Mack, Kristen, and John Palfrey. 2020. "Capitalizing Black and White: Grammatical Justice and Equity." *MacArthur Foundation*, August 26, 2020. https://www.macfound.org/press/perspectives/capitalizing-black-and-white-grammatical-justice-and-equity.

MacLean, Lauren M., Elliot Posner, Susan Thomson and Elisabeth Jean Wood (2021). "Research Ethics and Human Subjects: A Reflexive Openness Approach." In The Qualitative Transparency Deliberations: Insights and Implications (QTD Steering Committee report, eds. Alan Jacobs and Tim Büthe. *Perspectives on Politics*. 19, no.1 (March): 188-189.

Mallon, Florencia E., ed. 2011. *Decolonizing Native Histories: Collaboration, Knowledge, and Language in the Americas.* Durham: Duke University Press.

Norton, Jack. 1979. *Genocide of Northwestern California: When Our Worlds Cried.* San Francisco: Indian Historian Press.

Proudfit, Joely, and Nicole Quinderro Myers-Lim, eds. 2017. *On Indian Ground: A Return to Indigenous Knowledge—Generating Hope, Leadership and Sovereignty through Education.* Charlotte: Information Age Publishing.

Sheoran Appleton, Nayantara. 2019. "Do Not 'Decolonize' . . . If You Are Not Decolonizing: Progressive Language and Planning Beyond a Hollow Academic Rebranding." *Critical Ethnic Studies* (blog) February 4, 2019. http://www.criticalethnicstudiesjournal.org/blog/2019/1/21/do-not-decolonize-if-you-are-not-decolonizing-alternate-language-to-navigate-desires-for-progressive-academia-6y5sg

Thomson, Susan. 2021. "Reflexive Openness as Ethical Research Practice." *PS: Political Science & Politics* 54, no. 3 (July): 530-4. https://doi.org/10.1017/S1049096521000159.

Tuck, Eve, and Wayne K. Yang. 2012. "Decolonization is Not a Metaphor." *Decolonization: Indigeneity, Education & Society* 1 (1): 1-40. https://jps.library.utoronto.ca/index.php/des/article/view/18630.

Tuhiwai Smith, Linda. 2012. *Decolonizing Methodologies: Research and Indigenous Peoples.* New York: Zed Books.

Wilson, Shawn. 2008. *Research is Ceremony: Indigenous Research Methods.* Winnipeg: Fernwood Publishing.

Overcoming the Laws-in-Translation Problem: Comparing Techniques for Translating Legal Texts

Anthony J. DeMattee
Emory University

Takumi Shibaike
University of Calgary

Nick Gertler
Concordia University Montreal

Elizabeth A. Bloodgood
Concordia University, Montreal

Statutes and regulatory rules (henceforth "legal text") are invaluable forms of primary data for comparative analysis of regime transitions, political participation, policy diffusion, and the realization of justice and rights. Governments carefully curate and archive their legal codes, making it possible to trace a law's history and track its diffusion to other jurisdictions. Governments do not consistently enforce laws as they are written (Pound 1910; Law and Versteeg 2013); to understand the distinction between laws in the books and laws in action, we first must be able to read the law. Legal texts are thus invaluable for historical analysis and comparative studies, as laws in one jurisdiction can have important similarities and differences across national contexts (Glasius, Schalk, and De Lange 2020; Hummel, Gerring, and Burt 2021; Berinzon and Briggs 2019).

Despite their importance and accessibility, legal texts are not studied in a comparative perspective as frequently as quantitative indicators because of the limitations of language. The European Union and United Nations are

leading efforts to collect and translate legal texts, but those collections are not comprehensive across topics, countries, or time (Giampieri 2016). We refer to this as the "laws-in-translation problem," a term that we use to refer to the fact that legal translations are scarce or incomplete, and available strategies to translate such texts are either cost-prohibitive or error-prone to an unknown degree.

