Young-Shin Jun, Ph.D.
Professor, Energy, Environmental & Chemical Engineering (ECEE), Washington University in St. Louis

Young-Shin Jun is a professor in the Department of Energy, Environmental & Chemical Engineering (ECEE) at Washington University in St. Louis (WUSTL). Professor Jun earned her Ph.D. from Harvard University (2005) and conducted postdoctoral research at the University of California-Berkeley/Lawrence Berkeley National Laboratory (2005–2007). She received a U.S. National Science Foundation CAREER Award in 2011. She was named a New Faculty Fellow by the U.S. National Academy of Sciences in 2013, and a Frontiers of Engineering Fellow by the U.S. National Academy of Engineering in 2016. Her research group interests include solid-water interfacial reactions and nucleation and bio-inspired photothermal membrane materials development.

Deoukchen Ghim
Ph.D. Student, Energy, Environmental & Chemical Engineering (ECEE), Washington University in St. Louis

Deoukchen Ghim received his B.S. and M.S. in chemical engineering from University of Seoul, Republic of Korea in 2014 and 2016, respectively. He is currently a Ph.D. student in the Department of Energy, Environmental & Chemical Engineering (ECEE) at WUSTL. His research interests include the design and synthesis of nanomaterials to produce clean water by using solar thermal water purification, photocatalysts, and pressure-driven membranes.

Xuanhao Wu
Ph.D. Candidate, Energy, Environmental & Chemical Engineering (ECEE), Washington University in St. Louis

Xuanhao Wu received her B.S. in environmental science and technology from Fudan University, China (2015). She is currently a Ph.D. candidate in the Department of Energy, Environmental & Chemical Engineering (ECEE) at WUSTL. Her research interests include solid-water interfacial reactions and nucleation and bio-inspired photothermal membrane materials development.

Sisi Cao
Ph.D. Student, Mechanical Engineering and Materials Science, Washington University in St. Louis

Sisi Cao received her B.S. in materials science and engineering from Hunan University, China (2013) and her M.S. in materials science from WUSTL (2015). She is currently a Ph.D. student in the Department of Mechanical Engineering and Materials Science at WUSTL. Her research focuses on the fabrication of inorganic and organic nanostructures to produce clean water by using solar thermal water purification, photocatalysts, and pressure-driven membranes.

Srikant Singamaneni, Ph.D.
Professor, Mechanical Engineering and Materials Science, Washington University in St. Louis

Srikant Singamaneni is a professor in the Department of Mechanical Engineering and Materials Science at WUSTL. He obtained his Ph.D. in polymer materials science and engineering from Georgia Institute of Technology in 2009. He received a 2011 Translational New Investigator Award from Congressionally Directed Medical Research Programs and a 2013 U.S. National Science Foundation CAREER Award. His research group’s interests include the design, synthesis, and self-assembly of plasmatic nanostructures for various biomedical and energy applications.

Leah C. Windsor
The language people use can reveal a tremendous amount of insight into their preferences, beliefs, and attitudes. When we examine language use through quantitative or computational analysis, we can uncover patterns of behavior that are not discernable through evaluation of their content alone.

This article explores ways in which computational discourse analysis can help the U.S. Department of Defense (DoD) and the Intelligence Community (IC) better understand and predict problems related to international events and crises, such as democratization and democratic backsliding, security and defense intelligence analysts, and longevity of foreign leaders in political office.  

Quantitative analysis of speech can aid en-
al actors [16]. The discourse analysis tech-
niques demonstrated below can help DoD
and IC intelligence analysts foresee or an-
alyze political instability, explain changes
in state governance or military capacity,
and decode intentions behind bellicose rhetoric.

Language in the International System

Scholars of international relations already
use annual, aggregate observational data to
analyze and make predictions about global
political dynamics—including gross domestic
product per capita, governance, annual mili-
tary expenditures, conflict involvement, trade
volume, and population demographics. To
this discussion, I add robust language data
using text data from UNGA general debate
sessions from 2004 to 2018 as a sample
dataset (the techniques presented below are
applicable to other datasets and corpora).

Some of the standard covariate indicators
are dynamic and responsive to exogenous
shocks, such as civil wars or interstate con-
flicts. Others, however, are slower moving
and exhibit little variation over time—such
as indicators of governance. For example,
while some countries experience many in-
ternal changes in leadership and coalition
dynamics, the level of governance may not
change much, if at all, except in the cases of
coup d’états [17].

Figure 1 illustrates this point: the blue hori-
zontal line shows that the United States
has remained consistently classified as “free”,
while the language used in the UNGA gener-
debate has fluctuated over time, as shown
by the red line on the second y-axis, which I
will discuss in the next section.

