Introduction to NLP and Social Science

Case Studies: NLP in Social Sciences

2022L-WB-NLP, 24.02.2022

MI EDUCATION

Stanisław Giziński

https://github.com/MI2-Education/2022L-WB-NLP

s.gizinski84@gmail.com

Points

100 points overall

- 45 work during project
  - 9 for each milestone / homework (5 milestones/homeworks)
- 3 for activitiy
- 32 final reports
- 16 final presentation
- 4 good GitHub practices

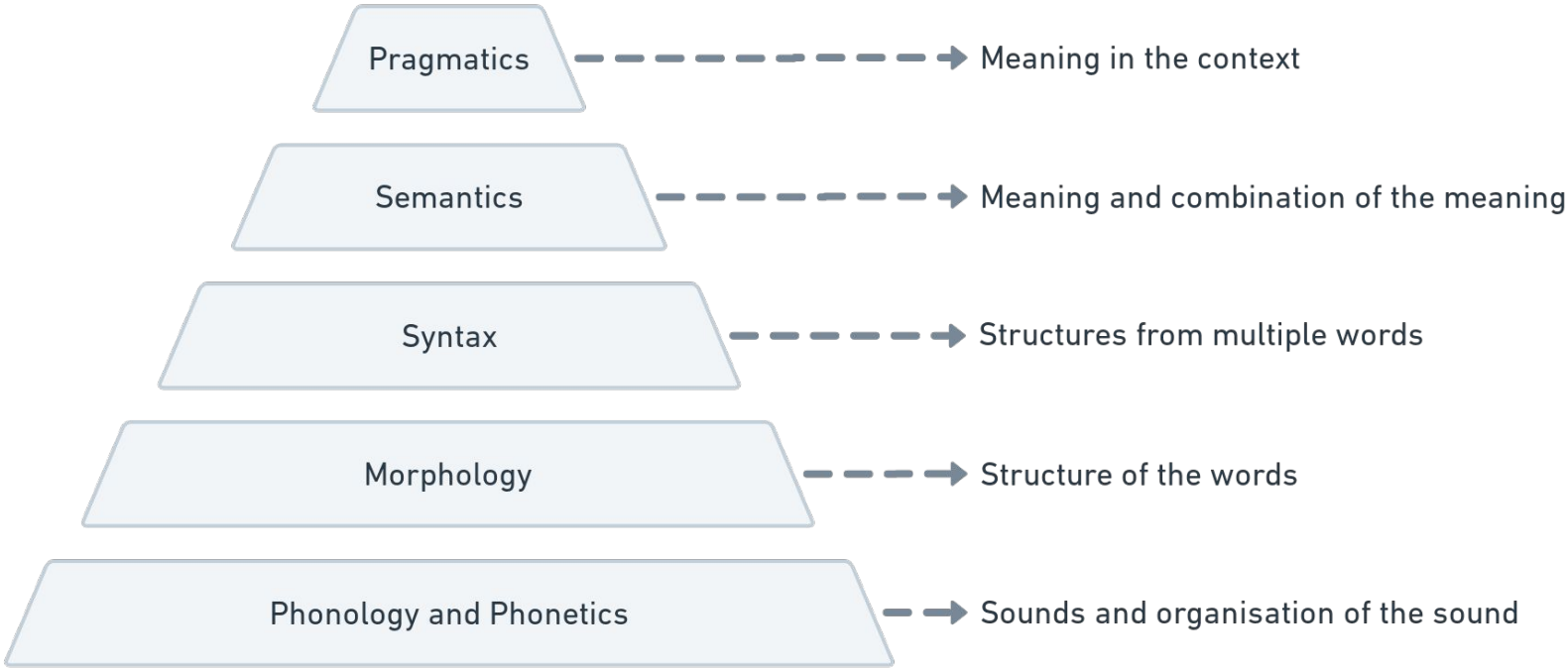
Natural Language Processing

What is **natural language processing**?

