Language Patterns in International Relations

Leah Windsor, Ph.D.
Languages Across Cultures Lab
Institute for Intelligent Systems
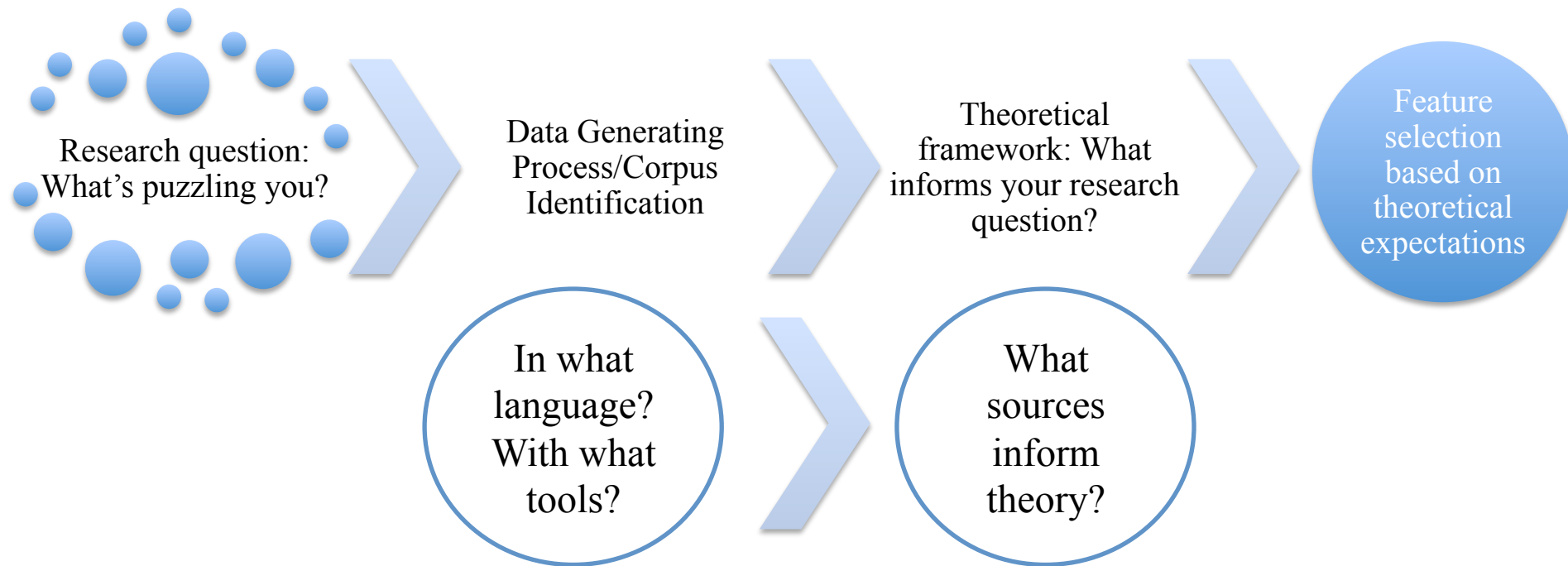April 9, 2019

THE UNIVERSITY OF
MEMPHIS.

How does what we say reveal who we are?

Is it what we say, or how we say it?

