Text Mining for Congressional Policymaking in the Philippines: A Public Forum

Ronald A. Pernia, PhD

Words are an integral part of politics. Politicians and citizens alike employ words to convey opinions, propose initiatives, and justify their actions. Moreover, laws and regulations are predominantly *articulated* through written language (Wilkerson and Casas 2017). In today's digital age, the vast amount of unstructured text (i.e. raw freeform text that is not broken down into a fixed record format) available in various formats offers significant potential for policy insights and decision-making. However, much of the information about our political world is found not in neat datasets but in various unstructured texts. These include presidential national addresses, social media posts, city council minutes, local ordinances, judicial decisions, and Congressional speeches.

How do scholars then *extract* meaning from such an abundance of textual data, which is too extensive for any single researcher to read and analyze, and transform it into *structured* and *valuable* information for scientific inquiry? This challenge has previously hindered the use of textual data in political science research but has been increasingly addressed through technological advancements (Grimmer and Stewart 2013; Grimmer, Roberts, and Stewart 2022).

Recent advancements in technology have made it feasible to *quantitatively* analyze large volumes of text. *Text mining*, a process of deriving meaningful information from text, has emerged as a powerful tool in fields including political analysis and policy development. The application of these machine learning tools to social scientific data allows us to *discover* new concepts, measure their prevalence, assess *causal* effects, and even make *predictions* (Burscher, Vliegenthart, and De Vreese 2015; Grimmer, Roberts, and Stewart 2021). This has enabled political scientists to access an unprecedented amount of data, and the research community offers accessible text analysis software packages, along with training and support (Wilkerson and Casas 2017).

In this public forum, we will explore the increasing use of *text-as-data* research in political science (Grimmer and Stewart 2013; Grimmer, Roberts, and Stewart 2021). This approach has been applied across various disciplines and contexts, from health and medicine to political communication and legislative studies. With its flexibility and broad array of technological applications, students and scholars of empirical social science are tapping into new data sources, employing diverse methods, and becoming critical consumers of findings based on these methods (Wilkerson and Casas 2017; Ke et al. 2024; Lucas et al. 2015).

Overall, the inclusion of machine learning tools in social science research, as demonstrated through our real-world data example, highlights how text mining and text analysis can be effectively applied within the increasingly *data-driven* Philippine socio-political landscape.

This forum aims to provide participants with a comprehensive understanding of these techniques and their applications, fostering insightful discussions on the potential benefits and challenges of text-as-data research in political science. Ideally, this forum will serve both as a *pedagogical tool* and a valuable *starting point* for students and teachers to initiate their own text-as-data projects. Through this forum, we aim to bridge real-world data with novel quantitative techniques, enriching participants' methodological toolkits and inspiring innovative research approaches.

References

Bertelli, Anthony M., and Jeffrey B. Wenger. 2009. "Demanding Information: Think Tanks and the US Congress." *British Journal of Political Science* 39 (2): 225–42. https://doi.org/10.1017/S0007123408000410.

Burscher, Bjorn, Rens Vliegenthart, and Claes H. De Vreese. 2015. "Using Supervised Machine Learning to Code Policy Issues: Can Classifiers Generalize across Contexts?" *The ANNALS of the American Academy of Political and Social Science* 659 (1): 122–31. https://doi.org/10.1177/0002716215569441.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18 (1): 1–35. https://doi.org/10.1093/pan/mpp034.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24 (Volume 24, 2021): 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921.

———. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. https://doi.org/10.1093/pan/mps028.

Ke, Zheng Tracy, Pengsheng Ji, Jiashun Jin, and Wanshan Li. 2024. "Recent Advances in Text Analysis." *Annual Review of Statistics and Its Application* 11 (Volume 11, 2024): 347–72. https://doi.org/10.1146/annurev-statistics-040522-022138.

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 (2): 254–77. https://doi.org/10.1093/pan/mpu019.

Maestas, Cherie. 2003. "The Incentive to Listen: Progressive Ambition, Resources, and Opinion Monitoring among State Legislators." *The Journal of Politics* 65 (2): 439–56. https://doi.org/10.1111/1468-2508.t01-3-00008.

Panao, Rogelio Alicor L., and Ronald A. Pernia. 2022. "Fear and Loathing or Strategic Priming? Unveiling the Audience in Duterte's Crime Rhetoric." *Journal of East Asian Studies* 22 (1): 77–98. https://doi.org/10.1017/jea.2022.1.

Thompson, Gabriela, Anthony Staddon, and Rick Stapenhurst. 2020. "Motivation of Legislators and Political Will." *Public Integrity* 22 (2): 134–53. https://doi.org/10.1080/10999922.2018.1511669.

Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20 (Volume 20, 2017): 529–44. https://doi.org/10.1146/annurev-polisci-052615-025542.