

What did I sign? A study of the impenetrability of legalese in contracts

Eric Martínez (ericmart@mit.edu)

Brain & Cognitive Sciences, MIT
Cambridge, USA

Francis Mollica (mollicaf@gmail.com)

School of Informatics, University of Edinburgh
Edinburgh, UK

Yufei Liu (yl2@wellesley.edu)

Psychology, Wellesley College
Wellesley, USA

Anita Podrug (apodrug@mit.edu)

Electrical Engineering & Computer Science, MIT
Cambridge, USA

Edward Gibson (egibson@mit.edu)

Brain & Cognitive Sciences, MIT
Cambridge, USA

Abstract

Legal documents, in the form of terms of service agreements and other private contracts, are now an increasingly prevalent part of everyday life. While legal documents have long been acknowledged to be difficult to understand without training, it remains an open question whether the ever-increasing exposure to contracts might have mitigated this difficulty. Moreover, insofar as this difficulty has persisted, there remains no systematic analysis of which linguistic structures contribute most heavily to the processing difficulty of legal texts, nor whether this difficulty is heightened for those with less language experience. Here, we investigate these issues, and in a well-powered experiment find evidence that (a) both recall and comprehension of legal propositions in a contract are hindered by use of a legal register relative to plain-English translations; (b) certain linguistic structures, such as center-embedding, hinder recall to a greater degree than others, such as passive voice; and (c) language experience influences comprehension of legal propositions. Surprisingly, language experience did not influence recall, nor was there an interaction between legal register and language experience on recall or comprehension. These findings suggest that legal language poses heightened difficulties for those with less language experience—who tend to be of lower socioeconomic status and with diminished access to the justice system—and that eliminating complex features of legalese would benefit those of all reading levels.

Keywords: legal language; language processing; memory

The mismatch between the ubiquity and impenetrability of legal documents has been a long-standing and seemingly intractable issue. Dating as far back to the plain language movement in the 1970s, government officials have acknowledged the need to simplify public legal documents for the benefit of society at large. Since then, there has been a sizeable literature exploring how to best simplify public-facing legal language, such as jury instructions (Charrow & Charrow, 1979; Elwork, Sales, & Alfini, 1982; Heuer & Penrod, 1989; Diamond, Murphy, & Rose, 2012; Randall, 2019) and

Miranda warnings (Goldstein, Condie, Kalbeitzner, Osman, & Geier, 2003; Rogers, Harrison, Shuman, Sewell, & Hazelwood, 2007).

While these studies have successfully demonstrated the efficacy of identifying and replacing problematic features of legal text (such as archaic legal jargon and complex syntax) with “plain English” equivalents to increase comprehension rates among laypeople, they only apply to a very small portion of the total corpus of legal language. 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: Refo, 2004; Rakoff, Daumier, & Case, 2014). The majority of individuals’ contact with legal language takes place outside the context of criminal or civil suits and involves more than just public-facing documents, such as contracts and other private-facing documents.

In the case of private legal documents, the literature in this regard has been more scarce. One notable exception to this is Masson and Waldron (1994), which investigated how clerical staff ($n = 24$) and age-matched students ($n = 24$) recalled and comprehended aspects of four short (≈ 400 words) legal contract texts. Masson and Waldron (1994) found that less complex—“plain English”—texts were recalled and comprehended better than more complex actual legal texts.

In the decades since these findings, contracts have become an increasingly ubiquitous part of the modern era, particularly with the rise of the internet and the constant exposure to online terms of service agreements. It remains an open question whether this increased exposure might have mitigated the difficulty of reading legal texts.

While the research on processing difficulties in legal language has been sparse, our understanding of language pro-

cessing difficulties in terms of the surface properties of text has greatly improved. As it turns out, several structures that have been identified as difficult to process in the psycholinguistic literature (e.g. clausal embedding: Gibson, 1998; Pinker, 2003; low-frequency words: Rayner, Ashby, Pollatsek, & Reichle, 2004; passive voice: Ferreira, 2003) are reputedly abundant in legal texts (e.g. P. M. Tiersma, 1993; P. Tiersma, 2008), and some are especially prevalent in contracts (e.g. passive voice: Goźdz-Roszkowski, 2011; all-caps: Arbel & Toler, 2020).