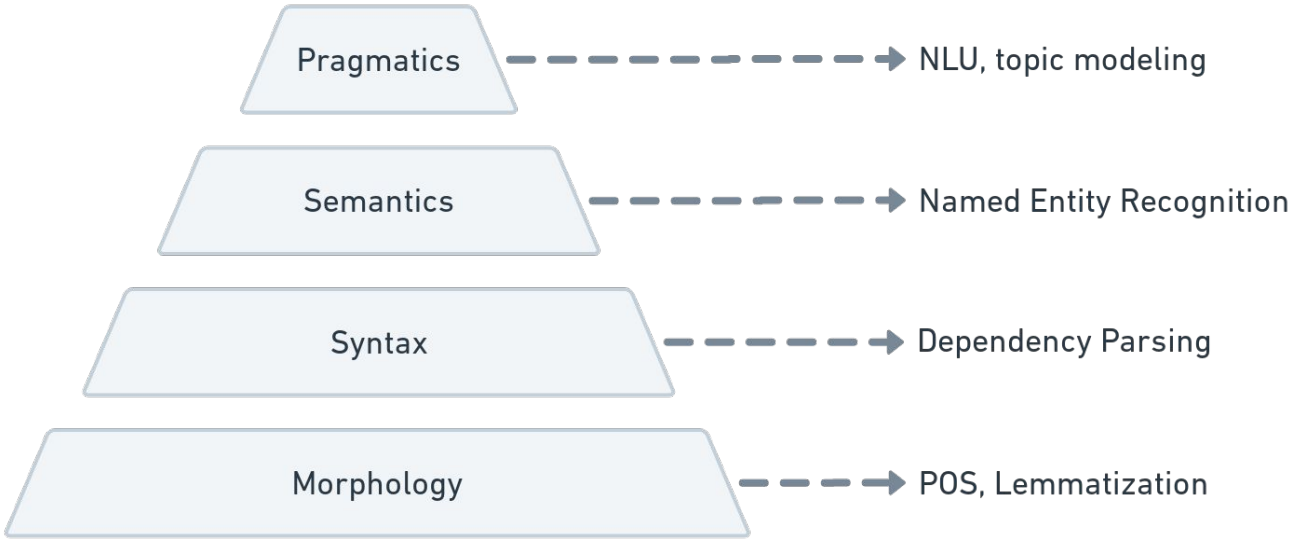
Natural language – any language that has evolved naturally in humans through use and repetition without conscious planning or premeditation.