Researchers can deploy solutions to the laws-in-translation problem, but each presents new issues. Professional human translators remain the "gold standard" for accuracy (Lucas et al. 2015, 259-60), but human translation services quickly become expensive for large legal codes. Alternatively, machine translation (MT) software provides speed, accessibility, and affordability, but there are concerns about their accuracy and reliability. We evaluate DeepL, Google, and Microsoft MT applications for their effectiveness in translating legal texts on five dimensions: generalizability, flexibility, presentation, simplicity, and reliability. For a controlled comparison of machine and human translations, we translate legal content from similar laws enacted in Brazil, China, France, Japan, and Mexico. We find that MT tools are not sufficiently nuanced for legal practice or fine-grained analysis, but maintain that MT's accuracy—when used in a hybrid approach—is sufficient for researchers conducting comparative socio-legal and policy research. Our recommendation to those who seek accuracy and cost-efficiency is to use MT applications in tandem with human translators. As we show in the following sections, combining these translation strategies increases transparency and accuracy while lowering costs and decreasing time spent.

Our hybrid approach to translating legal texts razes methodological barriers and expands the number of cases available for comparative analysis. The method has implications beyond law and policy scholars. The language in which a government writes its laws and executive orders has no relationship with whether it follows, bends, or violates those legal rules. Whether legal texts appear in Russian or English matters little if the government enforcing those rules seeks to threaten fundamental freedoms, undermine elections, or loot personal property. Yet language barriers profoundly hinder our ability to evaluate whether laws authorize illiberal practices, disguise them as legitimate actions, or simply fail to prevent such phenomena. Overcoming these barriers allows us to read local reports covering protests in foreign capitals and access the ideas and reactions of those affected by war, migration, or disaster.

Motivation: The Missing Standard for Translating Legal Texts

The social sciences contain numerous examples of scholars who have successfully navigated the laws-in-translation problem. Unfortunately, published research often lacks a clear explanation of *how* the translation process unfolded. We are guilty of this ourselves (DeMattee 2022a; Bloodgood, Tremblay-Boire, and Prakash 2014). For research covering a multilingual legal corpus, readers must assume that the researchers are polyglots or that translation is a rigorous part of data collection and research design phases. Not disclosing the translation process is not the same as being unwilling to be transparent. The publishing process has limited authors' ability to communicate how they translated texts and verified translations. Innovative technology such as the Qualitative Data Repository's Annotation for Transparent Inquiry (ATI) eases some limitations. ATI enables researchers to digitally link their article to analytic notes, allowing authors to provide more information about the data and analytic choices, including extended excerpts of original and translated legal text (Kapiszewski and Karcher 2021; Elman, Kapiszewski, and Lupia 2018). Clear explanations and standards for the translation of laws increase the replicability and credibility of findings. Reviewing sociology, political science, and law and society journals reveals a surprising lack of comparative law and policy research across geographic regions and languages. This research gap is likely related to a missing process for effective and efficient translation. Solving the laws-in-translation problem is thus likely to open opportunities for research into new questions as well as new approaches to old puzzles.

One approach is to limit case selection to regions whose countries publish laws in a common language. Scholars have successfully compared the development of penal codes in French West African countries (Berinzon and Briggs 2019) and corporate law in China, Hong Kong, and Taiwan (Lin and Chang 2018). Similarly, Lemon and Antonov (2020) compare legal text in five post-Soviet countries to show that independent countries enact laws with nearly identical language. Researchers can encounter the laws-in-translation problem even when maintaining a regional focus. For example, when studying immigration laws in the Americas, Cook-Martín and FitzGerald (2019) needed to analyze a multilingual corpus that included countries colonized by Britain, France, Spain, and Portugal. While researchers might need to limit their scope to monolingual corpora or regional analyses if a deep and nuanced interpretation of the particular legal text is necessary, scholars who focus on global phenomena, including diffusion effects

and postcolonial change, need competencies in multiple languages or an alternative approach.

Secondary sources offer scholars one solution to expand their geographic and linguistic range. They provide a valuable foundation, and merging multiple sources can improve comprehensiveness. Glasius, Schalk, and De Lange (2020, 457) and Hummel, Gerring, and Burt (2021, 873) assemble their corpora from various public, private, and academic sources. Secondary sources have hidden costs as they are neither as comprehensive nor as accurate as researchers often require, especially if secondary sources are limited to specific languages or periods. Echoing previous caveats of off-the-shelf data (Bennett 2007), to what degree can researchers trust secondary sources' accuracy, rigor, and objectivity? The remedy suggested still applies: "Take between five and ten random observations from the dataset and attempt to code the variables *from the ground up*" (Goemans 2007, 12; emphasis added). This prescription to recode a random sample of legal texts returns us to the laws-in-translation problem.