Here we seek to build upon these findings, as well as those of Masson and Waldron (1994), to better understand to what extent (a) contractual language continues to present difficulties in recall and comprehension (both overall, and relative to a simpler alternative), (b) such difficulties are mitigated or exacerbated by one's experience and/or skill with general reading ability, and (c) particular linguistic features give rise to this difficulty in the first place.

To that end, we conducted a high powered conceptual replication of Masson and Waldron (1994) where we constructed short legal contracts that varied linguistic structures known to incur processing difficulties. We also implemented the author recognition task (ART; Stanovich & West, 1989; Moore & Gordon, 2015), which measures individual differences in experience with language. We replicate Masson and Waldron (1994)'s main finding that information presented in a legal register is recalled and recognized less frequently and accurately than when presented in simple language. Further, we find that comprehension of legal language in general is related to experience with language. Lastly, we provide a descriptive analysis of how individual linguistic structures relate to the difference in performance across legal and simple registers.

Methods

Materials

Our primary materials consisted of 12 pairs of short contract excerpts of roughly 150 words each (see Figure 1). First, twelve excerpts were constructed in a standard "legalese" register by the first author, a lawyer, who modeled the content and form of the materials after common naturalistic contracts. Each of the 12 texts for each condition corresponded to one of three types of common contract provisions (with four texts pertaining to each genre), including: (1) general contract provisions, specifying the basic terms of a contractual agreement; (2) liability and warranty provisions, specifying to what degree each party could be sued or held accountable for not adhering to the terms of the agreement; and (3) jurisdiction, venue and choice-of-law provisions, specifying how and where parties could sue or be held accountable for not fulfilling the terms of the agreement. Each legalese text was drafted to contain the following language properties that have been identified as difficult to process and common to legal texts:

(a) *Low-frequency legal terms*—Words that are infrequently used in everyday speech provide processing difficulties for

readers relative to higher-frequency synonyms (Marks, Doctorow, & Wittrock, 1974). Legal texts are laden with "archaic words" such as *aforesaid*, *herein*, and *to wit* (P. Tiersma, 2008), which have been shown to be frequently misunderstood by laypeople (e.g. P. M. Tiersma, 1993). Each legalese text was constructed to contain several instances of legal jargon, which were replaced with high-frequency synonyms in the plain-English versions.

(b) *Center-embedded clauses*—Center-embedded structures have long been observed to pose processing difficulties on a reader (Miller & Chomsky, 1963; Gibson, 1998; Pinker, 2003). The tendency for lawyers to "embed" legal jargon "in convoluted syntax" has been observed not only to be prevalent in legal texts but as a potential badge of honor for those who wish to "talk like a lawyer" and be accepted by their profession (P. Tiersma, 2008). Each legalese text was constructed to contain multiple center-embedded clauses ("*Artist and Tour, said parties being hereinafter referred as...*"), which were written as separate sentences in the corresponding plain-English version.

(c) *Passive-voice structures*—Relative to their active-voice counterparts, passive-voice structures are acquired later by children (Baldie, 1976), and may continue to pose difficulties for adults (Ferreira, 2003). Goźdz-Roszkowski (2011) found passive structures to be more prevalent in contracts relative to other legal and non-legal genres (such as newspapers). Our legalese texts each contained multiple passive-voice structures ("*This agreement has been formed by the parties*"), which we converted into active-voice structures in the corresponding plain-English versions.

(d) *Capitalization*—Non-standard capitalization is ubiquitous in provisions such as warranty disclaimers and limitations of liability, which "must be conspicuous" in order to be legally upheld (American Law Institute and National Conference of Commissioners on Uniform State Laws, 2002). Arbel and Toler (2020) found that most standard form agreements used by major companies contain a provision in all-caps. Although the use of all-caps provisions is ostensibly for the benefit of the reader, evidence suggests that they do not aid comprehension (Arbel & Toler, 2020). Here we included at least one chunk of all-capitalized text in each legalese passage ("*THE WARRANTY IS HEREBY DISCLAIMED*"), which was replaced with standard capitalization in the simple version.