Main source of problems: **ambiguity**

Levels of human language understanding

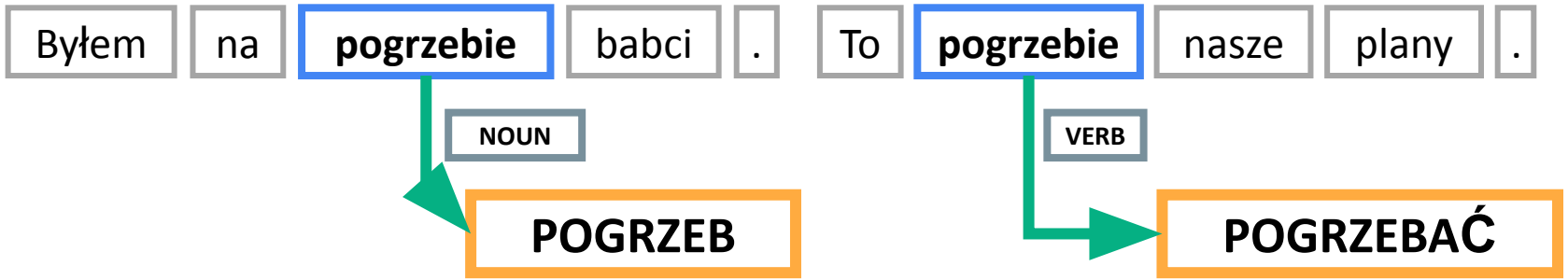


Which each level, there are NLP tasks associated



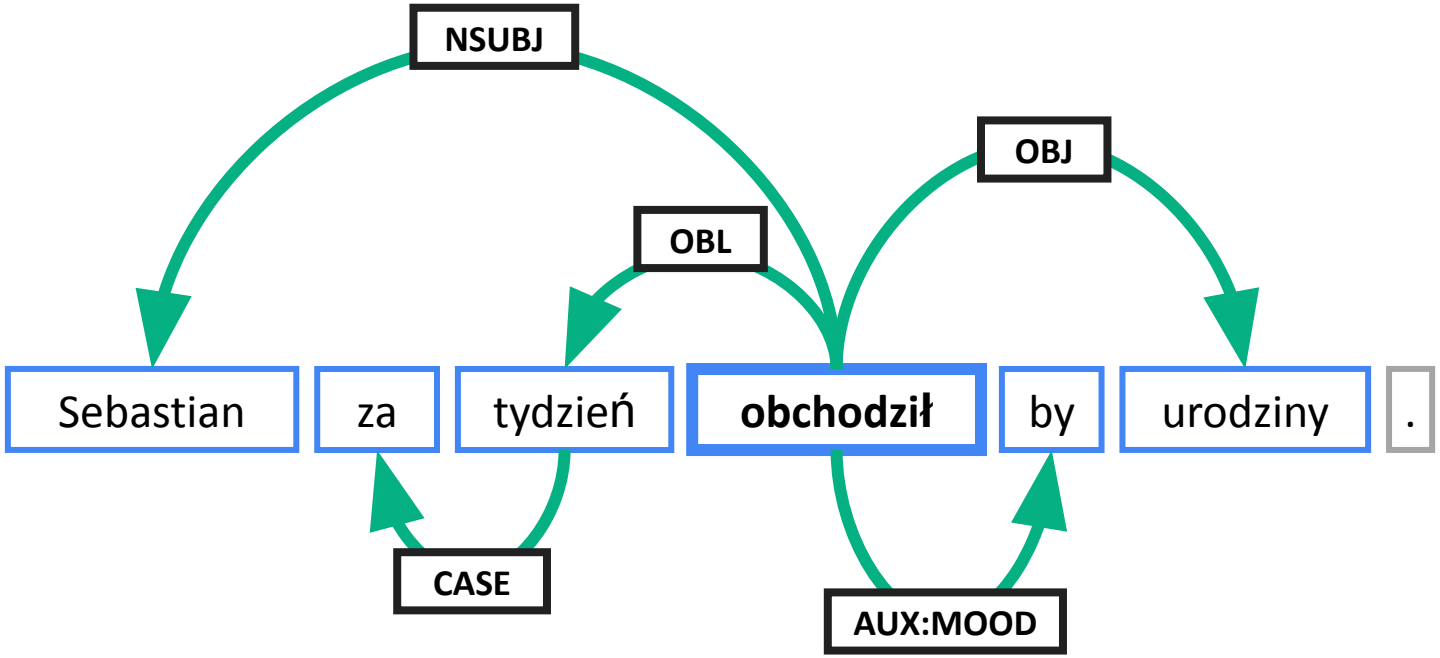
Lemmatization

Converting word to its *lemma* (dictionary form)