Brute force is another pathway to obtaining a large, multilingual corpus. With adequate resources, researchers can collect, translate, and code primary sources from the ground up. With financial commitments from Google, the National Science Foundation, and the United States Institute of Peace, the Comparative Constitutions Project (Elkins and Ginsburg 2021; Elkins, Ginsburg, and Melton 2009) is the exemplar for overcoming the laws-in-translation problem. Bradford et al. (2019, 416) used the collective skills of 70 law school students over six years to code competition laws in 131 jurisdictions between 1889 and 2010. DeMattee (2020) spent 567 person-hours coding a six-language corpus of 285 laws enacted by seventeen countries between 1872 and 2019. Researchers who wish to embark on similar quests should not underestimate the financial and human resources necessary to execute such projects.

The relative lack of comparative legal data, combined with the costs in time and money to create it, demonstrate why it is vital to discover new processes to allow researchers to accurately and economically translate legal texts. Appropriately using MT applications opens legal research across languages for more innovative comparative studies and increases access to less common case studies. Defining best practices for using these tools is a new and valuable contribution to qualitative and mixed-methods research.

Methods

To test the comparative performance of the MT applications against one another and human translators, we use similar legal texts from five countries that are written in major international languages. We examined three languages that use the Roman alphabet and two that use logograms. We used the legal definition of civil society organizations as specified in laws enacted by governments in Brazil, China, France, Japan, and Mexico to find comparable legal text across countries. These legal definitions vary slightly in content and length. Supplemental information (DeMattee et al. 2022) contains all original legal texts and the translated versions produced by DeepL, Google, and Microsoft.

We compared the original and translated versions of these legal definitions to assess which MT application has the highest usability and reliability across languages. First, we evaluated each translation application according to its flexibility in accepting and outputting file types, its ability to preserve the document's structure, and the available number of languages. Second, we tested the reliability and quality of the translations by testing the MT versions against human translations. Two native speakers per language of interest first translated the source text into English. Then, these same multilingual speakers evaluated four translated versions of a single source text: three MTs and one human. We randomized the order and anonymized the source of the translated texts and instructed evaluators to conduct their single-blind evaluations independently and without the assistance of other tools. Each evaluated the translated texts according to whether they required minor or critical edits for grammar and meaning. We defined minor edits as corrections made to the translated text that maintained the meaning of the source text, even if the translation is mediocre, and critical edits as corrections made to the text that did not maintain the original meaning.

Defining Five Measures: Generalizability, Flexibility, Presentation, Simplicity and Reliability

We define five measures to evaluate the quality of MT applications based on our past experiences with comparative research on association and charity laws (DeMattee 2022a, b; Bloodgood, Tremblay-Boire, and Prakash 2014) These measures—generalizability, flexibility, presentation, simplicity, and reliability—vary in importance depending on the research project.

Generalizability refers to the number of languages available for text-to-text translation. MT applications may be available for transliteration, translating the text in images, or text-to-speech translation. We expect that generalizability will increase with time as the demand grows for automated translation within and across applications, particularly among smartphone users. Increased generalizability does not guarantee that other measures of translation effectiveness will improve at the same pace. We argue that it is equally important, if not

more so, to know which MT application provides the most accurate translation for a given language.

Flexibility reflects the compatibilities of file formats in MT applications. Limitations on flexibility come in three forms: intake format, file size, and output format. When files are not compatible with MT applications, researchers must first convert them into appropriate formats. Another consideration for flexibility is file size, as some MT applications limit a file's upload size. Finally, output format typically corresponds to input format, but some MT applications allow more options for output formats. Flexibility in output format allows researchers to choose appropriate formats for subsequent analysis, reducing the incidence of human error.

Presentation is the degree to which the MT application preserves the source text's formatting and layout. Preserving identifiers accurately (e.g., alpha or numeric ordering) allows researchers to reference sections of a law correctly while facilitating replicability. If the translation process suppresses or distorts identifiers, researchers may need to invest considerable energy re-identifying articles, sections, subsections, and paragraphs for correct citations. This risks introducing human errors. Maintaining formatting (e.g., alignment and hanging indentations) is another consideration. Consistent formatting allows researchers to easily navigate and compare the original and translated versions. Likewise, protecting page breaks is stylistically desirable and valuable when a translation application strips identifiers from the document.