From the set of legalese materials, each passage was encoded in terms of legally relevant propositions. From these propositions, each passage was then translated into a "plain-English" version, which differed only with respect to the four surface properties described above, resulting in 24 total passages.

For each contract pair, 12-15 comprehension questions were drafted. The questions were multiple choice with four options. These questions both targeted comprehension of specific important legal propositions, as well as more general understanding of the legal content. To reduce a response bias for a given register, we controlled the overlap in form be-

It is understood by *Lessees*, standing liable for violating *obligations inter se, hereinbefore set forth* in Clause 3 of this real estate agreement, that **Lessor**s shall be exempt from liability for any damages, to the **maximum extent not prohibited** by law, unless **Lessor**s knew of the possibility of such damage and acted **with scienter**. LESSEES' AGGREGATE LIABILITY INTER SE FOR ALL CLAIMS, INCLUDING THOSE BASED ON TORT OR STATUTORY LIABILITY, IS LIMITED TO \$1000 BY THIS AGREEMENT. PERSONAL INJURY DAMAGES, LIMITATIONS OF LIABILITY OF WHICH, **EXCEPTING** THOSE FOR EMOTIONAL DISTRESS, **THIS JURISDICTION PROHIBITS, ARE NOT AFFECTED BY THE FOREGOING PROVISIONS.**

Tenants understand that Landlords will be exempt from liability for any damages, to the **extent allowed** by law, unless **Landlords** knew of the possibility of such damage and acted **willfully**. *Tenants will be liable for violating their duties to each other, described in Clause 3 of this real estate agreement, above.* This agreement limits **Tenants' combined liability to each other** for all claims to \$1000. This includes claims based off tort or statutory liability. This jurisdiction prohibits limitations of liability of personal injury damages, **except for emotional distress damages.** The above section does not affect personal injury damages.

Figure 1: An example stimulus pair in legalese (left) and simple (right) register. The differences in surface properties across registers are depicted by font style. Bold denotes word frequency. Italic denotes embedded clauses. Underlined denotes voice. Unfortunately, we have run out of font styles to make differences in capitalization more apparent.

tween contract excerpt and comprehension question. Both types of comprehension question were drafted in a “neutral” register. Passive/active structures were replaced by nominalizations. For example, “*shipment of the goods on the part of merchant*” instead of “*the goods were shipped by merchant*” or “*merchant shipped the goods*”). High or low frequency synonyms were replaced with a third synonym (e.g. “*renter*” instead of “*lessee*” or “*tenant*”).

In addition to our main experimental materials, we administered the Author Recognition Task (ART; Moore & Gordon, 2015) as a proxy for language experience.

Participants and Procedure

Based on a pilot study ($n = 32$), we found that 100 participants would provide us sufficient power ($> 80\%$) to detect our main effects in both recall and comprehension. Due to concerns about data validity with online collection, we actively assessed data quality during the experiment. Participants first completed three trials and were only allowed to complete the experiment if their comprehension was above chance performance. In total we recruited 186 participants for the first half, but we only retain 108 participants who completed the entire experiment for our analysis. All participants self-identified as native English users.

Retained participants were pseudorandomly assigned to six trials (3 legalese; 3 simple). Participants did not see the same contract in both a simple and legal register. Assignment of stimuli to participant was pseudorandom to ensure that across participants every trial was administered with approximately the same frequency. The order of trials was randomized for each participant.

A trial consisted of (a) reading an excerpt, (b) a subset of the ART, (c) recalling the excerpt, and (d) answering comprehension questions. For the reading component, participants were presented with exactly one excerpt, written in either legalese or plain English. They were asked to carefully read the text twice, and were given as much time as needed to do so. For the ART component, participants were given the names of 50 individuals and were asked to select which

names corresponded to real authors. The actual ART is 130 items but we needed 300. Therefore, the items were from the actual ART for the first three trials. For the remaining trials, the participants were administered “filler” items that looked virtually the same as authentic materials (half of the names corresponding to real authors, the other half corresponding to high-school track stars) but which had not been validated the same way. We do not use these fillers in our analysis. After being shown the ART materials, participants were asked to recall as much of the excerpt they had read as possible. They were told that they could use their own words, but that their version should stay true to the original. Finally, each trial ended with the comprehension questions corresponding to the excerpt.