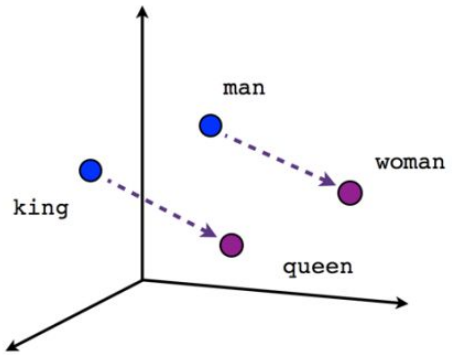
Dependency parsing

Machine-interpretable semantics

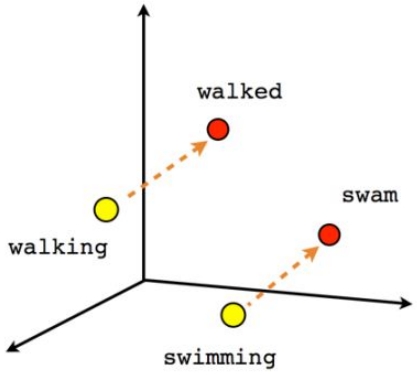


Word vectors

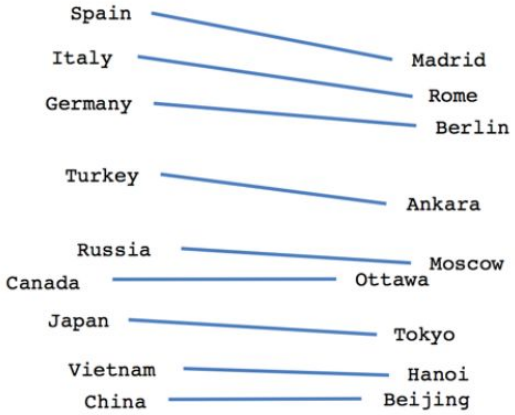
Machine-interpretable semantics



Male-Female



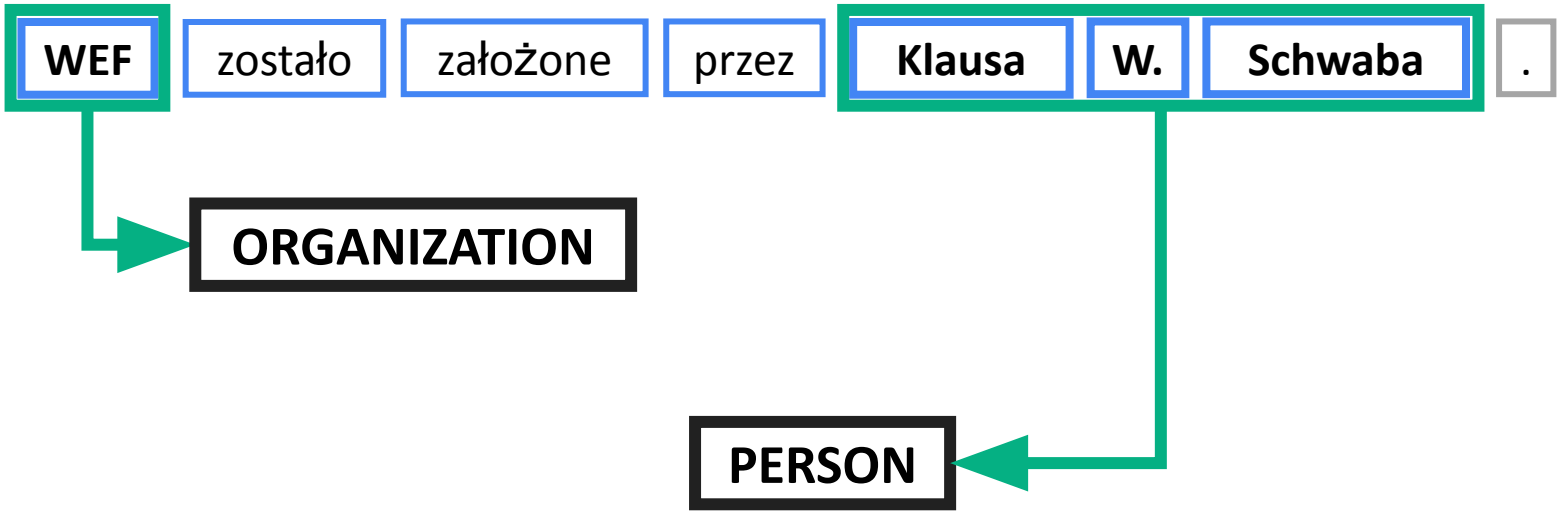
Verb tense



Country-Capital

Named Entity Recognition

Classifying text spans as specific classes of entities



Topic modeling

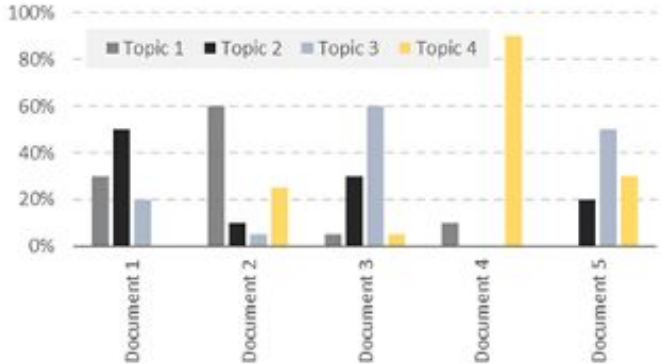
Documents

LDA

Creation of topics

	weight	words
Topic 1	3%	flower
	2%	rose
	1%	plant
...		
Topic 2	2%	company
	1%	wage
	1%	employee

Topics allocation to documents



Social science, political science

Social science investigates how people construct reality.