Syntactic and Semantic
Elements of Language

I examine five elements of language to
demonstrate how they convey information
about politics in the international system:
syntactic simplicity, word concreteness,
narrativity, deep cohesion, and referential
cohesion [18]. Syntax simplicity describes
how syntactically simple or complex a sen-
tence, paragraph, document, or corpus is.
This refers to the grammatical structure of
the textbase, where more simple syntax is
more easily understood and less cognitively
demanding on the audience. On the other
hand, complex grammar requires more cog-
nitive effort to parse. Syntactic simplicity can
indicate the relative status of a member of an
organization, as complex syntax can mark hi-
erarchy. For example, a junior member may
use more complex language deferentially to-
ward senior members as a sign of respect.
Complex syntax can also identify in-group
and out-group status.

Word concreteness measures tangibility
and intangibility. Concrete words correspond
to real-life referents (i.e., concrete nouns such
as people, places, and things). Abstract
words can be emotionally evocative, such
as hope, fear, community, terrorism, support,
or peace. Using abstract words can help an
audience connect to broader themes, frame
complex issues, and potentially build con-
sensus without necessarily identifying spe-
cific details.

A narrative text tends to follow the traditional
narrative, storytelling arc: introduction, rising
action, climax, denouement, and resolution.
Information presented in this format is easi-
er to remember than expository (list-like)
presentation as it helps the audience to
contextualize the information within familiar
heuristics. Expository language may instead
present as a list or set of related concepts,
relying on the audience’s working memory to
sort the information.

Deep cohesion can be understood as global
cohesion; in other words, the extent to which
the entire body of information is broadly the-
atically related. Semantically and concep-
tually related ideas may thread throughout
the textbase, but are presented in a more
nuanced and complex format.

Referential cohesion, on the other hand,
is more locally cohesive. Particular words,
phrases, or ideas may be repeated in prox-
imate sentences, or subsequent pronouns
may refer back to antecedent concrete
terms. Texts with high referential cohesion
tend to be more quotable, memorable, or
useful as soundbytes, whereas texts with
depth cohesion often need to be summarized
in order to be communicated succintly.

These five syntactic and semantic aspects of
language combine to form a composite fea-
ture called formality. Low levels of formality
indicate familiarity, shared experiences and
lexicon, and status parity. In short, people
who use less formal language may view
themselves as a part of a social system or
community. High levels of formality may in-
dicate that the speaker perceives himself or
herself as an out-group member—showing
dherence to an in-group with higher status
or hierarchy in the system.

How Democracies Speak

Regime type influences leaders’ use of formal
language in three ways: through institutional
constraints from the domestic bureaucracy,
regimes that exercise oversight from the leader’s trusted inner circle, and
accountability to both domestic and interna-
tional audiences. Democratic regimes are
defined by several political features, includ-
ing meaningful competition for publicly held
positions, an independent judiciary, and regu-
lar and fair elections [18–21]. These features
serve to constrain democratic leaders in their
daily activities, as well as in the internation-
al commitments they make, and the rules of
the political system that influence democratic
leaders’ freedom of speech. Legislatures and
political advisors in open societies function
as checks and balances. Policy change, vetting
changes in foreign policy, tempering hasty
and/or unilateral decisions, and encouraging
consensus-building in the international com-
munity.

Democratic leaders also face more direct
constraints from a trusted circle of advisors,
scholars, and speechwriters. In democratic
regimes, advisors should be less likely than
in authoritarian regimes to blindly concur with leaders’ proposed policies. Rather, they are
more likely to offer candid and contrary opin-
ions about the leader’s foreign policy plans.
Democratic leaders often solicit diverse opin-
ions for policy speeches, which is especially
useful given that they are accountable to a
diverse constituency.

The role of audiences should also be con-
sidered when it comes to choices made by
leaders. For example, democracies are likely
to engage in consensus-building and diplo-
matic persuasion. In 2004, the Japanese
delegation pronounced, “Peace and security,
economic and social issues are increasing-
lly intertwined. The response of the United
Nations must be coordinated and compre-
sensive. UN agencies and organs must be
effective and efficient [23].” Similarly, in 2012,
the UK delegation declared: “The building
blocks of democracy, fair economies and
open societies are part of the solution, not
part of the problem. Nations must be coordi-
nated and comprehensive. International
organizations must step up our efforts to support
the people of these countries as they build their
own democratic future [24].”

These features of democracies’ language
are in part due to the influence of the national
leader’s team of advisors and the extensive
linguistic and ideological vetting that takes
place before the speech is given. It is also
partly due to the distribution of power be-
tween different branches of government; in other
words, the leader generally does not make
international claims, threats, or commitments
without consulting and gaining approval from
advocates and scholars. For example, the legis-
lature, and by exten-
sion, the general public, who can remove
the leader from office for poor foreign policy
performance.

Language in Non-democracies

Leaders from non-democratic countries
generally face different and often fewer

Figure 2. Syntactic Complexity in the World

Figure 3. Word Abstractness in the World

Figure 4. Expository Language in the World

Figure 5. Deep Cohesion in the World

Figure 6. Referential Cohesion in the World

Figure 1. Variation in language and governance over time (Country: USA)
constraints than democratic leaders. Government types and governance can be compared in different ways and along varying lines, including the Polity IV’s ‘scale’ [17], the Freedom House typology [25], and varying approaches to classifying types of non-democracies [26–29]. Map linguistic features onto institutional features to provide a context for interpreting language in the international system. The depth and robustness of public institutions characterize the level of formality in public venues; of the non-democratic regimes, party-based ones have the most bureaucratic accountability, with personalist regimes having among the least [19, 26, 28, 30].