Analysis Plan

Two trained research assistants coded whether a proposition was successfully recalled. In doing so, they were presented with a participant’s retelling of a passage and then asked whether each legally relevant proposition of the passage was (a) fully recalled; (b) partially recalled; or (c) not recalled. Coders were told that for a response to count as “fully recalled,” it did not have to be recalled verbatim (i.e. they can use their own words or syntax), so long as they were confident that the meaning of what subject wrote is the same as the proposition.

For example, suppose the original text said “A court in Boston will resolve the dispute,” and the participant wrote “something will be resolved by a court.” When coding responses, a coder might see three propositions that say: (i) “A court in Boston,” (ii) “will resolve,” and (iii) “the dispute.”

For (i), the coder would put a 0.5 for “partially recalled” (since “in Boston” was missing from “a court”); for (ii), the coder would put a 1 for “fully recalled” (since “will be resolved” means basically the same thing as “will resolve”), and for (iii), the coder would put a 0 for “not recalled” (since “the dispute” was not in the response).

To reduce potential bias, coders were unaware of whether a participant had seen or recalled the simple or legalese ver-

sion of a text. Of the roughly 650 retellings, each coder was responsible for coding roughly 60 percent (≈ 390) of the retellings, such that (a) every retelling/proposition would be coded at least once, and (b) 20% of the retellings would be coded by both coders so as to assess inter-rater reliability. Coder reliability was assessed with Cohen's kappa coefficient (Cohen, 1960; McHugh, 2012).

We adjudicated ties as follows: (i) a tie between one "fully recalled" judgment and one "not recalled" judgment resulted in a final "partially recalled" judgment; (ii) a tie between one "fully recalled" judgment and one "partially recalled" judgment resulted in a final "fully recalled" judgment for a given proposition; and (iii) a tie between one "partially recalled" judgment and one "not recalled" judgment resulted in a final "not recalled" judgment. For our regression analyses, we perform both a conservative analysis (recoding "partially recalled" as "not recalled") and an anti-conservative analysis (recoding "partially recalled" as "fully recalled"). Our results do not qualitatively change, so we will only report the conservative analysis here.

Results

We set out to address three questions: 1) can we replicate Masson and Waldron (1994)'s result that comprehension and recall of contracts presented in a legal register is worse than that of contracts presented in a simple register? 2) If so, is the decline in performance worse for people with less written language experience? 3) How do individual linguistic structures purportedly prevalent in legal text contribute to the decline in recall of legal propositions? We will take the first two questions in turn for comprehension and recall. Then, we will provide an exploratory description to address the third.

Comprehension

Following Masson and Waldron (1994), we predict comprehension will be greater for text presented in the simple register than in the legalese register. We additionally predicted that the effect of register will be larger for participants who scored low on the ART task. Figure 2a illustrates the comprehension accuracy across registers and Figure 2b depicts comprehension accuracy as a function of ART score. Descriptively, participants were more accurate in the simple register (73.5%) than in the legalese register (67.7%).

We first conducted a mixed effect logistic regression, with register, standardized ART score and their interaction as fixed effects and comprehension question, excerpt, and participant as random effects, each with a random slope for register. Using likelihood ratio test to compare to a model without the interaction term, we found no significant interaction between standardized ART score and register. Therefore, we report the results of the model fit without the interaction term. Replicating Masson and Waldron (1994), we find a significant decrease in comprehension accuracy for a legalese register compared to a simple register ($\beta = -0.179$, $SE = 0.052$, $p < 0.05$). While we did not find an interaction between language experience and register, we do find that participants with less

language experience (lower ART scores) have worse comprehension accuracy than participants with more language experience ($\beta = 0.219$, $SE = 0.076$, $p < 0.05$).