Simplicity is the number of actions necessary to translate a single document, including preparing files to be readable in MT applications. This factor becomes more relevant as the number of files, or the size of the files, in the legal corpus increases. Greater simplicity means less work to prepare files prior to translation and fewer opportunities for mistakes. Some PDF files are document images or scans of computer-generated text. These situations require an additional step. Optical character recognition (OCR) software converts images into machine-encoded text. The original composition of the text (e.g., manual versus computer typesetting) and image quality can affect OCR detection. In both cases, researchers may benefit from specialized software to improve OCR detection and obtain better results.

Reliability represents accuracy. The most important aspect of reliability is that translation maintains the meaning of original texts. If the translated text's meaning differs from the original, any analysis that follows will be severely flawed. Another consideration for reliability is grammatical and syntax accuracy; however, such errors may be minor enough not to alter the meaning of the original texts.

We use two tests to evaluate *reliability*. Our first measure uses native language speakers to evaluate translations of civil society laws in five languages. For each language, two single-blinded evaluators independently assess the source text and translated text at three levels. "Minor edits" are small, stylistic changes that polish the text to improve flow or readability. "Critical edits" involve substantive changes necessary to realign the translated text with the source text. Substantive changes go beyond slight improvements in readability and correct key errors that may otherwise jeopardize research findings. Not all critical edits require large-scale changes. For example, incorrectly translating a deontic such as "may" for "must" can seriously impact research findings. Finally, each evaluator made an overall assessment of whether the translated text maintained the source text's original meaning without any editing.

The second reliability measure is a series of round-robin translations. Here, we translate the source text to other languages and then translate it back to the original language. Translation applications, we assume, treat each translation as an independent task. Errors will therefore compile through multiple translations. Reliability is the similarity between the source text and the final translation reverted to the original language. We use English and German—two languages outside those studied—to conduct the backwards translations. We vary the number of translations from one to two foreign languages to further test each translation application's stability. We use a similarity score to assess these backwards translations. Similarity scores are the percentages of words in the backwards translation that identically, nearly, or relatedly match the words in the source text.

Overview of Machine Translation (MT) Tools

There are multiple applications available to researchers seeking automated translations. While their user interfaces may be similar, the algorithms that generate the translations vary. A number of human decisions feed into the final algorithms, which makes them objects of human creation capable of producing biased and fallible outputs (Diakopoulos 2013, 10; Salminen et al. 2020). Time and resources constrain programmers' ability to validate and update an algorithm's performance. This means that time and resources are additional factors that we expect will affect an MT application's accuracy. Older tools, or those maintained by organizations with greater resources, may be more accurate because of the greater availability of time and resources to train and debug the algorithms. Our research compares both free and proprietary algorithms provided by big and small companies, both old and new.

Analysis

By our assessment, Google is the most *generalizable* MT application because it has the most languages available for text-to-text translation at 108. With 72 languages, Microsoft is the next most generalizable. Between these two tools, even researchers working in regional languages (e.g., Haitian Creole) have options for translation applications. All three applications offer eleven common languages: Chinese, Dutch, English, French, German, Italian, Japanese, Polish, Portuguese, Russian, and Spanish.

DeepL and Google Translate are the most *flexible* MT applications, and are the only translators that accept PDF files, which is the format used most often in legal texts. Microsoft Translator requires researchers to convert PDFs to another file type for translation. This rigidity in file formatting is not a critical factor when a legal corpus is composed of a small number of text documents, but could quickly become a significant issue. As the size of the legal corpus expands, researchers can use programs like R or Python to manage file conversions, but this adds new technical requirements. Applications also vary in their ability to process files of different size. DeepL can process bigger files and larger quantities of files if users upgrade to a premium service. Google Translate limits file size, but does not offer a subscription service to process large documents. Microsoft does not provide explicit file size limits on its Translator; however, our experience is that the Microsoft application tends to fail in files above 100 pages. Output file types are as inflexible as input types. Google Translate currently does not offer an export function; all translations are displayed as a webpage that researchers must save as a PDF. DeepL and Microsoft render translations as a new file that researchers can save in various file types.