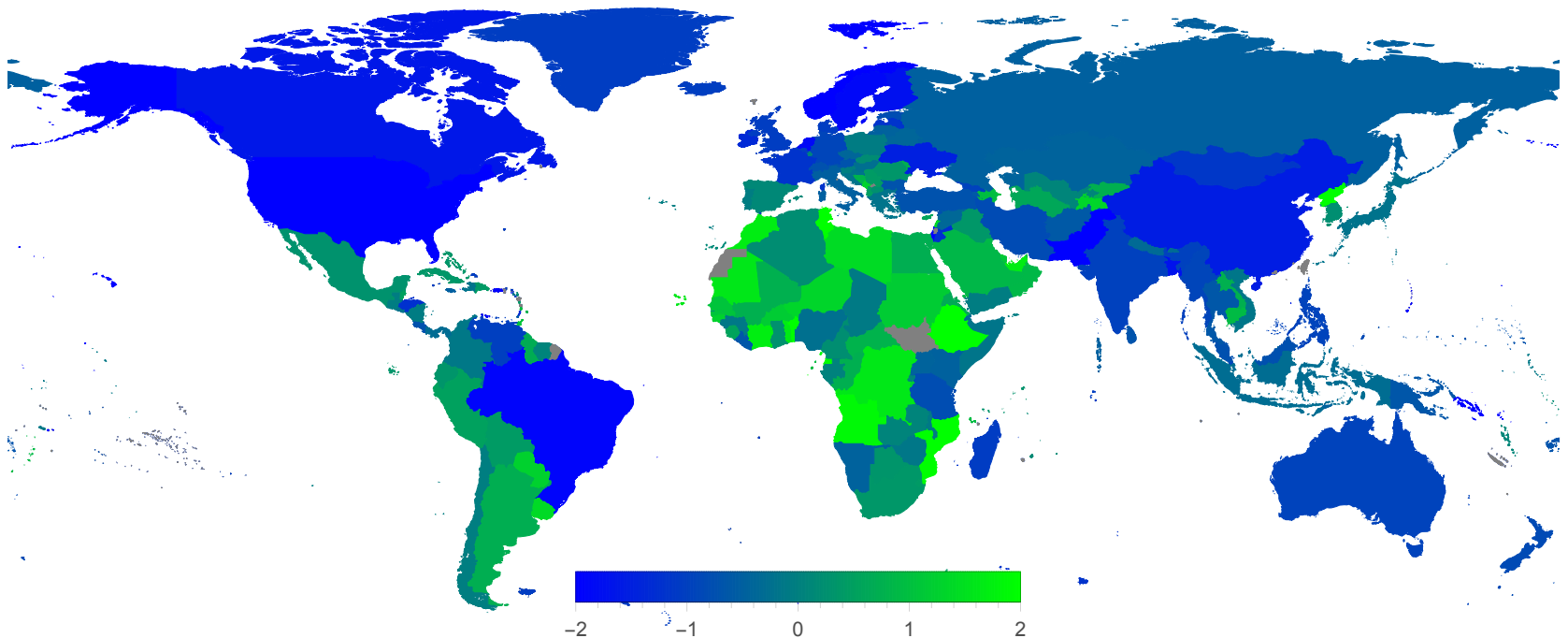# What biases influence our understanding of international relations?

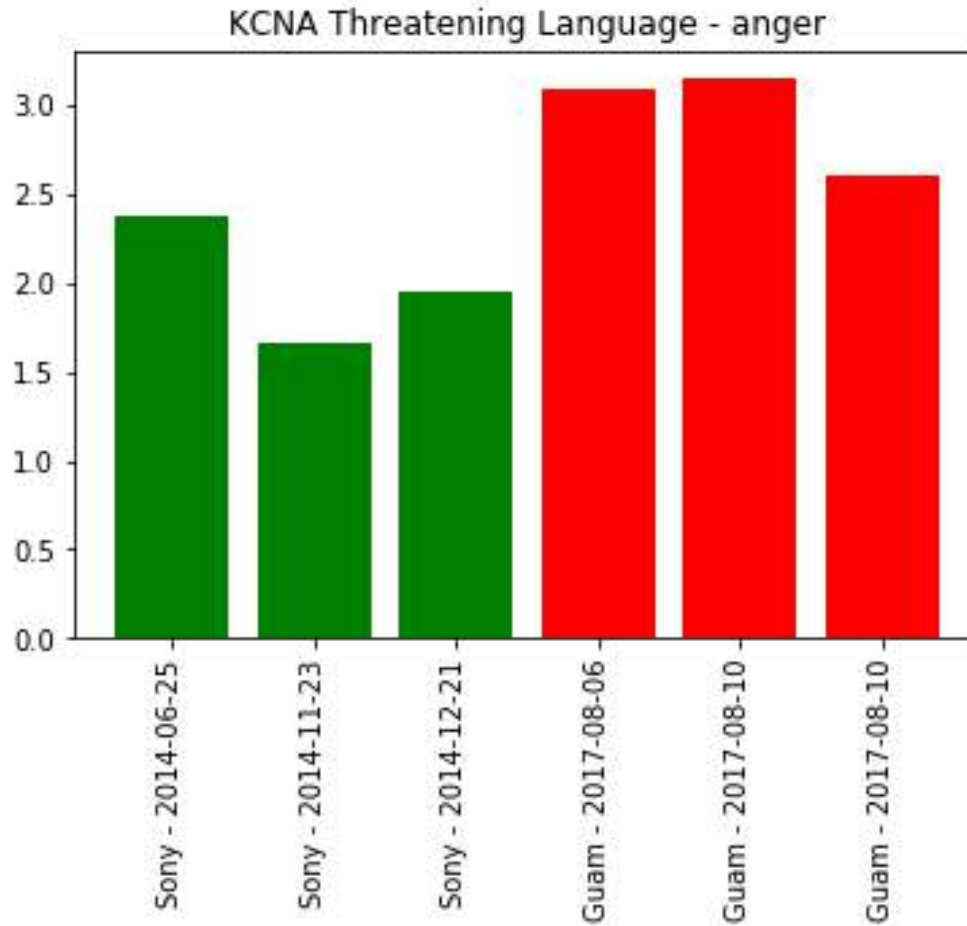# How we learn about the world through text analysis

Research question: What's puzzling you?