Recall of Legal Propositions

Before we analyse our predictions, we assessed coder agreement. Our two coders agreed on approximately 85% of overlapping judgments. Cohen's Kappa (unweighted) was measured to be 0.719 ($z = 47.1$; $p < 0.05$), indicating substantial agreement.

Again following Masson and Waldron (1994), we expect participants to recall fewer legal propositions when the text is presented in a legal register compared to a simple register. We further predicted an interaction with language experience, such that recall would be worse for people with less language experience. Figure 2c displays the proportion of propositions recalled across registers and Figure 2d renders the proportion recalled as a function of ART score. Overall, the average recall among participants was 41.1%, which is slightly better than recall rates for previous studies using text materials but a longer delay (Bergman & Roediger, 1999). Descriptively, propositions from excerpts in a simple register (45.9%) were recalled more than propositions presented in a legalese register (38.4%).

As for comprehension, we first conducted a mixed effect logistic regression with register, standardized ART score and their interaction as fixed effects and excerpt and participant as random effects with register as a random slope for each. Using likelihood ratio tests, we fail to find a significant interaction and, thus, report here a simpler model without the interaction term. Again replicating Masson and Waldron (1994), fewer legal propositions were recalled when they were presented in a legalese register compared to a simple register ($\beta = -0.219$, $SE = 0.081$, $p < 0.05$). Unlike our comprehension results, we do not find an effect of language experience on recall.

Exploring the effect of linguistic structures

While surface properties of a text seem to be forgotten relatively quickly (e.g., within an hour; Fisher & Radvansky, 2018) compared to propositional content (lasting weeks), it seems intuitive that they might appreciably influence memory for more abstract representations of content. If a reader can't understand or mis-parses the text, it's highly unlikely that they make the intended inferences and have a full grasp of the situation. Therefore, we expect linguistic structures known to incur processing difficulties to reduce the proportion of legal propositions recalled. Here, we focused on four kinds of structures purportedly common in a legal register and manipulated in our materials: embedded clauses, passive voice, frequency of lexical choice and capitalization. We do not have sufficient power to assess the generalizability of each structure's influence on recall. Instead, we provide a description of each structure's effect using Bayesian mixed effect logistic regression.

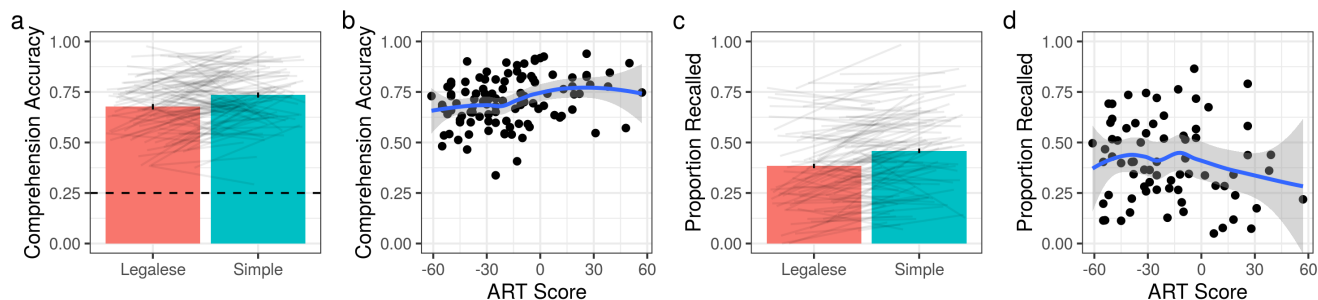


Figure 2: (a) Mean comprehension accuracy across registers. (b) Comprehension accuracy as a function ART score. (c) Mean proportion of legal propositions recalled across registers. (d) Proportion of legal propositions recalled as a function of ART score. In (a) and (c), lineranges reflect 95% bootstrapped confidence intervals and faded lines reflect individual subject averages. In (b) and (d), points reflect individual participant proportions.