DeepL and Microsoft slightly outperform Google on the *presentation* measure. While all three applications protect font styles, Google Translate is prone to dropping identifiers, such as alpha or numeric section markers that allow researchers to navigate a law's contents. Such omissions may force researchers to re-identify articles, sections, subsections, and paragraphs to accurately reference passages. Our experience working with large and multilingual legal texts raises two concerns. First, sections of laws may refer to one another or point to schedules for further information. This means that researchers may need to translate entire laws rather than individual sections, necessitating accurate section markers. Second, laws can be long and the ability to accurately translate a large document is a critical matter.

Simplicity varies depending on the quality of the original document. If the starting point is a corpus of DOCX files, the three tools (DeepL, Google Translate, Microsoft) have equal simplicity, as they do not require users to reformat files before translation. If the corpus contains machine-encoded PDF files, then DeepL, Google, and Microsoft require those files be converted to a compatible format, which varies across applications. If PDF files contain scanned images of text, researchers must first convert them to text before translating the document. Researchers can use external OCR programs, such as Adobe Acrobat and R, to automate the conversion processes. In our comparison, Google Chrome always requires one more step than the other applications because researchers must convert files to HTML format.

Reliability estimates the precision of each translation application and the degree to which it is a reliable research tool. We discuss two measures of reliability: human evaluators' assessments of the translations, and round-robin similarity scores. We begin with human evaluators who edited and analyzed the machine translations at three levels: minor edits, critical edits, and whether the translation maintained the meaning of the source text. The applications' average scores are similar concerning minor edits; as Table 1 shows, we found approximately four minor edits for every 100 words. Note that human translators amended other human translations at nearly twice the rate—seven minor edits per 100 words—as the machine translations. Our tests on civil society laws suggest that, for most research purposes, minor errors do not prevent researchers from understanding and using the translation. In many cases, minor errors simply require researchers to work through legal text that is wordy or convoluted.

Our remaining two measures are more consequential. Critical edits are substantive changes that are necessary to realign the translation with the source text. Critical changes correct errors that may jeopardize research findings. Google Translate, which averaged 0.5 critical edits per 100 words, was the top performer on this measure. DeepL and Microsoft translators produced over twice as many critical edits, on average. The third measure is whether a translation maintains a text's original meaning. Overall, native language speakers had perfect inter-coder reliability on this measure. Google Translate was the top performer in maintaining original meaning. Eight out of ten evaluators graded Google Translate positively on this measure (with the exception of the two Japanese assessments). Indeed, the Japanese language speakers agreed that all three translation applications drifted from the text's original meaning. The two Portuguese-speaking evaluators found that DeepL Translator likewise lost the text's original meaning, and the two Chinese-speaking evaluators observed a similar error in Microsoft Translator.

Table 1: Native Language Speaker Reliability Analysis

	DeepL Translator	Google Translate	Microsoft Translator	Human Translators
Brazilian Legal Text				
Minor Edits	2.5 edits	2.5	1.5	4.2
Critical Edits	2.5 edits	0.8	0.8	0
Maintained Meaning	No	Yes	Yes	Yes
Minutes to Edit Translation	9.0 minutes	4.1	4.5	5.1
Words Added/Removed	+1.6 words	+2.5	-2.3	+2.6
Chinese Legal Text				
Minor Edits	5.9	4.7	5.1	8.5
Critical Edits	0	0	1.7	0
Maintained Meaning	Yes	Yes	No	Yes
Minutes to Edit Translation	7.8	7.5	11.9	8.5
Words Added/Removed	-2.9	0	-11.9	+12.3
French Legal Text				
Minor Edits	5.6	3.6	1.2	11.5
Critical Edits	0	0	0	0
Maintained Meaning	Yes	Yes	Yes	Yes
Minutes to Edit Translation	3.3	3.6	4.8	8.7
Words Added/Removed	+1.1	0	+1.2	+13.4
Japanese Legal Text				
Minor Edits	0.4	0.8	2.3	3.4
Critical Edits	3.1	1.6	3.5	0
Maintained Meaning	No	No	No	Yes
Minutes to Edit Translation	6.6	12.3	8.74	8.5
Words Added/Removed	-45.2	+2.5	-8.7	0
Mexican Legal Text				
Minor Edits	6.4	8.1	9.0	6.4
Critical Edits	1.3	0	0	0
Maintained Meaning	Yes	Yes	Yes	Yes
Minutes to Edit Translation	19.2	41.9	21.8	18.9
Words Added/Removed	+1.3	+5.4	+2.6	+1.3
Five-Law Average				
Minor Edits	4.2	3.9	3.8	6.8
Critical Edits	1.4	0.5	1.2	0
Maintained Meaning	Yes (6/10)	Yes (8/10)	Yes (6/10)	Yes (10/10)
Minutes to Edit Translation	9.2	13.9	10.3	9.9
Words Added/Removed	-8.8	+2.1	-3.8	+5.9