Data Generating Process/Corpus Identification

Theoretical framework: What informs your research question?

Feature selection based on theoretical expectations

In what language? With what tools?

What sources inform theory?

Language shows country-level differences

Average Std. Formality by Country

Learning about politics from words: Forecasting

KCNA Threatening Language - anger

Learning about politics from words: Hard-to-observe groups

Stayed in Power

- King Hamad bin Isa Al Khalifa (Bahrain)
- King Abdullah (Jordan)
- Bashar al-Assad (Syria)
- King Mohammed VI (Morocco)
- Abdullah bin Abdulaziz Al Saud (Saudi Arabia)

Lost Power

Twice as many negative emotion words, more risk-associated words, and more third-person pronouns

- Zine al-Abidine Ben Ali (Tunisia)
- Muammar Gaddafi (Libya)
- Ali Abdullah Saleh (Yemen)
- Hosni Mubarak (Egypt)

More positive emotion words, and fewer negative emotion, anxious, and angry words

Institute for Intelligent Systems
THE UNIVERSITY OF MEMPHIS

Sources of Bias in Text-as-Data

- *How do we learn about political processes in opaque societies and extremist groups?*
- *How do we know what we know about the world?*
  - Sources describe events and ideas differently based on language, culture, geography, etc.
- *Research about language and politics broadly has deficiencies which affect conclusions and generalizability of findings, including:*
  - Information deficiencies
  - Document processing
- *Why it's important to dig and work for broader source texts*
  - Relatively few groups and regimes dominate policy space
  - Generalizability, via multilingual computational tools

Corpus generation

Information deficiencies

Linguistic pipeline

Corpus selection bias
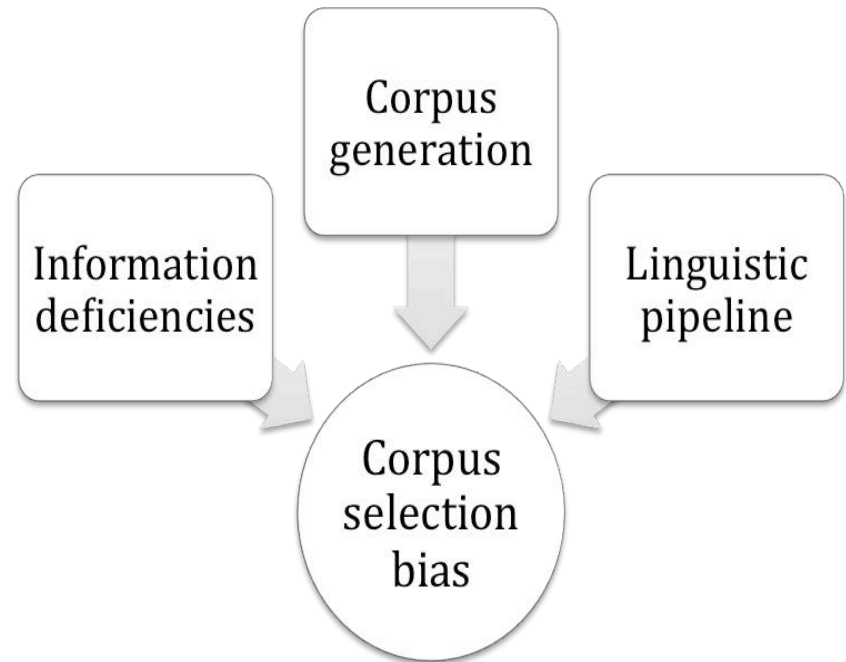
Topic Modeling:
Insights on the Rwandan Genocide
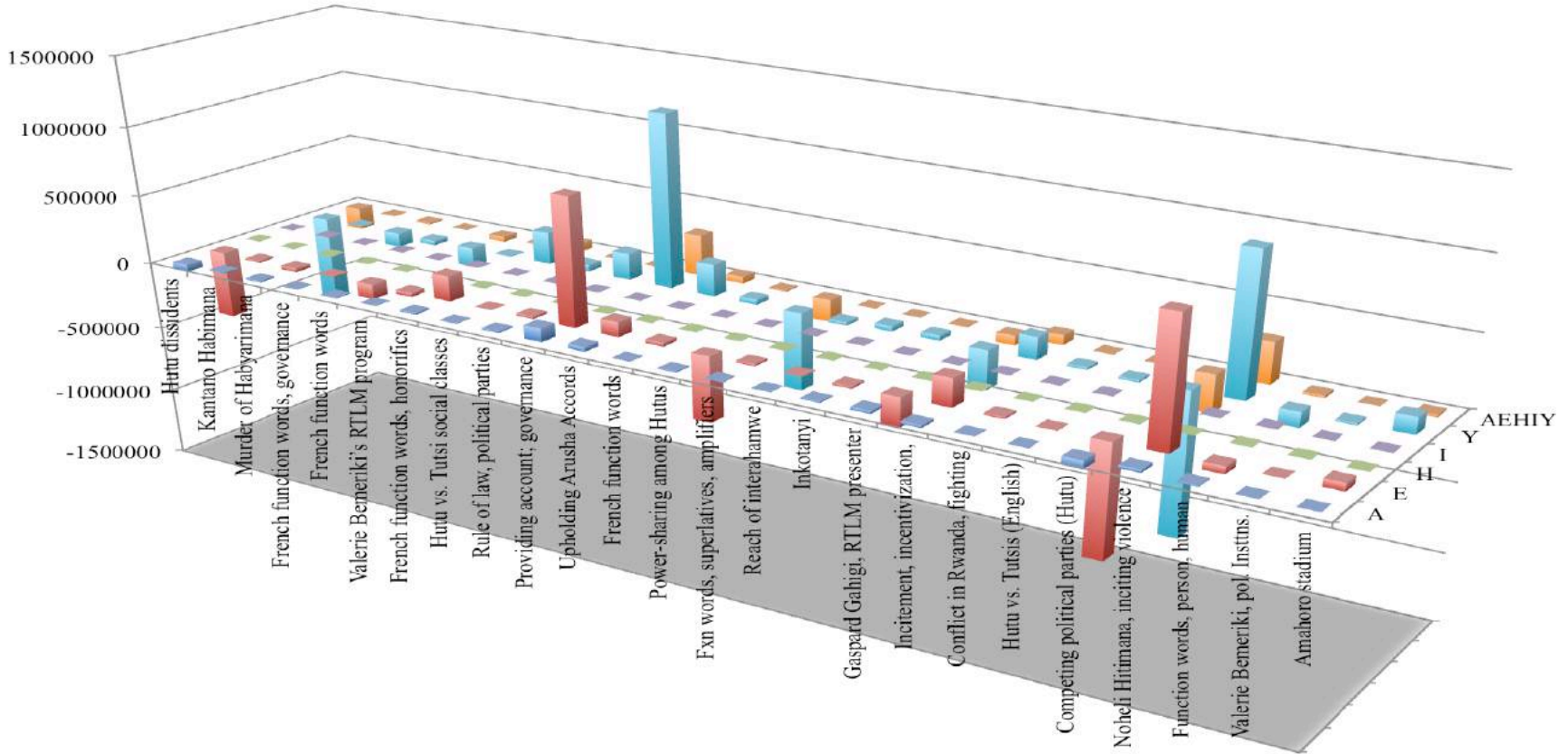
De-biasing Linguistic Analysis: Document Processing