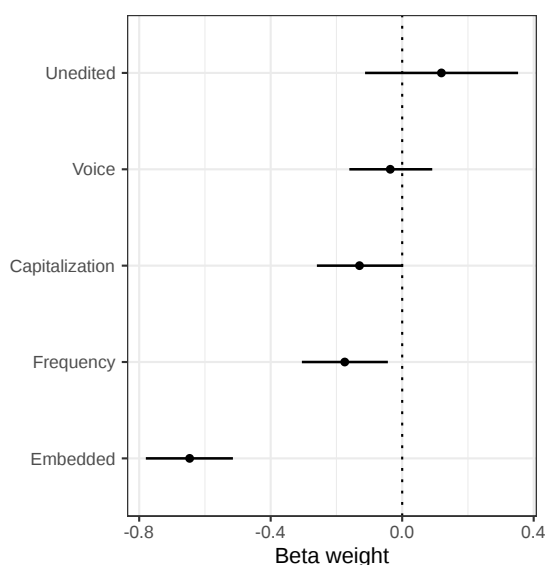


Figure 3: The influence of each surface property on the effect of register. Line ranges reflect the 95% credible intervals over the interaction term and points reflect medians.

For every proposition, we coded if the surface form across registers was unedited or differed in clausal embedding, voice, word frequency or capitalization. We conducted a mixed effect logistic regression predicting recall with surface form as a fixed effect and random intercepts for excerpt and participant. Our fixed effect was coded so that each coefficient reflects either an increase or decrease in recall rate for a legalese register relative to the average recall rate of a simple register. Figure 3 represents the 95% credible interval over the regression coefficient for each surface property. As a sanity check, the propositions that were unedited across register do not differ in rate of recall. We find the strongest effects of register for propositions that differed in embedding or frequency, a smaller effect of capitalization and perhaps surprisingly, no effect for voice.

Discussion

Our study aimed to assess whether (a) contractual legal language continues to present difficulties in recall and comprehension (both overall, and relative to a simpler alternative), (b) such difficulties are mitigated or exacerbated by one's experience and/or skill with general reading ability, and (c) particular syntactic or semantic features give rise to this difficulty in the first place. Here we discuss in turn the extent to which our results successfully answer these questions.

With regard to (a), we replicate Masson and Waldron (1994)'s findings that redrafting legal texts into a simpler register can lead to increases in both comprehension and recall over that of the original text. This suggests that the complex language used in contracts continues to present difficulties to one's ability to absorb and remember the content in such documents, despite the increased exposure to these documents over the last couple of decades.

With regard to (b), our results extend the findings of (Masson & Waldron, 1994) by revealing that language experience—as measured by ART—predicted comprehension performance, suggesting that those with less language experience have a harder time understanding legal texts. Given that those with less reading experience as a group tend to be of lower socioeconomic status (Kieffer, 2010; Bradley & Corwyn, 2002), and those of lower SES face greater disenfranchisement from the legal system (Legal Services Corporation, 2017), this suggests that simplifying contracts may have non-trivial access to justice implications, particularly as their prevalence increases.

At the same time, there was no correlation between ART and recall performance, nor was there a significant interaction between register and performance on ART in predicting recall or comprehension; that is, those who scored higher on the ART did not have a necessarily easier time on the legalese texts relative to the plain English texts, nor did they have an easier time recalling texts of either register relative to poorer readers. Taken together, our results suggest that redrafting texts into a simpler register has beneficial effects on comprehension for readers of all reading levels.

With regard to (c), our results suggest that some of the fea-

tures that are common to legal texts—such as clausal embedding and low-frequency words—present greater difficulties in the context of recall than others, such as passive voice, suggesting that lawyers interested in simplifying legal texts for the benefit of readers ought to prioritize unpacking clauses into separate sentences and opting for higher frequency synonyms when possible.

Having established that the average person's processing difficulties with legal text remain despite the presumed uptick in experience, we should aim to understand why lawyers choose to write in such an esoteric manner in the first place? One possibility is that legal language must be written so to maintain communicative precision. This possibility is partially ruled out by our results and previous findings that show comprehension of legal content with a simplified register (e.g., Masson & Waldron, 1994). While it seems entirely plausible that certain legal jargon is inevitable, our results demonstrate that in many instances such jargon can be replaced with simpler alternatives that increase recall and comprehension while preserving meaning.