All measures were averaged across native language speakers (2 per country, 10 total). For comparability and interpretability, the measures of edits, minutes to edit, and words added or removed are all standardized to a common unit: per 100 words. Google Chrome was excluded due to extensive overlap with Google Translate.

The second reliability assessment measures the similarity between the source text and a final translation that we reverse-translate to its original language. This process determines a translation's stability over multiple machine translations and thus the potential robustness of the application. We calculated similarity scores using the online plagiarism checker CopyLeaks at three different levels of precision, from exact matches to translations

involving synonyms, rounded to the nearest whole number. Translations were checked through one (English or German) and two (English then German and German then English) translations before being returned to their original language.

The stability scores show that DeepL generally outperformed Google and Microsoft in these round-robin exercises (DeMattee et al. 2022, 14). DeepL appeared most stable for Chinese, French, and Spanish content and produced average scores at or above 50%; when the translation was reverse-translated, at least half of the words were identical, near, or related matches. Google and Microsoft were most stable for Spanish, French, and Portuguese content. DeepL's strongest outperformance was with the Chinese legal text. The reliability tests suggest that scholars might prefer different MT software for some languages over others. DeepL performs better for translations from Chinese, while Google or Microsoft works better for Romance languages. The general underperformance of machine translators for Japanese is notable.[1]

Limitations of Machine Translation

While our analysis suggests many benefits of adopting MT applications, we highlight three issues that researchers should consider. First, MT is much easier with newer laws, often published on national websites as HTML or machine-readable PDF, which are quickly and easily read by any application. Older laws may only be available as an image file rather than as text, requiring retyping and careful reformatting, which can induce human error. Second, laws (at least as regards associations and charities) are becoming longer over time, and free MT applications limit file upload size. Most MT applications require legal texts to be divided into smaller individual PDF documents before translation. A potential time-saving solution is to use MT applications to translate the table of contents or index of the law in one PDF, then translate the relevant sections. Finally, it is difficult with current tools to present both original and translated texts in parallel (side-by-side pages), which makes it slightly more challenging to evaluate the quality of the translation by comparing relatable blocks of text.

Conclusion

The "laws-in-translation problem" exists because governments rarely translate their laws into multiple languages and translating these legal texts typically requires choosing between affordability and precision. While translation professionals provide high-quality translations, these services quickly become cost prohibitive.[2] MT applications are an attractive alternative.

They are increasingly accessible, fast, and affordable. Still, scholars may be reluctant to use these applications for research purposes because the accuracy of the translations is unknown. Moreover, other researchers—specifically journal reviewers—may not be convinced by research findings that depend solely on translation applications. This adds another dimension to the laws-in-translation problem and the challenge of transparently and rigorously translating legal text for comparative research purposes.

We used similar legal text from five countries with diverse languages to assess the performance of three translation applications using five measures: generalizability, flexibility, presentation, simplicity, and reliability. Our assessment found translation applications to be effective but not precise or consistent enough to warrant use without verification. Minor errors aside, these automated tools occasionally make critical errors and lose a text's original meaning. MT tools may thus adversely affect research outcomes. By comparison, humans produce translations without critical errors or deviations from the text's original meaning, although human translators do take issue with other translators' texts.

These findings lead us to recommend that researchers pair machine translations with human translators to produce reliable and affordable translations. Extrapolating from data from our human evaluators, it takes an individual fluent in the necessary languages three times longer to produce a translated text than it does to edit an MT translation of the same source text. This suggests it is far more efficient for a researcher to use *any* translation application to make an initial translation, and then employ a human translator to improve that translation by making minor edits and correcting critical errors as required. As legal texts become longer, more complicated, with more cross-referencing, the efficiency gains from using an MT tool will increase. The use of the recommended protocol also provides a clear and tested method for translation that can be easily explained in future publications, increasing replicability and transparency in comparative law and policy research. We also suggest that researchers who adopt this process deposit and share translations in the public domain. Scholarly repositories (e.g., the Qualitative Data Repository), independent organizations (e.g., the International Center for Not-for-Profit Law), and personal websites can each host translated versions of these public documents.

