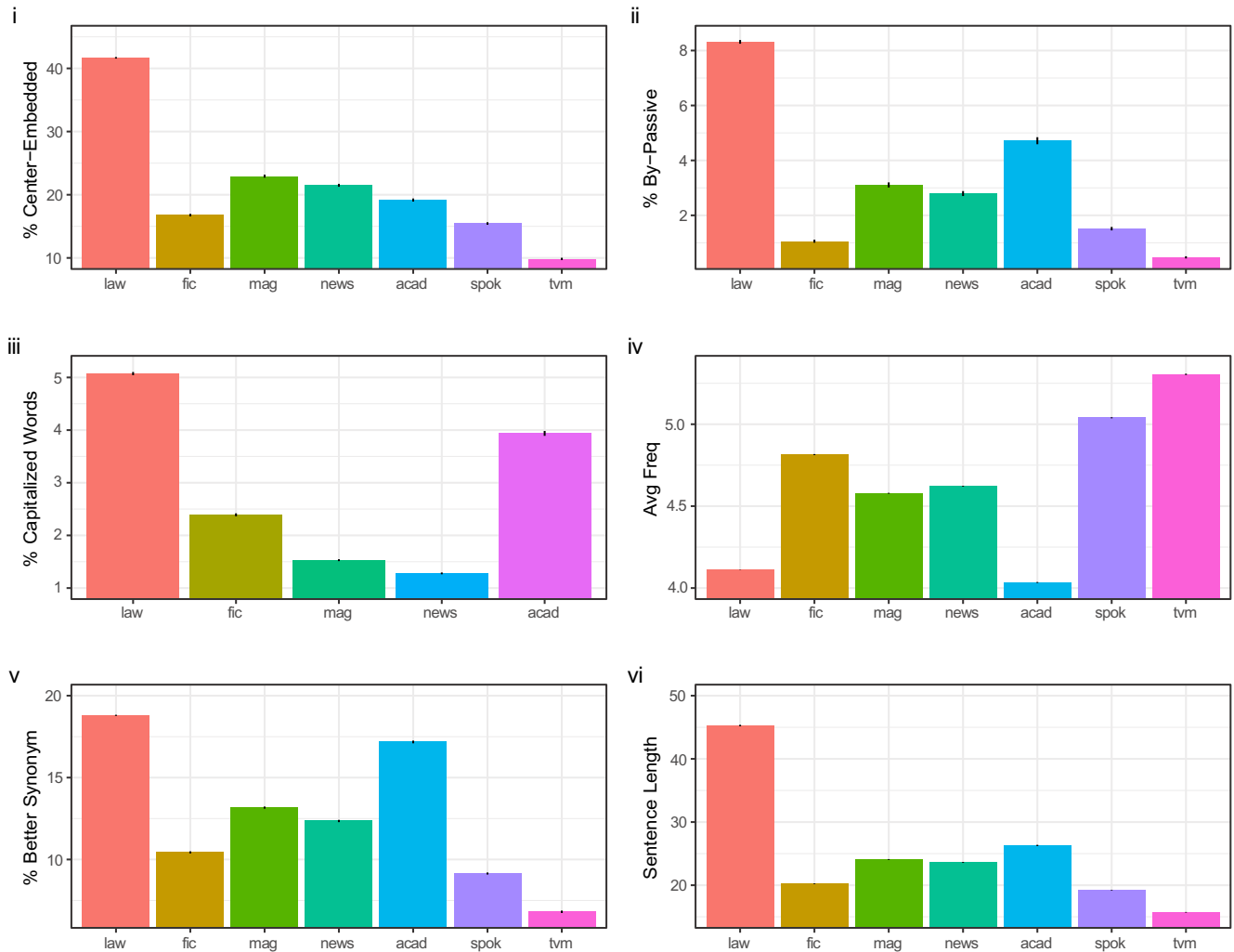
Note. Federal laws were taken from the 2021 edition of the United States Code. Baseline texts were taken from the Corpus of Contemporary American English. Height of bars reflects mean of index within a given genre, whereas error bars reflect 95% bootstrapped confidence intervals of the mean. With one exception (word frequency in academic texts), indices remain more prevalent in laws than any of the baseline genres. fic = fiction; mag = magazine; news = newspaper; acad = academic; spok = spoken; tvm = TV/movie; TV = television. See the online article for the color version of this figure.