These institutional features are also used to explain other political phenomena, like compliance with international treaties and participation in conflict. Using the typology set out by Lai & Slater [28], political scientist Olga Chyzh evaluates authoritarian regimes’ patterns of signing and complying with international treaties [31]. Of personalist regimes, Chyzh writes, “personalist leaders are effectively not constrained (or almost so) by the need to seek approval on all except for very particular policy issues” [31].

Notably, some authoritarian regimes, like party-based political systems, can display quasi-democratic traits that may make them more likely to speak and behave in a public fashion like democracies. Chyzh provides further useful insight into how these regimes are constrained, writing that, “In contrast, authoritarian leaders with larger domestic bases—oligarchic dictatorships—face decision-making constraints in more policy areas, as larger domestic bases have more diverse interests.”

In addition, domestic institutions, such as cabinets, juntas or politburos, common to oligarchic regimes, tend to institute a status-quo bias making policy change, such as entering into an international agreement, more difficult [31]. For example, the Chinese single-party political system has pseudo-democratic features: French, Spanish, German, Arabic, and Chinese. Linguistic features derived from Coh-Metrix include passive voice, latent semantic analysis, left-embeddedness, and age of lexical acquisition. Coh-Metrix and Coh-MetrixML analyze documents across more than 100 indices, and a principle components analysis reduces the language features to five dimensions: syntax simplicity; word concreteness; narrativity; deep cohesion; and referential cohesion. I discuss these features in more detail below.

The dependent variable from the empirical model comes from the Freedom House qualitative measurement of political rights and civil liberties, and categorizes countries as Free (0), Partly Free (1), and Not Free (2). Figures 2 through 6 show maps of the five categories of syntax and semantics in the world, divided by DoD Geocompmander Commands. In Figure 2, we observe more complex syntax in darker red color. In Figure 3, countries with darker shades of red use more abstract language, while those shaded lighter use more concrete terms. Figure 4 shows countries that use more list-like or enumerative language to convey their messages shaded in darker red, while those whose words follow the "narrative arc" more closely are shaded lighter. Figures 5 and 6 show deep and referential cohesion, respectively.

Table 1 shows the results of a generalized linear model with an ordered logit estimator, using Stata 15 software. Figure 7 shows the marginal effects of the covariates on the dependent variable, holding all covariates at their means. Countries that are more free use simpler syntax, whereas those that are less free tend to use more complex syntax. Similarly, countries that are rated more free tend to have higher levels of deep cohesion, and less free countries have less deep cohesion in their language.

This can be interpreted as follows: simpler syntax and deeper cohesion tend to suggest that there is a great deal of shared meaning among members of an in-group. In this way, individuals with genuine authority and power need not leverage their language to overcome perceptions about their legitimacy. Similarly, individuals who are members of the in-group tend to use more simple language, and in this way both syntax and semantics can provide clues to which members in the international system perceive themselves as members of the in-group, or the out-group.

Figure 8 shows syntax and semantics patterns by country type in the world according to the Freedom House Freedom in the World ratings. More free countries use simpler syntax and have higher levels of deep (or global) cohesion than do partly free or not free countries. Figure 9 represents these relationships geographically: darker blue indicates informal language and in-group language can help us make sense of complex political events, such as elections and candidate behavior; the rise of new regional threats; the emergence and dissolution of alliances, and in-group language will tend to use more complex syntax and semantics in their language. This in-group will tend to use more complex syntax and semantics in their language. This in-group will tend to use more complex syntax and semantics in their language.
Future Directions

Computational discourse analysis may also aid DoD in its efforts to develop automat-ed technologies for the extraction of useful information from very large datasets for defense analysis. For example, in 2012, the Defense Advanced Research Projects Agency (DARPA) began work on a project called Deep Exploration and Filtering of Text (DEFT), which seeks to develop a deep natural language processing architecture for text and audio analysis. According to the DEFT documentation, the program’s purpose is “to find and rep-resent key information, including information on entities, relations, events, sentiment, beliefs, and intentions” from multiple streams of data [35]. Developing a truly automat-ed solution will require additional steps to move from data extraction to intelligence production, and computational discourse analysis can play a critical role in identifying key patterns in behavior, with human analysts playing a critical role in interpreting results from computer-aided analysis [36]. Individual analysts remain the most important part of the analytical process. While computers excel at sorting and categorizing information, humans have the unique ability to contextu-alize, interpret, and communicate that infor-mation.

Acknowledgements

This work was supported by the Minerva Initiative (DoD Award #FA9550-14-1-0038). United Nations for the new era. Statement by the Permanent Mission of Japan to the UN at the Fifty-Ninth Session of the General Assembly of the United Nations. Retrieved from https://www.un.org/japan/Go/content/ jsten/69584921.html.