- Archived in user-unfriendly formats
- Automated document processing technology exists but requires labor-intensive human supervision
- Linguistic and translational barriers
  - Some "bag of words" programs are language-agnostic, such as topic modeling
  - Windsor et al., (under review) finds very small effect sizes for dictionary-based programs between human and computer-translated documents

THE UNIVERSITY OF MEMPHIS.

Language and culture shape how we see the world (What we say reveals who we are)

Culturally Implicit Associations: Differing Perspectives on the Same Phenomenon
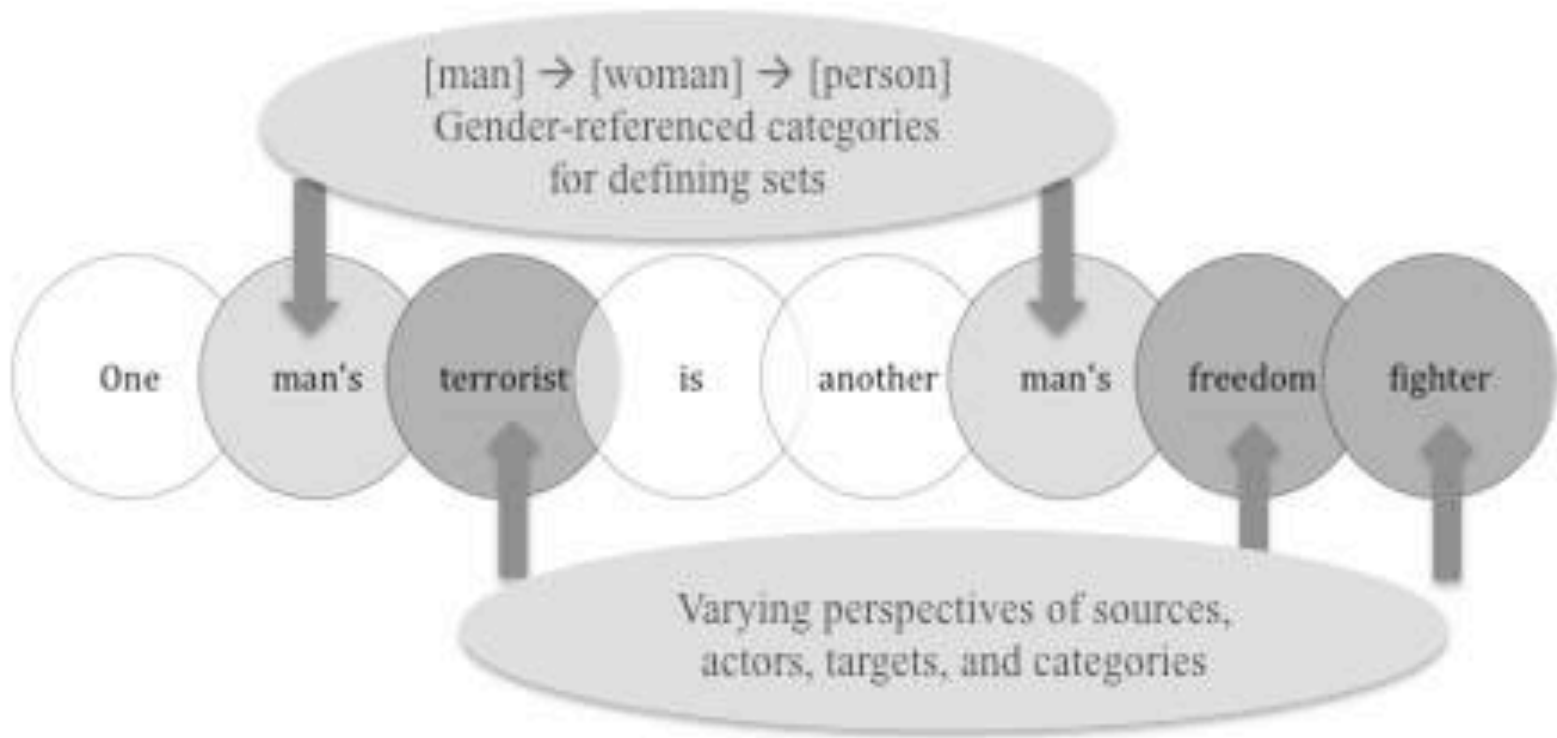
It's what you say *and* how you say it
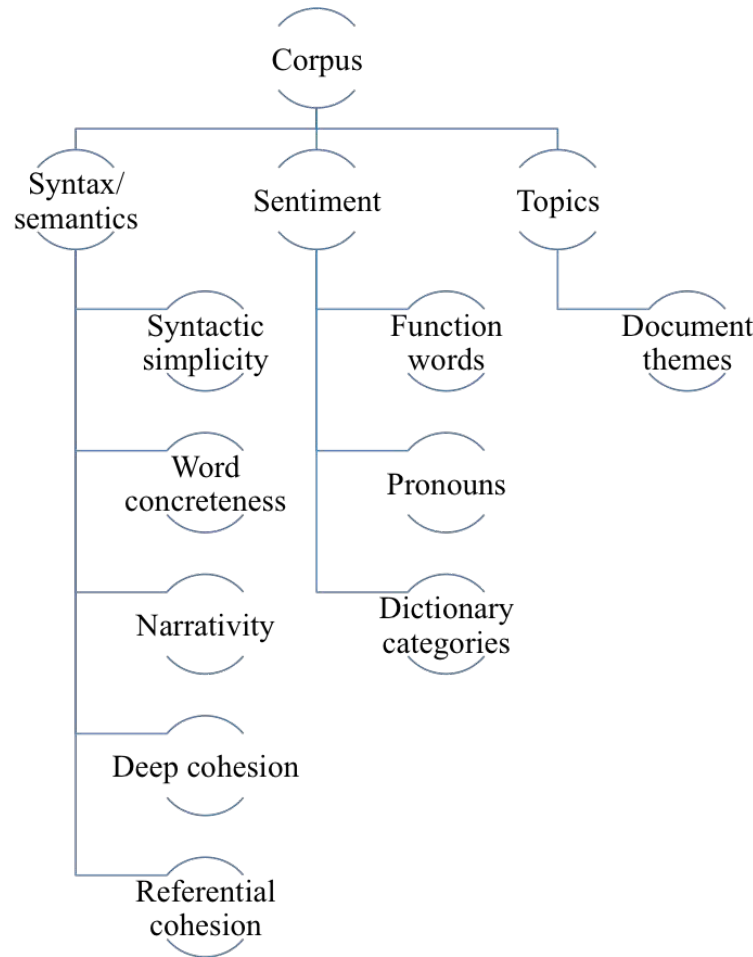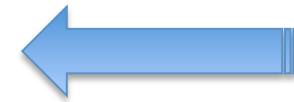
Syntactic simplicity/complexity

connor

@connorhannigan4

text: i can't
email: i cannot
essay of 2000 words: i am unable
to can

4:06 · 26 Feb 19 · Twitter for iPhone

41.1K Retweets   **191K** Likes

Word concreteness/abstractness

*"China will enhance **friendly** relations and **cooperation** with all other **countries** on the basis of the Five **Principles** of **Peaceful Coexistence** and strive to build a new type of **relations** between major countries based on mutual **respect** and win-win **cooperation**. Following a policy of building **good-neighborly relationships** and **partnerships** with neighboring countries, China has actively expanded exchanges with countries in its **neighborhood**."*

