The Predictive Potential of Political Discourse

Language and discourse analysis has the potential to reveal the innovation and diffusion of global ideas, particularly in closed societies and organizations where intelligence about internal group dynamics is sparse and unreliable. Political discourse is under-utilized as a predictor for socio-political instability and crisis. However, automated computational linguistics programs can leverage the power of big data sets to generate new insights into patterns of political behavior at the individual (leader), group, and state levels.

I. Challenges, questions, and needs

Challenges: The unpredictable behavior of authoritarian leaders and violent non-state actor groups presents challenges in deciphering their capabilities and intentions. Yet their written and spoken discourse can reveal latent patterns of power transitions, internal group dynamics, organizational hierarchy, larger group membership, and threat credibility. For example, most research on bluffs and threats evaluates aggregate data at the country-year level, rather than at the level of speech act. At its core, a threat is a speech act: leaders make and break international treaties, and engage in the activities of war-making, through the use of speech acts called performative utterances – statements that not only describe a situation, but can also change the social and political reality. For example, when a leader declares war, it is not just a string of empty words, but rather it carries the implication that the country intends to act aggressively. Leaders intend bluffs to be indistinguishable from credible threats; computational linguistics can help distinguish between them.

Other challenges include mapping the hierarchy of individuals within politically opaque regimes and non-state organizations, including changes to the leadership structure. Innovations in group communication dynamics can help to reveal emerging power struggles, strategies for identifying saboteurs, and crafting counter-messaging narratives to interrupt the process of radicalization. The intelligence and social scientific communities also face challenges in predicting both micro- and macro-level changes, like coup-proofing in authoritarian regimes, and the contagion of democratization and democratic backsliding like that which occurred during the Arab Spring. For example, research has demonstrated marked linguistic differences in the leaders who lost power and those who remained in power during the Arab Spring.[1]

Questions: The challenges facing scholars and intelligence analysts yield important agenda-setting questions that computational analyses of political language can help answer:

- *When are leaders bluffing, and when are they making credible threats? Specifically, what linguistic features of threatening and deceptive communications characterize credible threats and bluffs?*[2]

- *What linguistic features indicate democratization and democratic backsliding as well as transitions toward and away from improvements in human rights practices?* One approach to answering this question is using semantic analysis to evaluate the differences between authoritarian and democratic leaders' speeches. Linguistic formality indicates common ground amongst speakers, identifying in-groups and out-groups, and regime characteristics that matter to this designation.[3] Recent work suggests that leaders of non-democracies use more formal language in public forums like the United Nations General Assembly general debate because they share less common ground – both with each other and with democracies – and because they leverage their language to bolster their credibility and legitimacy and overcompensate for their out-group status.

- *What type of language indicates shifts in internal regime dynamics, like coup-proofing, as well as waxing/waning allegiances in the international system?* Changes in linguistic formality are also useful in identifying coalitions and patterns of association in international politics. Several countries, including South Africa, Gambia, Burundi, and Russia, have indicated their intent to withdraw from the International Criminal Court (ICC). The linguistic style matching literature would suggest that countries with similar priorities and repeated interactions should align their speaking styles. Latent semantic analysis (LSA) can help identify these types of trends and help predict which vulnerable countries might exit the ICC or other international organizations.

- *How does language change over the course of an individual's political radicalization?* The proliferation of online social communities in forums like Tumblr, Twitter, Reddit, Facebook, and Weibo provides ample opportunities for observing individuals' process of radicalization as well as interrupting it.[4]

- *How does language help us understand changes in political systems that affect government's ability (and willingness) to satisfy the needs of citizens (including political participation)?* In particular, social media provides valuable insights into public opinion and depth of discontent that can indicate political instability and potentially leadership change.[5]

- *Which persuasion strategies do leader of countries and non-state organizations use in propaganda?[6]*

Needs: From the challenges and questions, we can identify gaps in capacity that collaborations between the social science and intelligence communities can fill:

- Computational programs to assist human intelligence in deciphering bluffs from credible threats, understanding changes in organizational structure including personnel and goals, and detecting shifting regional and global allegiances.