Political science investigates how power – ability to maneuver agents to act in a specific way – express itself in reality.

Public policy investigates how **the state** cooperating with **other agents** solve problems in the **public sphere**.

Research process

1. Research question
2. Hyphothesis formulation
3. Hypothesis testing

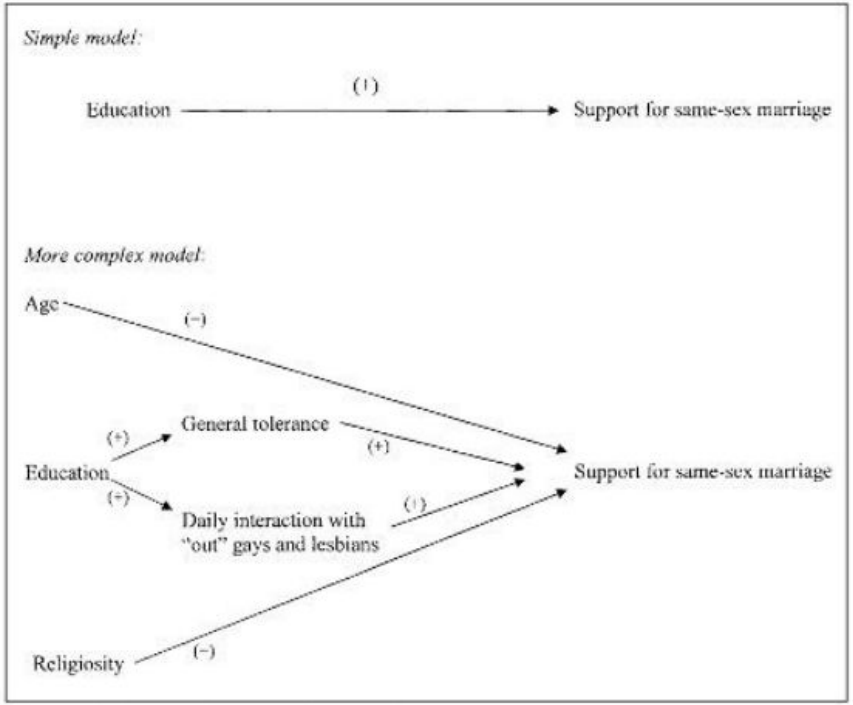


Figure 3.1. Diagramming a causal argument.

Example 1: advocacy coalitions

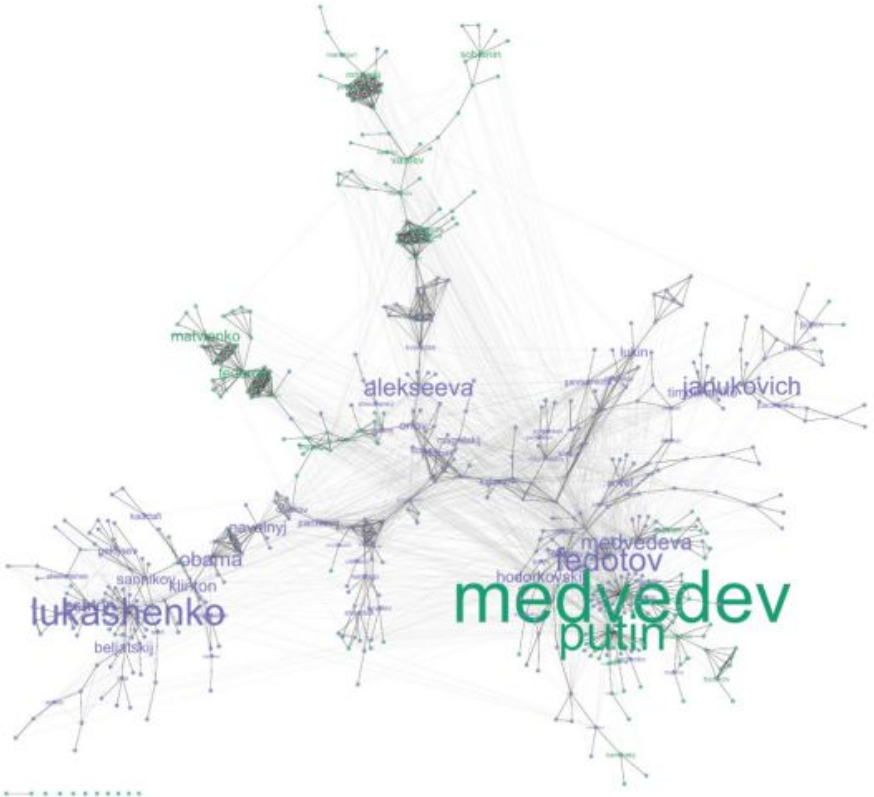
Hyphothesis:

