

Computational linguistics, critical discourse analysis, and contemporary Korea:

A digital dialogue between humanities and social sciences

Professor PARK, BYUNG-SUN

- ▶ Professor and Chair of the Department of Korean Language and Culture, NCCU
- ▶ Ph.D., Computational Linguistics, Korea University

Text-as-data has become increasingly popular in political science and international relations. It is essential for text analysis and natural language processing to have linguistic fundamentals, which will be delivered in this guest lecture. A critical discourse analysis approach will be also introduced in the context of the humanities so that we can better understand contemporary Korea.

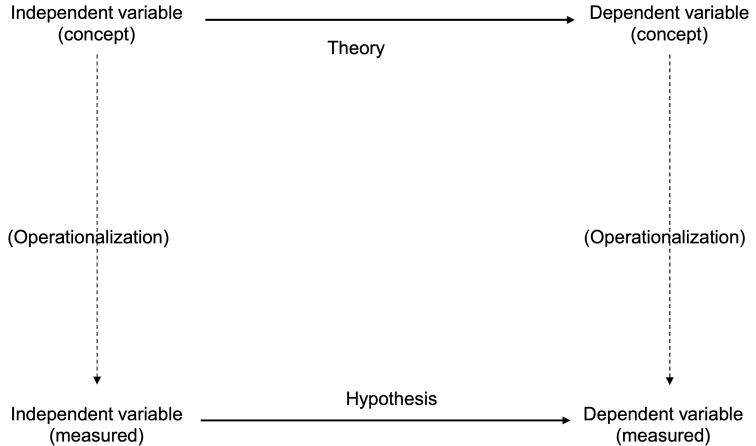
Today

- ▶ A long introduction on the status of discipline: political science and text-as-data
- ▶ A toy example: text analysis during the pandemic in Taiwan and South Korea
- ▶ Professor Park on computational linguistics, critical discourse analysis, and more

Typical research process in political science

1. Theory and hypothesis about 'y' and 'x'
2. Check the data and measurement
3. Get to know 'y' and identify a stochastic component of relevant regression model
4. Define a systematic component of statistical model of 'x'
5. Estimate it
6. Interpret and present the result.

Measurement is needed



Text analysis: old and new

- ▶ “Systematic, objective, quantitative analysis of message characteristics” by Neuendorf (2002, p1)
- ▶ Different types of content analysis: human coding, supervised, and unsupervised
- ▶ Sources: almost everything, but I'm familiar with speech, bill content, parliamentary debate, newspaper, SNS content, ...
- ▶ Type of outputs: ideal point, (un)known category classification, sentimentality, dimensionality, ...

Example

Text-based ideal point

- ▶ The location of political actors, e.g. politicians, parties, countries, or SNS influencers along one or two dimensional space
- ▶ Proksch and Slapin (2010) apply their Wordfish software to EU parliament

Example

Classification

- ▶ Mostly using Latent Dirichlet Allocation, these days
- ▶ Titus C. Chen on PRC

New frontiers

- ▶ Incorporate within potential outcome framework of causal inference, e.g., Egami et al, 2022
- ▶ Unfolding something that we do not know
- ▶ How about voice or video as data?
- ▶ Get along with machine learning, and then, are we still within the disciplinary boundary? How about computational social sciences or digital social sciences?