General Trends in Accessibility of Legal and Nonlegal Language

Even if plain-language efforts have not coincided with a decrease in difficulty-inducing structures in legal texts, it may be the case that (a) difficulty-inducing structures became more prevalent in other texts relative to or as well as legal language, or that (b) legal language was not filled with very high indices of difficulty-inducing structures to begin with. To evaluate these alternative accounts, as well as to obtain a more general systematic account of the accessibility of federal laws—both temporally and relative to other genres of English—we first computed the descriptive statistics of each index within the corpora over time.[5] We found that for each year, the prevalence of virtually every metric was higher in federal laws than in any of the

four genres of the plain-language corpus (in most cases, the difference was striking). These results are visualized in Figure 2.

We then used Bayesian regression methods to estimate the influence of corpus (legal vs. baseline) over time (in years) for each of our indices of processing difficulty (results in Table 3). For every metric, our models revealed federal laws to contain more difficult to process structures than our baseline texts. While the credible intervals for our estimates of the main effect and interaction with time do not include zero, the parameters reflect very small effects (e.g., 20 years for change of

---

[5] Because the Corpus of Historical American English only extends to 2009, we did not directly compare laws from 2010 to 2022 with our baseline corpus.

Table 3
Estimates and 95% Confidence Intervals for the Intercept and Slopes of the Bayesian Regression Models, As Well As the BF Estimates in Favor of These Models Over Models Without an Interaction Term

Metric	Intercept	Corpus	Year	Corpus: year	BF
Word frequency	4.69 [4.69, 4.70]	−0.60 [−0.60, −0.60]	−0.00 [−0.00, −0.00]	0.00 [0.00, 0.00]	1.28
Word choice	−1.72 [−1.7, −1.72]	−0.24 [−0.25, 0.24]	0.00 [0.00, 0.00]	−0.00 [−0.00, −0.00]	Inf
All caps	−3.88 [−3.88, −3.88]	−0.13 [−0.13, −0.13]	0.02 [0.02, 0.02]	−0.01 [−0.01, −0.01]	Inf
Embedding	−0.97 [−0.97, −0.97]	−0.57 [−0.57, −0.56]	0.00 [0.00, 0.00]	−0.00 [−0.00, −0.00]	12.13
Sentence length	35.99 [35.60, 36.30]	−14.73 [−15.03, −14.34]	0.02 [−0.00, 0.04]	−0.04 [−0.06, −0.02]	Inf
Passive voice	−2.61 [−2.61, −2.60]	−1.00 [−1.01, −1.00]	−0.00 [−0.01, 0.00]	−0.01 [−0.01, −0.01]	Inf

Note.   BF = Bayes factor; Inf = infinite.

1%). If anything the tightness of the credible intervals is likely overestimated due to the large size of the corpora. Therefore, we interpret the results to suggest no meaningful influence of time on the prevalence of a given metric, nor of the interaction between time and corpus.

Accessibility of Contemporary Legal Versus Baseline Texts

Even if plain-language efforts failed to reduce the prevalence of complex psycholinguistic features in legal texts to the level of those in everyday text and speech, it is conceivable that this failure is a natural result of the higher conceptual complexity of legal texts relative to other texts. If so, one would predict that texts of similar conceptual complexity (such as academic articles) would have the same rate of psycholinguistically complex features as legal texts.

In order to account for this possibility, as well as to more generally compare laws to contemporary baseline texts, we conducted an additional comparison between (a) all U.S. federal laws in force as of 2021 (United States Code, 2021) and (b) a comparably-sized sample of academic texts, fiction books, magazine articles, newspaper articles, TV/movie scripts, and spoken language transcripts published in 2019 (Davies, 2009).

For all baseline genres except academic texts, each index of processing difficulty was disproportionately common in legal texts relative to the baseline texts. For academic texts, each index of processing difficulty except one (word frequency) was disproportionately more common in legal texts. In most cases, the difference was striking. Full results are reported in Figure 3.

Discussion

The present study first set out to investigate whether the plain-language movement succeeded in reducing certain features (a) that are associated with psycholinguistic complexity, (b) whose use is discouraged by plain-language advocates, and (c) that have been attested to be common in legal documents. According to our regression models, the slope of the line after 1972 suggested no change or harmful change, indicating that laws on balance had not gotten meaningfully simpler by our metrics since the onset of the plain-language movement. With regard to 2010, most (but not all) of our regression models did not reveal a positive change, indicating that the Plain Language Act of 2010 may have induced some modest improvements but did not coincide with a meaningful reduction of most of these features, either.