-UNGA General Debate, Chinese Representative, 2012

Institute for Intelligent Systems
THE UNIVERSITY OF MEMPHIS

THE UNIVERSITY OF MEMPHIS

Narrativity vs. Expository

Therefore, I direct all department secretaries and the heads of agencies **to reduce requirements** *and the* **processing time of all applications**, *from the submission to the release. I order all department secretaries and heads of agencies* **to remove redundant requirements** *and* **compliance with one department or agency**, *shall be accepted as sufficient for all. I order all department secretaries and heads of agencies to* **refrain from changing and bending the rules government contracts, transactions and projects already approved** *and awaiting implementation. Changing the rules when the game is on-going is wrong. I abhor secrecy and instead advocate transparency in all government contracts, projects and business transactions from submission of proposals to negotiation to perfection and finally, to consummation.*

-Rodrigo Duterte, 2016

Deep cohesion

- This dimension reflects the degree to which the text contains causal and intentional connectives when there are causal and logical relationships within the text.
  - These connectives help the reader to form a more coherent and deeper understanding of the causal events, processes, and actions in the text.
  - When a text contains many relationships but does not contain those connectives, then the reader must infer the relationships between the ideas in the text.
- If the text is high in deep cohesion, then those relationships and global cohesion are more explicit.
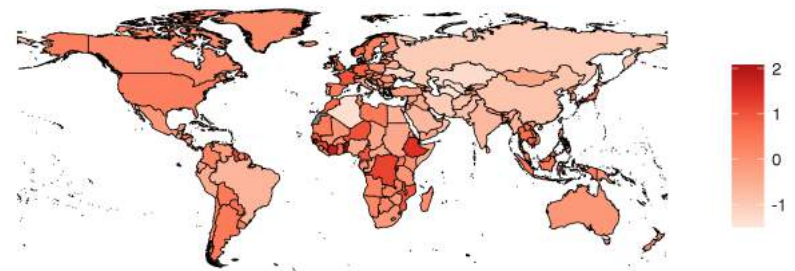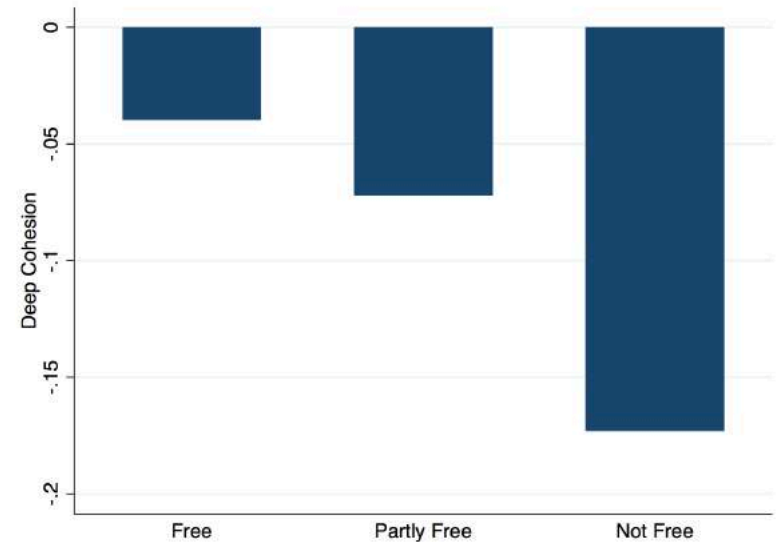
Referential cohesion

*"There are **moments** we can describe as historical turning points, **when nations** and **peoples** must decide where they **stand**. This is one of these **moments**, **when** history will judge us as **leaders**, and examine if we were democratic **leaders** that represented the will of our **peoples**. It is clear that the **people** of the **world** are taking a **stand**, **against** neoliberal economics and **war**. **They** are **fighting against** those who would impose their will by **military** and economic **force**. **They** are resisting those who would undermine and even overthrow the basic principles that founded the United Nations itself."*

-UNGA General Debate, Venezuelan Representative, 2004

Sentiment Analysis: Pronouns as language truth-o-meter

"Hey very sorry. Meant no harm. And I apologize for whatever I may have said or what was said. And my overtures are strictly professional. Again I apologize didn't mean to be inappropriate at all. Sorry that impression was left."