“The first coalition developed a policy that supported GONGO and those NGOs that did not raise any political and policy problems as an alternative to the governmental agenda. The second coalition developed a policy that supported independent NGOs articulating alternative political and policy agenda. We assume that this dynamic became visible in 2011, when the Bolotnaya protests took place, and led to the further stiffening of the policy towards civil society”

Testing:

“We used media texts from TV, radio, newspapers, and the Internet as the unit of analysis. We downloaded 10 000 news articles (maximum available for download) from various mass media sources (both online and print) from year 2011. (…) The articles were found in the archive with search request “civil society” (in Russian).”

“ (…) we have identified names of possible actors using a named entity recognition”


matvienko
alekseeva
jadukovich
navalnyj
obama
saakikov
klinton
lukashenko
belyatskij
medvedeva
fedotov
hodorkovskij
medvedev
putin

Example 2: policy diffusion

Question

How adopting policy in one state influences policies in other states over time?

Main hypothesis

Policy adoption in one state influence the discussion of the issue in other states.

Example 2: policy diffusion

Hypothesis

1. Diffusion: Adoption by other states predicts policy frames in the state

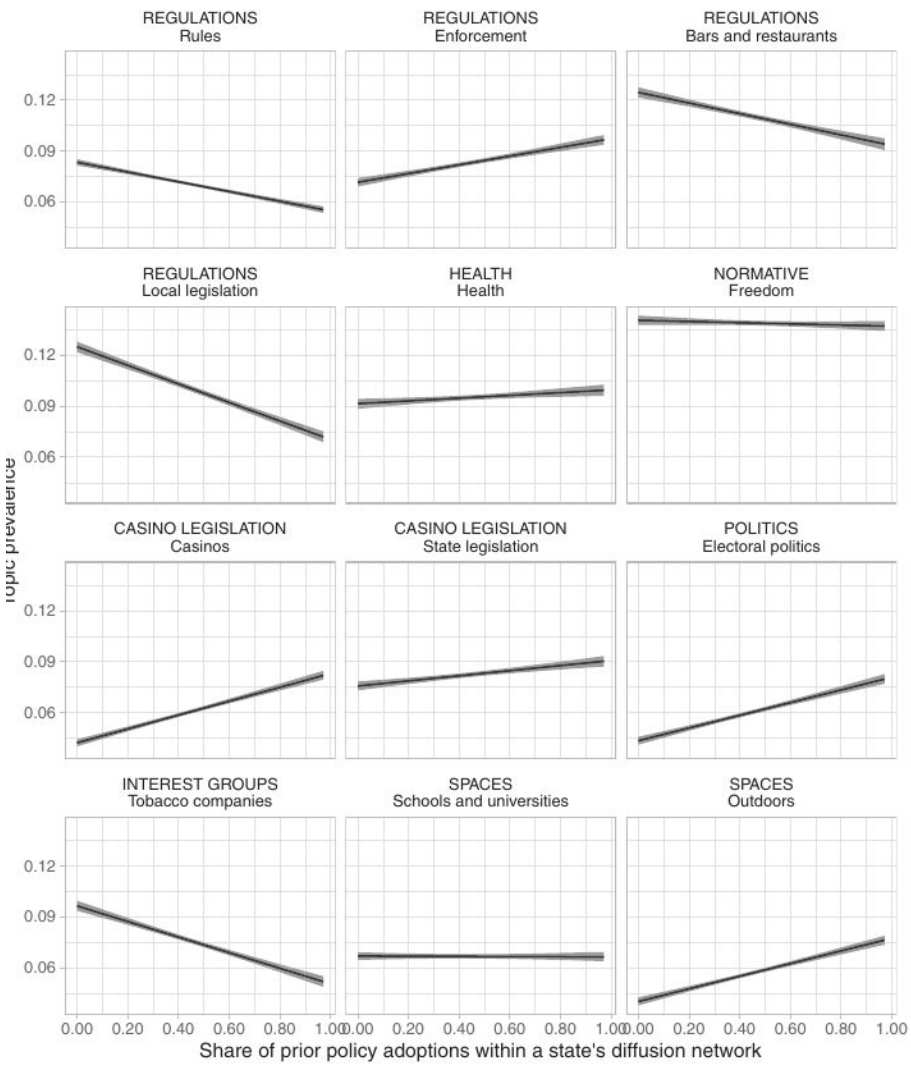
2. Learning: Adoption by other states predicts policy frames based on practical, empirical consequences

