Language Patterns in International Relations

Leah Windsor

Language, and the meaning behind it, represents an ongoing challenge to understanding the intentions of others. How can we know if a leader is making a credible threat, or just bluffing? When do leaders’ words signal their longevity in office, or the stability or popularity of their regime? How can language provide insight into opaque political environments, like dictatorships? How can we infer culture from language, and clarify biases in existing paradigms? What language is likely to attract and persuade individuals to radicalize? The use of text-as-data (i.e., computational discourse analysis) is evolving in its efforts to answer such pressing questions about international relations.

The North Korean regime, for example, periodically makes threatening statements, and occasionally carries through with them; sometimes these statements are intended for international adversaries, but sometimes they are directed toward an internal audience. We can evaluate patterns by analyzing linguistic content or style. Content refers to the topics a speaker includes in a speech or document; style refers to the manner in which they are delivered. Both approaches provide insight into patterns of conflict and cooperation. As we see in Figure 1, we observe trends in anti-imperialist language and political events between 1997 and 2015 in North Korea. Toward the end of Kim Jong-Il’s regime (in 2011), we see a precipitous decline in the use of saber-rattling rhetoric, and a corresponding increase in the number of cooperative events. During the transition time when Kim Jong-Un takes power (2011-2012), we see a decrease in cooperativeness and increase in anti-imperialist rhetoric. This suggests different strategic mechanisms at work: regime transition times are fraught and uncertain, and leaders may use less combative language.
Figure 1. Content: Combative language from North Korean state media

Figure 2. Style: Localized cohesion and event intensity in North Korea
A different indicator that measures linguistic style is called “referential cohesion” and it follows a similar pattern to the intensity scores, as shown in Figure 2. Referential cohesion captures the degree to which a speaker’s language is locally cohesive, with repeated words or phrases and semantically related concepts. Language with higher referential cohesion tends to be more conceptually simple, and less cognitively demanding on the audience. This linguistic style is straightforward and uncomplicated; as political events become more pacific, state media language becomes easier to parse and comprehend. On the other hand, as political relations worsen, the semantic presentation becomes more complex. It seems that linguistically, peace is more straightforward, whereas conflict is more complicated to communicate.

The implication for policymakers is twofold. First, we find utility in studying the language of state media, as systematic variation in language can be used to help predict and forecast political events. While some might regard North Korean state media as noisy propaganda, we find evidence that the content and style changes in meaningful ways, such as during times of crisis and domestic uncertainty. For example, the young new leader Kim Jong-Un needed to generate support in his new position of power, being relatively inexperienced and unknown to the North Korean elites in the “winning coalition” (Figure 1). During this time we observe a marked increase in belligerent rhetoric in their state media.

Second, we observe changes in the style and presentation of state media reports, which we have connected to theories of persuasion (Figure 2). Substance and style change during peaceful and conflict-ridden times, and we can gain insight into how North Koreans – including the leader – think about the nature of these problems, and make inferences about their worldview and state of mind.

**Language Conveys Meaning between Countries and Leaders**

Meaning varies across culture, language, context, and time. Leaders may use the same word, or semantically related concept, but with different intentions. In the international system, lexical and semantic ambiguities can lead to uncertainty about leaders’ preferences, goals, and strategies, especially in authoritarian regimes which tend to be more opaque than democracies. Linguistic relativity, i.e., the Sapir-Whorf Hypothesis (Whorf 1940), suggests that the structure of language influences actors’ worldview, including grammar and vocabulary that can be modeled computationally with multilingual (and monolingual, multicultural) data, such as leader speeches, debates, and social, traditional, and state media sources. We engage the ongoing debate over the Sapir-Whorf Hypothesis—how meaning varies across language and culture—using the examples of “war” (Figure 3, Column 1) and “conflict (Column2). As Figure 3 shows, there are similar but distinct patterns (between languages over time) among semantically related concepts. At a very basic level, countries and languages make distinctions between the phenomena of war and conflict, suggesting that we cannot always assume shared meaning even with the most basic of international relations concepts. In other words, a native French speaker speaking to a native Chinese speaker may have different conceptualizations of these two terms.

Varying interpretations may rest on even more nuanced elements of language. While conflict and war are relatively concrete ideas, function words such as articles and conjunctions, can be the source of misinterpretation and mistranslation. Function words are the glue that bind content words together; in the case of UN Resolution 242 following the Six-Day War, the French and English versions held discrepancies based on the French article “des”. While the English version referred to “territories”, the French version required an article (“des territoires”), leading to controversy about whether the language includes some or all of the Occupied Territories.
Figure 3. Changes in meaning (Y-axis: Term Usage Frequency) over time (X-axis: 1900-2008), by language: English, German, French, Russian, and Chinese (Source: Google Ngram)

To quantify differences in meaning and language usage, the Languages Across Cultures lab studies country-level discourse. Scholars already use annual, aggregate indicators to study international relations such as GDP, military expenditures, trade patterns, energy consumption, and conflict participation; to this observational data, we add linguistic information at the leader/country level. For example, we find variation in linguistic formality across the world, as we see in Figure 4, derived by analyzing the United Nations General Assembly (UNGA) General Debates with a syntax/semantics tool called Coh-Metrix. Five features comprise formality: syntactic complexity; word concreteness; narrativity; deep cohesion; and referential cohesion. For example, syntax simplicity describes how syntactically simple or complex a sentence, paragraph, or document is. This refers to the grammatical structure of the textbase where more simple syntax is more easily understood, and less cognitively demanding for the audience. On the other hand, complex grammar requires more cognitive effort to parse. Syntactic simplicity can indicate the relative status of a member of an organization, as this feature can mark
hierarchy. A more junior member may use more complex language deferentially toward more senior members as a sign of respect. Speakers using very formal language may perceive themselves as members of an out-group, or subordinates, whereas speakers using less formal language may perceive themselves as having an in-group identity with shared values, culture, experiences, and referents.