-email by Harold Ford, Jr.
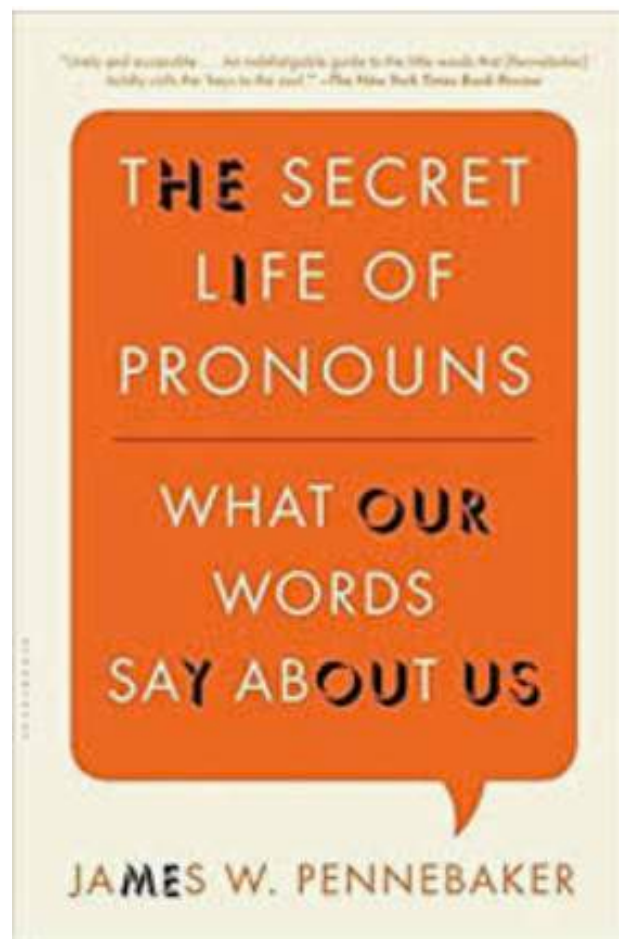
Why function words matter

- UNSC Resolution 242 controversy:
  - Withdrawal of Israeli armed forces from territories occupied in the recent conflict
  - Retrait des forces armées israéliennes **des** territoires occupés lors du récent conflit.

Minerva 2018 Project: Capstone and Coh-MetrixML

Source text
• Pre-processing

Language
• Identification

Sentence Boundary Disambiguation

Syntactic Feature Identification
• Tokens/Tags
• Lemmatization
• Polysemy/hypernymy
• Latent Semantic Analysis