- Uniform data streams for speech acts like threats, overtures for cooperation, and changes in both state and non-state actors' policies.

- Access to declassified Department of Defense data and documents, like those that existed in the archives of the Conflict Records Research Center at the National Defense University.

- Increased attention to the valuable role that political discourse and computational linguistics can play in modeling international behavior and forecasting trends.

II. What makes these challenges and questions important at this time and in the foreseeable future?

Nearly thirty years ago at the end of the Cold War, global instability increased through countries' transitions to democracy, shifts in alliances and organizational memberships, and the rise of violent non-state actor groups. Localized conflicts like civil wars displaced interstate wars as the primary type of large-scale violence. Many global developments have generated a similar type of turmoil in the international system; whereas the post-1989 world issued in multinational cooperation in human rights, trade, and the environment, recent trends point toward a reversal of liberal phenomena. This includes worsening human rights practices, departures from international agreements and organizations, and democratic backsliding. Automated analysis of political language can help to reveal the dynamics of these complicated political processes, like the political processes in opaque regimes and organizations. It also adds value to computational methodologies like agent-based modeling. Political language provides dimensionality, and an alternative to using annual aggregate data often used in forecasting. While linguistic data is often noisy, it provides more variation within countries than other political indicators.

III. Security Benefits

The security benefits of rigorous study of trends in authoritarian leaders' language include the ability to track trends in tone, semantics, and topics, and apply these indicators in forecasting models. Political language pattern detection (PLPD) incorporates leader discourse and political events generated using analytical schemas like CAMEO alongside data sets like ICEWS and GDELT.[7] The relationship between political language and actual events is an under-explored area of research, allowing scholars to probe important questions: When leaders use more cooperative language, is this associated with more cooperative events? Or, when leaders use more belligerent language, do they in turn act more combatively in the international system? Additionally, models employing recurrent or deep neural architecture will be useful to generate predictive models that yield more nuanced political forecasts. State of the art deep learning models and the most recent advances in machine learning can inform inform the understanding of how events and language interact in the international system.

PLPD also helps to explain changes in governance, state capacity, and the viability of transnational regimes by leveraging linguistic information and rigorous statistical methodology to uncover latent relationships and trends.[89] It also reveals the innovation and diffusion of global ideas, including their uptake in different cultural and political contexts like the cascade of withdrawals from the ICC. Through language, scholars can track behaviors in response to strongly held beliefs, including the recruitment of terrorists and sympathizers, such as the case of Aqsa Mahmood, the young Scottish woman who made *hijrah* to Syria while chronicling her radicalization online. Her situation is of particular interest given the uniqueness of transnational recruitment of a young woman to radical political behavior; transnational recruitment has predominantly targeted young men. Political linguistics also has important implications for decoding propaganda and messaging, including gauging public well-being, trust in government and support for political leaders, and destabilization of states. In an analysis of state media in North Korea, scholars have uncovered changes in the distribution of topics related to national infrastructure, including health and education.[10]

The Department of Defense has an extensive collection of printed documents and conversations that require analysis. These digitized corpora are growing rapidly. As a consequence, there are tremendous demands on intelligence agencies to dramatically enhance their ability to analyze texts automatically and efficiently, and to assess the psychological, social

and cultural characteristics of communication. The speed of these automated analyses needs to be optimized in order to respond to threats in a timely manner. Computational linguistics tools will assist intelligence missions, whether the texts are extracted from formal or informal sources. Computational methods can provide very fast and efficient markers of basic social dynamics that may in the future be used to assess leadership dynamics in authoritarian regimes, cohesiveness of groups, social and ethnic identity of individuals, indicators of deception, or hostile threats. These tools will help analysts understand the past actions and cognitions of previous regimes.