The hybrid method for translating legal text that we have introduced can expand and improve comparative law and socio-legal research by drastically reducing

---

1 The mechanism causing this underperformance is beyond our scope, we simply note its existence.
2 The price for such services—even the most economical—begins at $25 per page or $0.05 per word.

linguistic knowledge as a constraint. The reduction in the time and translation costs opens new areas of research that would otherwise require large teams and grants to even consider. Accurate translations of new categories of law enables the systematic investigation of the evolution of law over time or the study of diffusion effects both within regions and globally. Generations of political scientists and legal scholars have emphasized the distinction between laws in the books and laws in action. This new method enables scholars to examine the extent to which *de facto* enforcement diverges from *de jure* rules. The large number of languages now available for machine translation, as well as the growth of international research societies providing networks with native speakers, enables research to redress the relative marginalization of some geographic areas that are either too poor to translate and publish their laws, or where legal text is written in a language that is unfamiliar to comparative legal scholars. Finally, the output of MT applications works easily as raw data for qualitative research software such as NVivo and ATLAS.ti.

References

Bennett, Andrew. 2007. "Symposium: Multi-Method Work, Dispatches from the Front Lines." *Qualitative Methods* 5, no. 1 (Spring): 9-28.

Berinzon, Maya, and Ryan C. Briggs. 2019. "Measuring and Explaining Formal Institutional Persistence in French West Africa." *The Journal of Modern African Studies* 57, no. 2 (June): 183-202. https://doi.org/10.1017/S0022278X19000077.

Bloodgood, Elizabeth A., Joannie Tremblay-Boire, and Aseem Prakash. 2014. "National Styles of NGO Regulation." *Nonprofit and Voluntary Sector Quarterly* 43, no. 4 (August): 716-36. https://doi.org/10.1177/0899764013481111.

Bradford, Anu, Adam S. Chilton, Christopher Megaw, and Nathaniel Sokol. 2019. "Competition Law Gone Global: Introducing the Comparative Competition Law and Enforcement Datasets." *Journal of Empirical Legal Studies* 16, no. 2 (June): 411-43. https://doi.org/10.1111/jels.12215.

Cook-Martín, David, and David Scott FitzGerald. 2019. "How Their Laws Affect our Laws: Mechanisms of Immigration Policy Diffusion in the Americas, 1790–2010." *Law & Society Review* 53, no. 1 (March): 41-76. https://doi.org/10.1111/lasr.12394.

DeMattee, Anthony J. 2020. "Domesticating Civil Society: How and Why Governments Use Laws to Regulate CSOs," PhD diss., Indiana University.

DeMattee, Anthony J. 2022a. "Are Constitutions Important in Diffusion Research?: How Constitutional-Level Rules, Autocratic Learning, and Emulation Shape the Development of Civil Society Laws in East Africa." Presented at the Workshop on Openness in Law-Related Research, New York City, New York, April.

DeMattee, Anthony J. 2022b. "A Grammar of Institutions for Complex Legal Topics: Resolving Statutory Multiplicity and Scaling-up to Jurisdiction-Level Legal Institutions." *Policy Studies Journal* (Forthcoming):1-34.

DeMattee, Anthony J., Nick Gertler, Takumi Shibaike, and Elizabeth A. Bloodgood. 2022. "Supplemental Information For: Overcoming the Laws-in-translation Problem: Comparing Techniques to Translate Legal Texts." OSF Preprints. February 23. doi:10.31219/osf.io/jc5p9.

Diakopoulos, Nicholas. 2013. "Algorithmic Accountability Reporting: On the Investigation of Black Boxes." *Tow Center for Digital Journalism, Columbia Journalism School*. https://doi.org/10.7916/D8ZK5TW2.

Elkins, Zachary, and Tom Ginsburg. 2021. "Characteristics of National Constitutions, Version 3.0." *Comparative Constitutions Project*. https://comparativeconstitutionsproject.org/download-data/.