Aggregate Syntactic Features
• Word Concreteness
• Syntactic complexity
• Cohesion
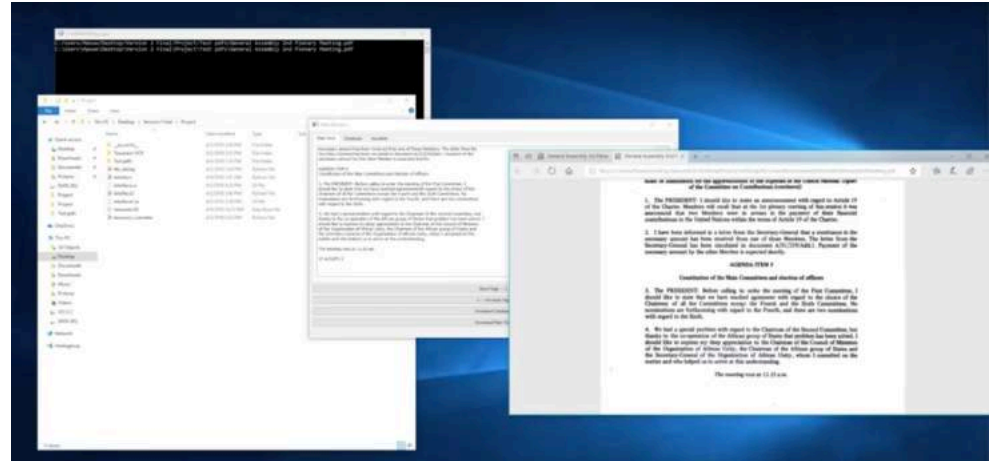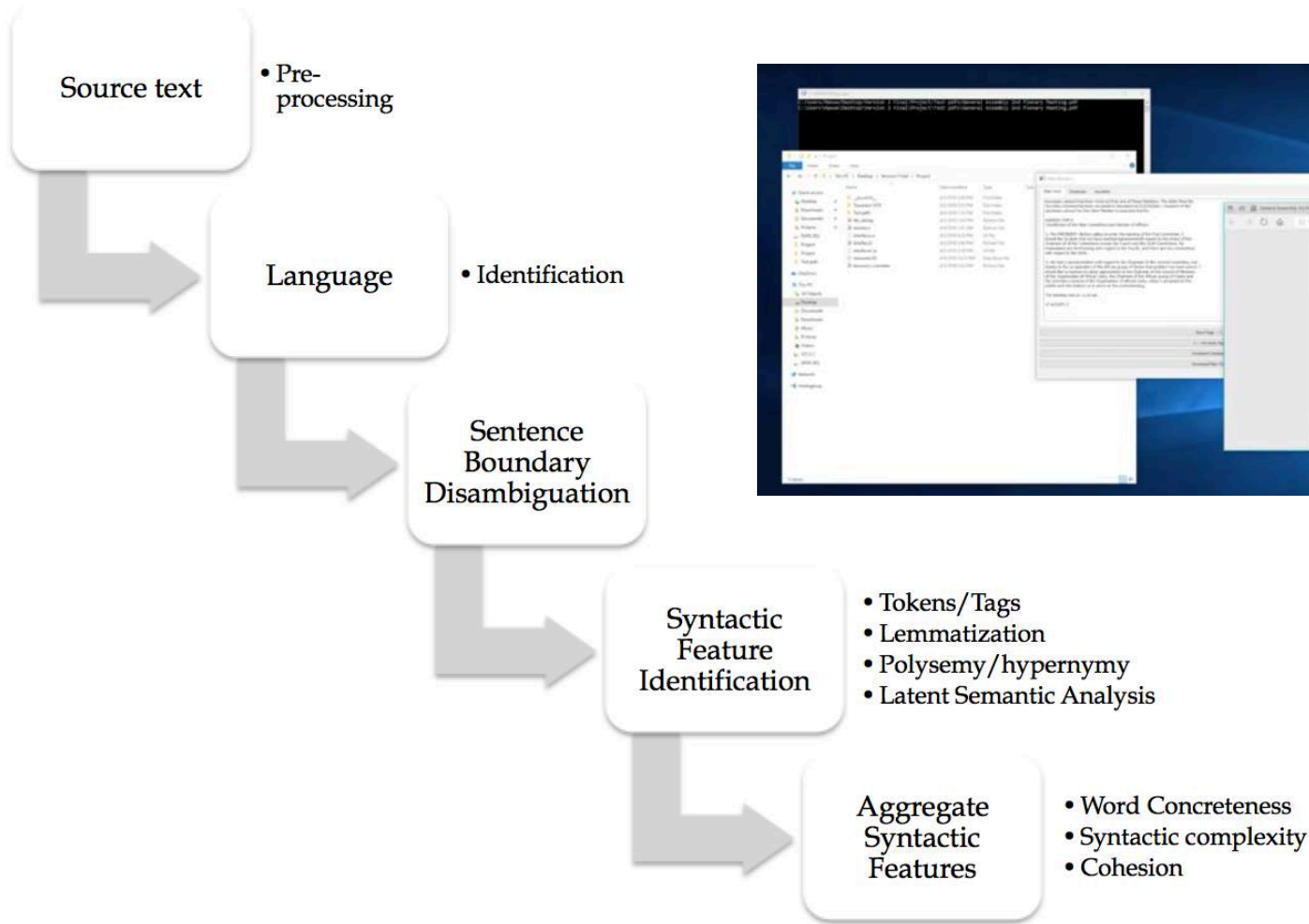
Coh-MetrixML
(ML = multi-lingual)

- This program has four major component: base, CoreNLP, WordNet and LSA.
  - The CoreNLP component is responsible for paragraph and sentence splitting, sentence parsing, named-entity recognition and syntax related measures.
  - The WordNet component is responsible for providing word information, such as word lexical type, polysemy, hypernymy, etc.
  - The LSA component is responsible for providing cohesion measures.
- In Chinese, French, Arabic, Spanish, German, and possibly Russian (extending beyond English-only)

Coh-MetrixML Advantages

- This program will allow researchers to analyze syntactic and semantic features in the original language.
  - Circumvents translations
  - Broadens the domain of corpora available for scholarly inquiry
  - Maintains fidelity to original intent/meaning of the corpus.
  - Can be aligned with other text processing methods

Future directions

- Evaluating systematic biases in language (similar to known identity-based biases in AI) and implications for social and political processes

- Deploying Coh-MetrixML for public use

- Recruiting participation from a broader network of global scholars and multilingual text analysis applications via QuantText

https://QuantText.com

Thank you!

Leah.Windsor@memphis.edu
http://leahcwindsor.com
http://quanttext.com

Institute for Intelligent Systems
THE UNIVERSITY OF MEMPHIS

THE UNIVERSITY OF
MEMPHIS.