Another possibility is that lawyers choose to write in a complex manner to convey their priorities. For example, if a lawyer prioritizes the user's responsibilities they may focus on making them clear at the expense of other content (e.g., company's obligations). If the lawyer's priorities differ from the reader's priorities they may even do this implicitly as opposed to engaging in an outright "conspiracy of gobbledygook" (Mellinkoff, 2004)

Lastly, lawyers may not *choose* to write in an esoteric manner. Similar to the "curse of knowledge" (Hinds, 1999; Nickerson, 1999), they may not realize that their language is too complicated for the average reader to understand (Azuelos-Atias, 2018). Further work into the plausibility of these hypotheses could yield insight into how best to persuade lawyers to integrate the findings of our and similar studies and help alleviate the growing mismatch between the ubiquity and impenetrability of legal texts in the modern era.

Acknowledgments

We would like to thank Alex Paunov, Anya Ivanova, Giuseppe Ricciardi, Rachel Ryskin and Ray Jackendoff for their extremely helpful comments at various stages of this paper.

References

- 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.
- Azuelos-Atias, S. (2018). Making legal language clear to legal laypersons. *Legal Pragmatics*, 288, 101.
- Baldie, B. J. (1976). The acquisition of the passive voice. *Journal of Child Language*, 3(3), 331–348.
- Bergman, E. T., & Roediger, H. L. (1999). Can bartlett's repeated reproduction experiments be replicated? *Memory & Cognition*, 27(6), 937–947.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology*, 53(1), 371–399.
- Charrow, R. P., & Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. *Columbia law review*, 79(7), 1306–1374.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- Diamond, S. S., Murphy, B., & Rose, M. R. (2012). The ket-tleful of law in real jury deliberations: Successes, failures, and next steps. *Nw. UL Rev.*, 106, 1537.
- Elwork, A., Sales, B. D., & Alfini, J. J. (1982). *Making jury instructions understandable*. Michie Company.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, 47(2), 164–203.
- Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language*, 102, 130–141.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Goldstein, N. E. S., Condie, L. O., Kalbeitzner, 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.
- Goźdz-Roszkowski, S. (2011). *Patterns of linguistic variation in american legal english: A corpus-based study*. Peter Lang Frankfurt am Main.
- Heuer, L., & Penrod, S. D. (1989). Instructing jurors. *Law and Human Behavior*, 13(4), 409–430.
- 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.
- Kieffer, M. J. (2010). Socioeconomic status, english proficiency, and late-emerging reading difficulties. *Educational Researcher*, 39(6), 484–486.
- Legal Services Corporation. (2017). *Justice gap report: Measuring the civil needs of low-income americans*.
- Marks, C. B., Doctorow, M. J., & Wittrock, M. C. (1974). Word frequency and reading comprehension¹. *The Journal of Educational Research*, 67(6), 259–262.
- 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.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.
- Mellinkoff, D. (2004). *The language of the law*. Wipf and Stock Publishers.
- Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In *Handbook of mathematical psychology* (pp. 2–419). John Wiley & Sons.

- Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior research methods*, 47(4), 1095–1109.
- 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.
- Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin UK.
- Rakoff, J. S., Daumier, H., & Case, A. C. (2014). Why innocent people plead guilty. *The New York Review of Books*, 20.
- Randall, J. (2019). How just is justice? ask a psycholinguist. In *Grammatical approaches to language processing* (pp. 275–308). Springer.
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the ez reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 720.
- Refo, P. L. (2004). The vanishing trial. *Journal of Empirical Legal Studies*, 1(3), v–vii.
- 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.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 402–433.
- Tiersma, P. (2008). The nature of legal language. In *Aila applied linguistics series: Vol. 5. dimensions of forensic linguistics* (p. 7-25). John Benjamins Publishing Company.
- Tiersma, P. M. (1993). Reforming the language of jury instructions. *Hofstra L. Rev.*, 22, 37.