To further contextualize these findings, the present study next sought to investigate to what extent federal laws have deviated from baseline texts with respect to the presence of these features, both (a) over time between 1951 and 2009 and (b) at present. With regard to (a), as visualized and documented above, all of the metrics we looked at were startlingly more prevalent in federal laws than each of our baseline texts, with the relative prevalence failing to decrease over the examined time interval. With regard to (b), with one exception, all of the features we looked at were startlingly more prevalent in the U.S. Code than each of our baseline contemporary texts. Insofar as these features are accurate proxies for processing difficulty, then, in line with common intuition and plain-language advocates and consistent with recent findings regarding private legal documents (Martínez et al., 2022), this suggests that U.S. laws have been and continue to be more difficult to understand than other genres of English, including documents of comparable conceptual complexity, such as academic texts.

Our study provides the first systematic large-scale account of the accessibility of public legal language—both longitudinally and compared to more standard forms of English—substantiating previous anecdotal accounts of the efficacy of plain-language efforts made by plain-language advocates, who have described progress as "way slow" and acknowledged that "much remains to be done to improve" (Plain Language Action Information Network, 2011).

Having documented the profile of public legal language over the last 70 years and demonstrated the inefficacy of plain-language efforts over the same time period, further extensions to this study—both with regard to academic scholarship and government advocacy—should seek to confirm the extent to which these findings hold for other types of government documents, such as federal regulations and informational pages on government websites. For example, it may be the case that the plain-language movement led to a simplification not of laws themselves, but of supplemental supporting documents that provided a layperson's explanation of the content contained in those laws.

In addition, future work could also seek to understand the cause of the complexity of legal language. In other words, not only how lawyers and lawmakers write but why they choose to write the way that they do. One possibility is that the style in which laws are currently written is necessary to maintain communicative precision. Prior to our study, this hypothesis had been undercut by previous findings showing comprehension of legal content with a simplified register (Martínez et al., 2021, 2022; Masson & Waldron, 1994). Our results further undercut this possibility, as our analysis focused on features that are known to have simpler alternatives with equivalent meaning (e.g. "mala fides" vs. "bad faith"). While it seems entirely plausible that certain legal jargon is inevitable, our results suggest that in many instances such jargon can be replaced with simpler alternatives that preserve meaning.