One takeaway from Figure 4 is that more democratic and more politically institutionalized countries tend to use less formal language. Note, for example, that China has the same relative formality as the United States, Brazil, Canada, and Norway. While classified as an autocracy by ratings systems such as the Polity Project and Freedom House, China is a robust single-party political structure with quasi-democratic features, and it is well-integrated into the social, political, economic, and technological fabric in the international system. As such, it is unsurprising that China ranks alongside democracies in its language use given socialization in the international system, including its stature as a global economic powerhouse and permanent member of the UN Security Council. Additionally, tracking these indicators over time can provide clues to the formation and dissolution of alliances, the rise of regional powers, internal regime dynamics, and trends in democratization and democratic backsliding. For example, leaders that use more syntactically complex language may be trying to demonstrate their credibility or legitimacy; in many cases, leaders seem to be letting their words do the work for them. This is likely the case in countries like North Korea, which use some of the most formal language in the international system.

Figure 4. Linguistic formality in the world

Another finding from this line of inquiry is that leaders holding office longer tend to use more formal language (Figure 5). Using the Archigos data set on leaders and linguistic data derived from the UNGA General Debates, we classified leader use of formal/informal language alongside their length of time in office (Goemans, Gleditsch, and Chiozza 2009).
Figure 5. Leader language formality by number of years in office (Source: UNGA and Archigos)

Figure 5 shows the country language and formality pattern presented differently. Leaders who are in office longer, and who represent non-democratic countries, tend to use more formal language. This is somewhat contrary to our expectations: usually, longer group membership and duration correlate with less formal language. However, in this case, the longer a leader remains in power (above and beyond the average time of 8 years), the more of an isolated outlier s/he becomes. Examined in this context, language can represent socialization within the international system, and tracking linguistic trends over time helps identify changes in the composition of affiliations, such as alliances and group membership. It can also indicate internal political instability, as shown in Figure 6. In a paper on political survival strategies in the Arab Spring published in International Interactions in 2018, we found differences in language use among leaders of countries experiencing sociopolitical unrest (Windsor, Dowell, Windsor, and Kaltner 2018). Leaders who remained in power used more positive and less negative, anxious, and angry language; leaders who lost power used twice as much negative language, more risk-associated words, and more third-person pronouns.

In Figure 6, we see that Jordan is an outlier among its regional peers, demonstrating consistent cohesive use of informal language. The outlier with the highest formality – Oman in 2013 – represents a significant departure from the country’s mean; in this year, the longest-ruling leader in the Middle East – Sultan Qaboos – made concessions following widespread civil disobedience and a government crackdown on civil liberties. This is an example of how a leader
can increase his or her formal register to manipulate perceptions of credibility, status, and legitimacy.

![Figure 6. Arab Spring language trends](image)

**Developing Computational Linguistics for Analyzing Language**

The above analyses were performed using English-language corpora. Coh-Metrix is a computational tool that analyzes written and spoken texts, producing over one hundred indices that represent text features at different levels (word, sentence, and document), such as density of certain word types (pronouns, nouns, verbs, adjectives, and adverbs). At present, most computational linguistics programs – with the exception of bag-of-words approaches like topic modeling – rely on English language sources (De Vries, Schoonvelde, and Schumacher 2018). If researchers are limited functionally and computationally to English-language sources, then we lack the ability to generalize beyond Western democracies and linguistic styles. In other words, we have good evidence that leaders of non-democratic states (and members of violent extremist organizations) have different institutional constraints and decision-making strategies than do leaders of democracies. This leads to different political outcomes in the world, such as variation in human rights practices, domestic and international conflict involvement, and ability to uphold organizational commitments. To address this problem, we developed Coh-MetrixML, an extension of the existing Coh-Metrix from a single language (English) to multiple languages (Windsor and Cai 2018). Currently, we are developing Coh-MetrixML for Arabic, Chinese,
French, Germany, Spanish, and Russian. The tool will be released for free use and will provide invaluable insight into the political dynamics of regimes and actors using original source languages.

Coh-MetrixML is a tool that we hope will help remedy the reliance on English-language sources, and encourage syntactic and semantic exploration of linguistic features in international politics. At present, there is a strong bias toward analyzing language of Western politicians and institutions, including legislatures and judiciaries, as this data is more readily available – and in more user-friendly formats – than language data from the developing world and non-democracies. While the institutional transparency in democracies can facilitate easier collection and analysis of language, it is also important to collect data from a representative sample of countries, as findings may not generalize across regime and governance types. Coh-MetrixML will provide more analytical insight and nuance for primary-source documents into the patterns of language and strategies that leaders use. As such, we can better understand where meanings diverge and help identify patterns to threats and bluffs, democratization and backsliding trends, and engagement in (or withdrawal from) commitments to international norms.

Associated Reading

Biography
Leah Windsor is a Research Assistant Professor in the Institute for Intelligent Systems at The University of Memphis. She runs the Languages Across Cultures lab and her interdisciplinary approach to understanding political language is situated at the intersection of political science, linguistics, and cognitive science. Her work examines governance, power, and communication to answer the question: How does our language reveal who we are? To learn more visit https://quanttext.com or https://www.leahewindsor.com.

Associated Minerva Project
Political Crisis and Language: A Computational Assessment of Social Disequilibrium and Security Threats

Supporting Service Agency
Air Force Office of Scientific Research

Nota Bene
Content appearing from Minerva-funded researchers—be it the sharing of their scientific findings or the Owl in the Olive Tree blogs posts—does not constitute Department of Defense policy or endorsement by the Department of Defense.