Elkins, Zachary, Tom Ginsburg, and James Melton. 2009. *The Endurance of National Constitutions*. Cambridge: Cambridge University Press.

Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." *Annual Review of Political Science* 21, no. 1 (May): 29-47. https://doi.org/10.1146/annurev-polisci-091515-025429.

Giampieri, Patrizia. 2016. "A Critical Comparative Analysis of Online Tools for Legal Translations." *The Italian Law Journal* 2 (2): 445-61.

Glasius, Marlies, Jelmer Schalk, and Meta De Lange. 2020. "Illiberal Norm Diffusion: How Do Governments Learn to Restrict Nongovernmental Organizations?" *International Studies Quarterly* 64, no. 2 (June): 453-68. https://doi.org/10.1093/isq/sqaa019.

Goemans, Hein E. 2007. "Qualitative Methods as an Essential Complement to Quantitative Methods." *Qualitative Methods* 5, no. 1 (Spring): 11-13.

Hummel, Calla, John Gerring, and Thomas Burt. 2021. "Do Political Finance Reforms Reduce Corruption?" *British Journal of Political Science* 51, no. 2 (April): 869-89. https://doi.org/10.1017/S0007123419000358.

Kapiszewski, Diana, and Sebastian Karcher. 2021. "Transparency in Practice in Qualitative Research." *PS: Political Science & Politics* 54, no. 2 (April): 285-91. https://doi.org/10.1017/S1049096520000955.

Law, David S., and Mila Versteeg. 2013. "Sham Constitutions." *California Law Review* 101, no. 4 (August): 863-952.

Lemon, Edward, and Oleg Antonov. 2020. "Authoritarian Legal Harmonization in the post-Soviet Space." *Democratization* 27 (7): 1221-39. https://doi.org/10.1080/13510347.2020.1778671.

Lin, Yu-Hsin, and Yun-chien Chang. 2018. "An Empirical Study of Corporate Default Rules and Menus in China, Hong Kong, and Taiwan." *Journal of Empirical Legal Studies* 15, no. 4 (December): 875-915. https://doi.org/10.1111/jels.12200.

Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer, and Dustin Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23, no. 2 (Spring): 254-77. https://doi.org/10.1093/pan/mpu019.

Pound, Roscoe. 1910. "Law in Books and Law in Action." *American Law Review* 44, no. 1 (January-February): 12-36.

Salminen, Joni, Maximilian Hopf, Shammur A. Chowdhury, Soon-gyo Jung, Hind Almerekhi, and Bernard J. Jansen. 2020. "Developing an Online Hate Classifier for Multiple Social Media Platforms." *Human-centric Computing and Information Sciences* 10 (1): 1-34. https://doi.org/10.1186/s13673-019-0205-6.

Back to the Field: Uncertainty and Risk in Field Research

Jannis J. Grimm
Freie Universität Berlin

Sarah E. Parkinson
Johns Hopkins University

Ellen Lust
University of Gothenburg

Isabell Schierenbeck
University of Gothenburg

Kevin Koehler
Sant'Anna School of Advanced Studies - Pisa

Dina Zayed
University of Sussex

The rapid spread of COVID-19 beginning in early 2020 caused global disruption. As the risk of infection rose and public health authorities around the world enacted measures to contain the virus, everyday life ground to a halt. Activities that seemed routine in late 2019 became fraught with uncertainty. Fieldwork was no exception. Most field researchers had to change or cancel at least some of their plans; some left their field in a hurry before travel was shut down while others had to lock down on site; most academic institutions restricted travel, with some even prohibiting all forms of international movement. In brief, many traditional forms of fieldwork became all but impossible during the pandemic.

Even as parts of the world begin to emerge from the pandemic, things have not returned to normal. Indeed, the emergence of the Omicron variant in November 2021 led to new restrictions, with some universities again moving to block field research.[1] It is important to note that such restrictions sometimes seem to be driven by factors other than the risk of infection alone. Infection rates in parts of Europe or the US were frequently just as high or even higher than in other parts of the world, yet restrictions seemed to be primarily aimed at preventing movement between the Global North and South. At the same time, global vaccine inequalities and vaccine resistance are threatening to relegate parts of the world to the category of places which are not safe

---

1 For example, Arantxa Rodriguez-Uribe (@MAranzazuRU, November 30, 2021) tweeted that Princeton had just banned all international fieldwork.