Figure 3

Posterior Predictive Draws from the Breakpoint Regression Models
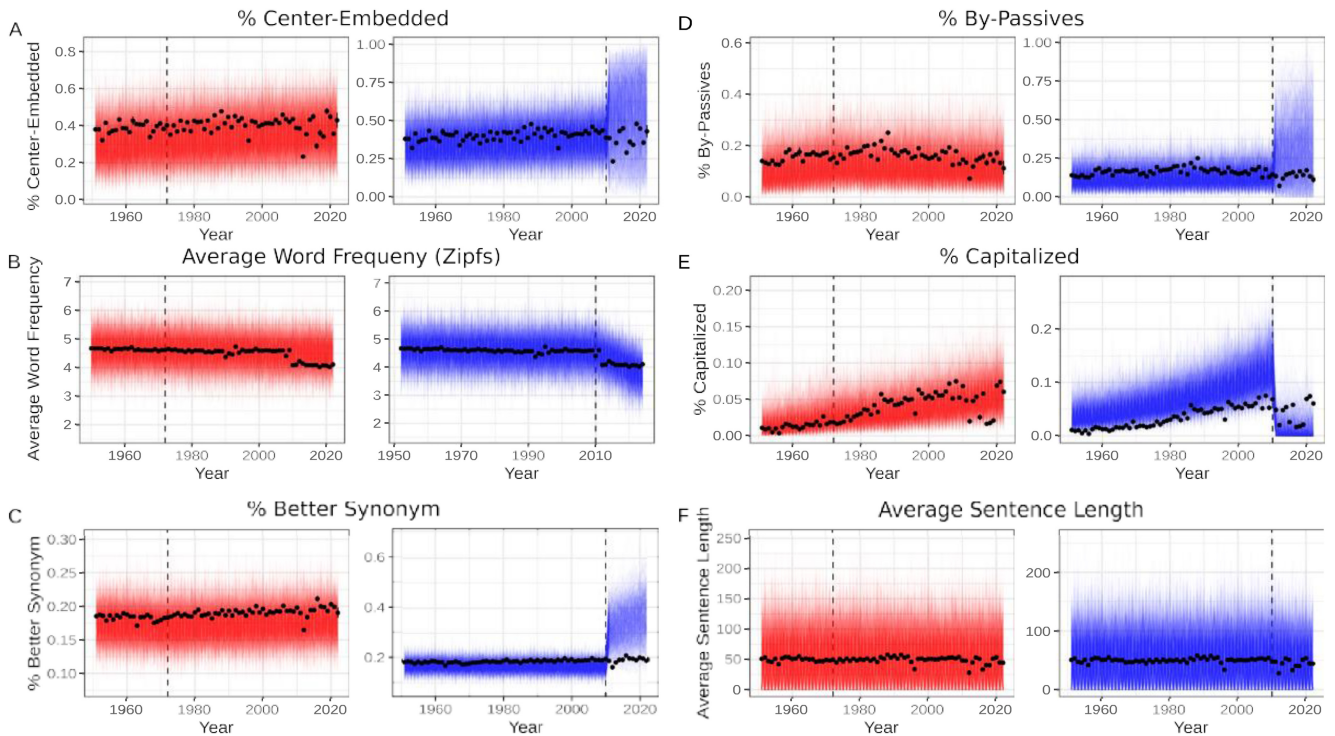
Note. Posterior predictive draws from the break-point regression models, with break-point set at 1972 (left, red [light gray] lines) and 2010 (right, blue [dark gray] lines) for (A) center-embedding, (B) average word frequency (in Zipf scale), (C) word choice ("better synonym"), (D) by-passives, (E) capitalization, and (F) average sentence length. The mean and binomial/bootstrapped confidence intervals for the corpus data are plotted as black points and line-ranges (which are very tight). Note the amount of corpus data aggregated for each time point is variable with higher variance post-2010. Based on Martínez et al. (2022, 2023a, 2023b), the efficacy of the plain-language movement would have material benefits to language processing if there are decreases in center-embedding (A) and word choice (C) and increases in word frequency (B). If the plain-language movement had made an effect at the break point, there would be a significant decrease (increase for word frequency) in the slopes of the best fit model. For 1972, the best fit model is a linear model (not shown; see Table 1); nevertheless, it is visually apparent in the break-point model (left panels, red lines) that there is no significant change in slope after 1972. For 2010, the best fit model is a linear model (not shown) for all predictors but sentence length (see Table 2). Looking at the break-point regression (right panels, blue lines), there is a general increase in variance after the break points most likely due to decreased data amounts. While the break-point model is a better fit for sentence length, the effect size is so small as to be negligible and its significance is driven by the high number of observations. For frequency, by-passive and word choice, the slopes post the 2010 break point are trending toward worse language processing outcomes contra the plain-language movement. For capitalization, the slope post the 2010 break point are trending toward better language processing outcomes, suggesting potential improvement. See the online article for the color version of this figure.

Moreover, to the extent that legal jargon is inevitable, the inaccessibility of legal language would still be problematic even according to the law's own aims, as much of legal doctrine either assumes or requires that laws be accessible to the typical layperson. For example, in U.S. constitutional law, the fair notice doctrine requires "that laws give the person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly" (*Grayned v. City of Rockford*, 1972; Love, 2011; *McBoyle v. United States*, 1931; Robinson, 2005). Insofar as laws are incomprehensible to the typical layperson, this would arguably imply that laws are not giving laypeople fair notice, which would in turn undermine both the constitutionality of those laws and the legitimacy of the fair notice doctrine.

Meanwhile, the ordinary meaning doctrine, which has been referred to as "the most fundamental principle of legal interpretation," not only of U.S. law but of jurisdictions across the world, requires that words in legal documents typically be interpreted according to how they are ordinarily understood by laypeople

(*Moskal v. United States*, 1990; *Richards v. United States*, 1962; Slocum, 2019; Slocum & Wong, 2021; *United States v. Turkette*, 1981). However, insofar as legal documents are not ordinarily understood by laypeople, the coherence and legitimacy of this doctrine would also be undermined.

Aside from the inevitability of legal jargon, another possibility for why lawmakers write the way that they do is that esoteric text arises out a mismatch between the priorities of the writer and reader of a law. If lawmakers' priorities differ from the reader's priorities they may even do this implicitly as opposed to engaging in an outright "conspiracy of gobbledegook" (Mellinkoff, 2004). This possibility seems to have been undercut by recent findings indicating that lawyers, like laypeople, disprefer complicated legalese to simplified legal language when tasked with reading and evaluating legal documents (Martínez et al., 2023a, 2023b).

