

# Penn State Political Science Team

- Pennsylvania State University
  - Department of Political Science
  - Quantitative Social Science Initiative (QuaSSI)
    - Big Data Social Science / Social Data Analytics training initiatives
  - Statistical Analysis of Political Text projects
    - Event Data Projects (Schrodt)
    - Legislative Speech Projects (Monroe)
- Core Team:
  - Philip Schrodt (Lead Investigator)
  - Burt Monroe

# Core Team Experience and Capabilities

## **Philip Schrod**

- Political Instability Task Force (multi-agency)
- Integrated Conflict Early Warning System (DARPA)
- Joint Warfare Analysis Center (DoD)
- 12 NSF grants, mostly for political forecasting
- Fellow and Past President, Political Methodology Society
- Published applications of computational methods include hidden Markov models, reverse-Wolfram models, genetic algorithms, neural networks, sequence and cluster analysis methods, as well as conventional time series

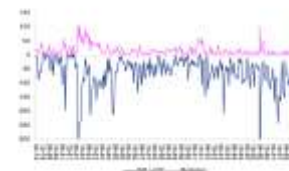
## **Burt Monroe**

- 4 NSF grants (>\$1m current), mostly in political text analysis, newest for Arabic in Middle East and North Africa
- Director, Penn State Quantitative Social Science Initiative (QuaSSI)
- Director, Penn State Big Data Social Science (BDSS) & Social Data Analytics (SoDA) training initiatives
- Winner, Gosnell Prize for Excellence in Political Methodology
- Published applications of computational and statistical methods in topic modeling, feature extraction, information reduction, discussion dynamics

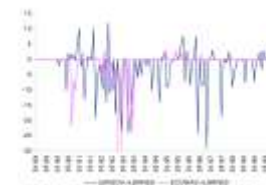
# Schrodt Core Project: TABARI

- Text Analysis by Automated Replacement Instructions
  - Automated coding of political events from newswire sources
- Coded 3-million events from 26Gb of text ICEWS-I
  - models based on TABARI data were the only ones passing the Phase I benchmarks
- Codes 5,000 events per second and easily adapted to parallel processing.
- CAMEO coding ontology
  - 15,000 phrase event ontology has proven highly robust in ICEWS tests
  - Flexible actor ontology handles sub-state and transnational actors
- Open source C++ (possible enhanced version in Python)
- Ancillary programs for named-entity recognition and rapid dictionary development
- <http://eventdata.psu.edu>

Israel-Lebanon: Conflict and mediation  
1979-98



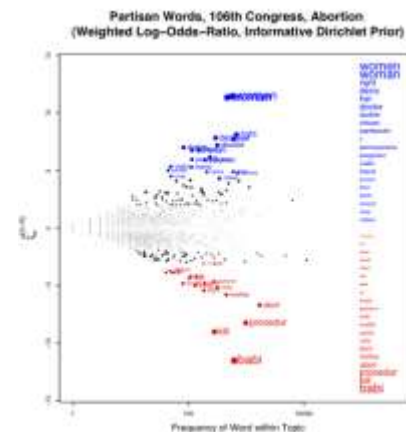
Goldstein series for Liberia and ECOWAS  
actions toward rebels, 1989-99



# Monroe Core Project: Legislative Speech

- Modeling of political landscapes / attention / behavior from web and digital archives of parliamentary texts
- Current speech data encompass 20b+ words in 40+ languages
- Ancillary data on politicians, legislation, parties, elections, named entities including places, events
- Ancillary software for tasks like extraction of Arabic from pdfs
- Open source text processing and statistical software (Python, R, C++) for core NLP and modeling tasks

Negotiations	Jobs	Religion / Identity	Education	Housing/ Immigration
אוסלו	העבודה	בשבת	החינוך	הקליטה
ערפאת	והרוחה	המשטרה	הספר	דירות
הפלשתינים	עובדים	הדתות	והתרבות	דיוור
ירושלים	התעסוקה	השואה	המורים	העלייה
השלוש	שכר	היהדות	התלמידים	העלים
Oslo	work	Saturday	education	absorption
Arafat	welfare	police	school	apartments
Palestinians	workers	religions	culture	housing
Jerusalem	employment	Holocaust	teachers	immigration
peace	wages	Judaism	students	immigrants



# Contact Information

- Philip Schrodtt
- Professor, Political Science
- Pennsylvania State University
- [schrodtt@psu.edu](mailto:schrodtt@psu.edu)
- 814-863-8978
- <http://eventdata.psu.edu/>