3. Emulation (imitation): There is a negative correlation between policy adoption by other states and frames based on narrative arguments.

Testing

49 newspapers from 49 states (1996-2013)

Topic modeling with **covariates**



Example 3: political agendas exploration

Exploring the political agenda of the European Parliament using a dynamic topic modeling approach [Greene & Cross, 2017]

"This study takes up the challenge of extracting latent thematic patterns in political speeches by developing a dynamic topic model to investigate how the plenary agenda of the EP has changed over three parliamentary terms (1999–2014)."

- How the political agenda of the EP has evolved over time and reacted to stimuli in the period 1999-2014
- They show how a dynamic topic modeling approach, based on Non-negative Matrix Factorization is better suited than LDA
- Able to capture the attention of EP to external events (e.g., the Euro Crisis)

Example 3: political agendas exploration

Exploring the political agenda of the European Parliament using a dynamic topic modeling approach [Greene & Cross, 2017]

A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases [Grimmer, 2010]

- Measure how US senators explain their work in Washington to constituents using a collection of over 24,000 press releases from senators from 2007.
- The Expressed Agenda Model measures priorities of each author
- Ideal for comparing priorities

Bibliography

[1]“Computational Analysis of Political Texts.” Accessed: Feb. 24, 2022. [Online]. Available: https://docs.google.com/presentation/d/1Pm2obVYPjruc-zR2URnNVd5ndtAek2wwPn4JpX-Svx8

[2]D. Greene and J. P. Cross, “Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach,” Political Analysis, vol. 25, no. 1, pp. 77–94, Jan. 2017, doi: 10.1017/pan.2016.7.

[3]J. Lyons, Natural Language and Universal Grammar: Essays in Linguistic Theory, vol. 1. Cambridge: Cambridge University Press, 1991. doi: 10.1017/CBO9781139165877.

[4]C. Grabow, J. Macinko, D. Silver, and M. Porfiri, “Detecting causality in policy diffusion processes,” Chaos, vol. 26, no. 8, p. 083113, Aug. 2016, doi: 10.1063/1.4961067.

[5]J. Grimmer, “A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases,” Political Analysis, vol. 18, no. 1, pp. 1–35, ed 2010, doi: 10.1093/pan/mpp034.

[6]F. Gilardi, C. R. Shipan, and B. Wüest, “Policy Diffusion: The Issue-Definition Stage,” American Journal of Political Science, vol. 65, no. 1, pp. 21–35, Jan. 2021, doi: 10.1111/ajps.12521.

[7]D. Zaytsev, N. Talovsky, V. Kuskova, and G. Khvatsky, “The Entity Name Identification in Classification Algorithm: Testing the Advocacy Coalition Framework by Document Analysis (The Case of Russian Civil Society Policy),” in Analysis of Images, Social Networks and Texts, Cham, 2019, pp. 276–288. doi: 10.1007/978-3-030-37334-4_25.