Another alternative, similar to what has been dubbed the "curse of knowledge" (Hinds, 1999; Nickerson, 1999), is that lawyers may not

realize that their language is too complicated for the average reader to understand (Azuelos-Atias, 2018). Although this hypothesis appears to be supported by previous findings that show an effect of features such as prior knowledge and reading skill on the processing of specialized texts (Cain et al., 2004; Kendeou & Van Den Broek, 2007; Long et al., 2008; Noordman & Vonk, 1992; Ozuru et al., 2009), recent evidence in legal contexts has undercut this hypothesis. In particular, lawyer subjects in Martínez et al. (2023a, 2023b), like laypeople, were found to struggle to understand convoluted legal documents, and were not found to be disproportionately better at understanding convoluted legal documents relative to simplified legal documents compared to laypeople, nor were they found to underestimate the difficulty of convoluted legal documents relative to simplified legal documents.

An additional possibility is that legalese is a result of an iterative drafting process, in which conditions are often thought of after the creation of an initial draft and are more easily embedded within the center of existing sentences as opposed to separated out into a subsequent sentence. If so, this would predict that the complexity of legal language could be alleviated by thinking through the conceptual complexity of a legal document prior to writing as opposed to copying and iteratively editing documents over time.

A final possibility is that lawyers and lawmakers write in a convoluted manner in order to lend official legal documents a ritualistic, spell-like element of authority (cf. Hart & Green, 2012; Tiersma, 1999). If true, this could explain why the plain-language movement might have succeeded in spurring efforts to create unofficial descriptions of laws but not in the simplification of official legal documents such as legislation.

Further work into the plausibility of these hypotheses could yield insight into how best to persuade lawmakers to integrate the findings of our and similar studies and help alleviate the mismatch between the ubiquity and impenetrability of legal texts in the modern era.

Constraints on Generality

The research question of this study related to the accessibility of the federal laws of the United States (a) over time since the onset of the plain-language movement, and (b) relative to other texts applicable to the general population of the United States. The legal materials that we used were an exhaustive set of (a) all federal laws passed by Congress since before the onset of the plain-language movement and May 2022, and (b) all federal laws currently in effect as of 2021. Therefore, we can be confident that our results generalize to the target set of legal documents identified by our research question. Our results also converge with recent findings of the same complex features in other types of legal documents relative to nonlegal documents.

Although the complexity of legalese has been attested in other countries beyond the United States, and although other countries have had similar plain-language efforts, it is unclear whether our results would generalize to laws of other countries and other languages. It is also unclear to what extent these findings generalize to layperson summaries of legal documents within and beyond the United States.

With regard to our nonlegal materials, our sample included a large and wide-ranging set of baseline genres of English that varied in their intended audience and formality. We, therefore, expect that our findings would hold were we to compare legal texts with other baseline genres according to our metrics.

The metrics we looked at are generally considered by plain-language advocates, as well as within the psycholinguistics and reading literatures, as valid proxies for accessibility, and the tools we

used to measure those metrics have been validated as accurate beyond the present study. We can therefore be confident that our analyses reliably assessed the efficacy of the plain-language movement according to its own aims.

That said, it is possible that there are some indices of processing difficulty that we missed. There may be other ways in which laws are more complex than nonlaws, and there may be some ways in which laws are less complex than nonlaws.

Similarly, it is unclear to what extent the psycholinguistic complexity of laws can be dissociated with their conceptual complexity. Previous studies we ran have found that both lawyers and nonlawyers recall and understand more content in legal documents drafted without these features compared to legal documents of equivalent meaning drafted with these features (Martínez et al., 2021, 2022). Similarly, in the present study, we found that laws had higher indices of complex psycholinguistic features than texts of plausibly similar levels of conceptual complexity. However, it is plausible that some degree of psycholinguistic complexity in legal texts is a result of conceptual complexity, and it is unclear to what degree complex psycholinguistic features in legal documents can be removed without leading to a loss or distortion of meaning.

References

Adler, M. (2012). The plain language movement. In L. M. Solan & P. M. Tiersma (Eds.), *The Oxford handbook of language and law* (pp. 67–83). Oxford Academic.

American Law Institute. (1984). *Model penal code*.