IV. Conclusions

Traditionally, political outliers (i.e., leaders of relatively rare regime types like Kim Jong Un's North Korea, Bashar al-Assad's Syria, Saddam Hussein's Iraq, or Muammar Gaddafi's Libya) have defined the global narrative of security threats. Moments of crisis occur when politically unconstrained leaders demonstrate their ability to project power by engaging in dangerous rhetoric or displays of force, like conducting missile or nuclear weapons tests. In other words, relatively few actors set the tone for global issues, like nuclear weapons, state-sponsored terrorism, or interstate conflict. Somewhat ironically, these few regimes tend to be opaque and credible information difficult to obtain. However, state media and leader speeches are both valuable and underused sources of information about leaders' priorities, credibility, and intentions. On the other hand, the groundswell of both violent and nonviolent social movements, and the accompanying social media they generate, harness the power of citizens who readily provide streaming real-time information about political activities. This includes radicalization of young women to high-risk political activities, a rather new phenomenon in international politics. The Euromaidan protests in Ukraine, as well as the subsequent Russian incursion into Ukraine, found a voice and outlet online, and linguistic analysis has helped to decipher how state and independent media cover such activities.[11]

Finally, the first and second waves of democratization have begotten political progress toward liberal economic and human rights values and have largely been viewed as successes and progress. However, recent authoritarian trends in mature democracies, like France, Denmark, Germany, Poland, and the Philippines, indicate that democratic values may be losing a foothold, and that the previously small degree of support for authoritarian leaders may be increasing. Language serves as a leading indicator of this changing political climate, detecting shifts within domestic political dynamics as well as between states in the international system.

Works Cited

[1] Leah Windsor et al., "Leader Language and Survival Strategies in the Arab Spring," August 2015.

[2] Cindy K Chung and J. W. Pennebaker, "Using Computerized Text Analysis to Assess Threatening Communications and Behavior," in *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*, 2011, 3–32.

[3] Kate G. Niederhoffer and James W. Pennebaker, "Linguistic Style Matching in Social Interaction," *Journal of Language and Social Psychology* 21, no. 4 (December 1, 2002): 337–60, doi:10.1177/026192702237953; Nathanael J. Fast, Chip Heath, and George Wu, "Common Ground and Cultural Prominence: How Conversation Reinforces Culture," *Psychological Science* 20, no. 7 (2009): 904–11, doi:10.1111/j.1467-9280.2009.02387.x; William S. Horton and Boaz Keysar, "When Do Speakers Take into Account Common Ground?," *Cognition* 59, no. 1 (April 1996): 91–117, doi:10.1016/0010-0277(96)81418-1.

[4] Leah Windsor, "The Language of Radicalization: Female Internet Recruitment to Participation in ISIS Activities" (International Studies Association, Atlanta, GA, 2016).

[5] The AmericasBarometer by the Latin American Public Opinion Project (LAPOP), www.LapopSurveys.org.

[6] R. E. Petty and J. T. Cacioppo, "The Effects of Involvement on Responses to Argument Quantity and Quality: Central and Peripheral Routes to Persuasion.," *Journal of Personality and Social Psychology* 46, no. 1 (1984): 69.

[7] Conflict and Mediation Event Observations; Integrated Crisis Early Warning System; Global Database of Events, Language and Tone

[8] Gregory Love and Leah Windsor, "Alo' Populism!: Discourse and Presidential Approval in Chavez' Venezuela" (International Studies Association Annual Conference, New Orleans, LA, 2015).

[9] Leah Windsor, "Allies in Peace: Communication in the United Nations General Assembly" (International Studies Association – International Communication conference (ISA-ICOMM), Atlanta, GA, 2016).

[10] Leah Windsor, "Discourse and Discord in North Korea," 2016.

[11] Labzina, Elena and Mark David Nieman. 2016. Understanding the Use of State-Controlled Media: A Machine-learning Approach. Presented at International Studies Association-Midwest. St. Louis, MO, November 18.