American Law Institute and National Conference of Commissioners on Uniform State Laws. (2002). *Uniform Commercial Code (U.C.C.) s 2-216(2))*.

Arbel, Y. A., & Toler, A. (2020). All-caps. *Journal of Empirical Legal Studies*, *17*(4), 862–896. https://doi.org/10.1111/jels.12272

Arsanjani, M. H. (1999). The Rome statute of the international criminal court. *American Journal of International Law*, *93*(1), 22–43. https://doi.org/10.2307/2997954

Azuelos-Atias, S. (2018). Making legal language clear to legal laypersons. In B. Kryk-Kastovsky & D. Kurzon (Eds.), *Legal pragmatics* (Vol. 288, pp. 101–116). John Benjamins. https://doi.org/10.1075/pbns.288.05azu

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671–681. https://doi.org/10.1037/0022-0663.96.4.671

Charrow, R. P., & Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia Law Review*, *79*(7), 1306–1374. https://doi.org/10.2307/1121842

Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word corpus of historical American English. *Corpora*, *7*(2), 121–157. https://doi.org/10.3366/cor.2012.0024

Diamond, S. S., Murphy, B., & Rose, M. R. (2012). The "Kettleful of law" in real jury deliberations: Successes, failures, and next steps. *Northwestern University Law Review*, *106*(4), 1537–1608. https://doi.org/10.2139/ssrn.1641552

Elwork, A., Sales, B. D., & Alfini, J. J. (1982). *Making jury instructions understandable*. Michie Company.

Exec. Order No. 13648. (1979). *44 fr 69609*.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Flesch, R. (2007). *Flesch–Kincaid readability test*. https://rockstar-english.com/lessons/advanced/12-Flesch_Kincaid_Readability_Test.pdf

Garner, B. A., & Black, H. C. (2004). *Black's law dictionary*. Thomson West Publishing Co.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Goldstein, N. E. S., Condie, L. O., Kalbeitzer, R., Osman, D., & Geier, J. L. (2003). Juvenile offenders' Miranda rights comprehension and self-reported likelihood of offering false confessions. *Assessment*, *10*(4), 359–369. https://doi.org/10.1177/1073191103259535

Grayned v. City of Rockford (Vol. 408) (No. 70-5106). (1972). Supreme Court. https://www.law.cornell.edu/supremecourt/text/408/104

Hart, H. L. A., & Green, L. (2012). *The concept of law*. Oxford University Press.

Heuer, L., & Penrod, S. D. (1989). Instructing jurors: A field experiment with written and preliminary instructions. *Law and Human Behavior*, *13*(4), 409–430. https://doi.org/10.1007/BF01056412

Hiltunen, R. P. (2012). The grammar and structure of legal texts. In P. M. Tiersma & L. M. Solan (Eds.), *The Oxford handbook of language and law* (pp. 39–51). Oxford University Press.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, *5*(2), 205–221. https://doi.org/10.1037/1076-898X.5.2.205

Kendeou, P., & Van Den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567–1577. https://doi.org/10.3758/BF03193491

Kurzon, D. (1997). 'Legal language': Varieties, genres, registers, discourses. *International Journal of Applied Linguistics*, *7*(2), 119–139. https://doi.org/10.1111/j.1473-4192.1997.tb00111.x

Library of Congress. (2021). *United States statutes at large*.

Long, D. L., Prat, C., Johns, C., Morris, P., & Jonathan, E. (2008). The importance of knowledge in vivid text memory: An individual-differences investigation of recollection and familiarity. *Psychonomic Bulletin & Review*, *15*(3), 604–609. https://doi.org/10.3758/PBR.15.3.604

Love, J. A. (2011). Fair notice about fair notice. *The Yale Law Journal*, *121*(8), 2395–2403.

Marks, C. B., Doctorow, M. J., & Wittrock, M. C. (1974). Word frequency and reading comprehension. *The Journal of Educational Research*, *67*(6), 259–262. https://doi.org/10.1080/00220671.1974.10884622

Martínez, E., Mollica, F., & Gibson, E. (2022). Poor writing, not specialized concepts, drives processing difficulty in legal language. *Cognition*, *224*, Article 105070. https://doi.org/10.1016/j.cognition.2022.105070

Martínez, E., Mollica, F., & Gibson, E. (2023a). Even lawyers do not like legalese. *Proceedings of the National Academy of Sciences*, *120*(23), Article e2302672120. https://doi.org/10.1073/pnas.2302672120

Martínez, E., Mollica, F., & Gibson, E. (2023b). *Accessibility of legal texts over time*. https://osf.io/ambp4

Martínez, E., Mollica, F., Liu, Y., Podrug, A., & Gibson, E. (2021, Jul 26–29). *What did I sign? A study of the impenetrability of legalese in contracts*. 43rd Annual Conference of the Cognitive Science Society 2021, University of Vienna, Vienna, Austria (Vol. 43, pp. 140–146).

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *46*(5), 1138–1153. https://doi.org/10.1044/1092-4388(2003/089)

Masson, M. E., & Waldron, M. A. (1994). Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Applied Cognitive Psychology*, *8*(1), 67–85. https://doi.org/10.1002/acp.2350080107

Mattila, H. E. (2016). *Comparative legal linguistics: Language of law, Latin and modern Lingua Francas*. Routledge.

Mcboyle v. United States (Vol. 283) (No. 552). (1931). Supreme Court. https://www.law.cornell.edu/supremecourt/text/283/25

Mellinkoff, D. (2004). *The language of the law*. Wipf and Stock.

Miller, G. A. (1995). Wordnet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce (Ed.), *Handbook of mathematical psychology* (pp. 2–419). John Wiley & Sons.

Moskal v. United States (Vol. 498) (No. 89-964). (1990). Supreme Court. https://supreme.justia.com/cases/federal/us/498/103/

Plain Writing Act of 2010. (n.d.). 5 U.S.C. § 301. 2010.

Nickerson, R. S. (1999). How we know—And sometimes misjudge—What others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. https://doi.org/10.1037/0033-2909.125.6.737

Noordman, L. G., & Vonk, W. (1992). Readers' knowledge and the control of inferences in reading. *Language and Cognitive Processes*, *7*(3–4), 373–391. https://doi.org/10.1080/01690969208409392

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Plain Language Action Information Network. (2011). *Federal plain language guidelines*.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). *Stanza: A Python natural language processing toolkit for many human languages*. arXiv preprint. https://arxiv.org/abs/2003.07082.

Rakoff, J. S., Daumier, H., & Case, A. C. (2014, November 20). Why innocent people plead guilty. In *The New York Review of Books*, 20. https://www.nybooks.com/articles/2014/11/20/why-innocent-people-plead-guilty/

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732. https://doi.org/10.1037/0096-1523.30.4.720

Refo, P. L. (2004). The vanishing trial. *Journal of Empirical Legal Studies*, *1*(3), v–vii. https://doi.org/10.1111/j.1740-1461.2004.00029.x

Richards v. United States (Vol. 369) (No. 59). (1962). Supreme Court. https://supreme.justia.com/cases/federal/us/369/1/

Robinson, P. H. (2005). Fair notice and fair adjudication: Two kinds of legality. *University of Pennsylvania Law Review*, *154*(2), 335–398. https://doi.org/10.2307/25047590

Rogers, R., Harrison, K. S., Shuman, D. W., Sewell, K. W., & Hazelwood, L. L. (2007). An analysis of Miranda warnings and waivers: Comprehension and coverage. *Law and Human Behavior*, *31*(2), 177–192. https://doi.org/10.1007/s10979-006-9054-8

Shain, C., Clara, M., Tiago, P., Cotterell, R., & Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences*, *121*(10), 1–12. https://doi.org/10.1073/pnas.2307876121

Slocum, B. G. (2019). *Ordinary meaning: A theory of the most fundamental principle of legal interpretation*. University of Chicago Press.

Slocum, B. G., & Wong, J. (2021). The Vienna Convention and ordinary meaning in international law. *The Yale Journal of International Law*, *46*(2), 191–239.

Solnyshkina, M., Zamaletdinov, R., Gorodetskaya, L., & Gabitov, A. (2017). Evaluating text complexity and flesch-kincaid grade level. *Journal of Social Studies Education Research*, *8*(3), 238–248.

Tiersma, P. M. (1999). *Legal language*. University of Chicago Press.

Tobia, K. P. (2020, February 10). Legal concepts and legal expertise. *Synthese*. Advance online publication. https://doi.org/10.2139/ssrn.3536564

United States v. Turkette (Vol. 452) (No. 80-808). (1981). Supreme Court. https://supreme.justia.com/cases/federal/us/452/576/

United States Code. (2021). U.S. Government Publishing Office. https://www.govinfo.gov/app/collection/uscode/2021

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!