WZB

Text-mining international politics

Session 3
Dictionary-based text analysis

1

WZB

Basic idea of
dictionary-based text classification

➤ Presence of or rate at which a _set of predefined key words_
   occurs in a document is used to classify or scale
   the document into/along theoretically relevant categories

+ Intuitive and easy to apply
+ Replicable and expandable to various theoretical concepts

– Variety of human language: false positives _and_ false negatives
– Appropriate normalization required …

2

1

WZB

Example I: Debating the EU in national parliaments (Rauh & De Wilde 2018, EJPR)

- **Question:** Do national parliamentary debates enhance the public accountability of EU decision-making?

  (If so, they should mirror EU authority, decision-making, public demand, and feature a balance of government and opposition...but party strategies...)

- **Text analysis approach**
  - Build a full-text corpus of parliamentary debates in various EU member states (get it here: www.bit.ly/ParlSpeech)
  - Find typical ways of referencing the EU polity, politics, and policies on n-gram level (= reading lots of speeches!)
  - Generalize these examples by *regular expressions* and build a respective *dictionary*
  - Count, normalize and aggregate EU references to party-month level
  - Relate this to relevant external data

WZB

Rauh & De Wilde (2018): Text data

	Period available	N speeches	Ø speeches per month	Ø terms per speech	Unique terms
DE: *Bundestag*	1991-03 / 2013-09	149,553	607.94	550.82	600,925
ES: *Congreso*	1989-11 / 2015-10	131,986	515.57	526.92	360,012
NL: *Tweede Kamer*	1994-12 / 2015-11	787,879	3396.03	165.57	401,471
UK: *House of Commons*	1988-11 / 2015-01	1,463,637	5361.31	202.07	1,037,450

Table 1: Domestic plenary debate corpora

Rauh & De Wilde (2018): Dictionary
(English version, other languages more complex)

EU polity	EU politics	EU policy
(eu|ec)[1-9]{1,2}	ecb	(c|e)sdp
(european|europe's|eu's) constitutional treaty	ecj	(common|european) foreign and security polic(y|ies)
(rome|maastricht|amsterdam|nice|lisbon) treat(y|ies)	ep	(common|european) security and defen(s|c)e polic(y|ies)
ec('s){0,1}	european (official(s){0,1}|civil servant(s){0,1})	eurozone|euro zone|euro area
economic and monetary union	european (politics|policy)	cfsp
eec('s){0,1}	european central bank	european ([a-z]*){0,1}polic(y|ies)
emu	european commission(er|ers){0,1}	european ([a-z]*){0,1}(act(s){0,1}|bill(s){0,1}|law(s){0,1}|legislation(s){0,1}|statute(s){0,1}
eu('s){0,1}	european competenc(e|es|ies)	european ([a-z]*){0,1}(aim(s){0,1}|goal(s){0,1}|target(s){0,1}
euratom('s){0,1}	european council	european ([a-z]*){0,1}decision(s){0,1}
european ([a-z]*){0,1}(integration|unification|cooperation)	european court of justice	european ([a-z]*){0,1}directive(s){0,1}
european_communit(y|ies)	european election(s){0,1}	european ([a-z]*){0,1}engagement(s){0,1}
european (economic |atomic energy)communit(y|ies)	european executive	european ([a-z]*){0,1}guideline(s){0,1}
european institutions	european level(s){0,1}	european ([a-z]*){0,1}(measure(s){0,1}|action(s){0,1})
european project(s){0,1}	european member state(s){0,1}	european ([a-z]*){0,1}(provision(s){0,1}|prescription(s){0,1}
european treat(y|ies)	european parliament	european ([a-z]*){0,1}(requirement(s){0,1}|allowance(s){0,1}
european_union('s){0,1}	european procedure(s){0,1}	european ([a-z]*){0,1}(standard(s){0,1}|norm(s){0,1}
single european act	european summit(s){0,1}	european ([a-z]*){0,1}agenda(s){0,1}
treat(y|ies) of (rome|maastricht|amsterdam|nice|lisbon)	mep(s){0,1}	european ([a-z]*){0,1}budget(s){0,1}
treaty establishing a constitution for europe	policy on europe	european ([a-z]*){0,1}f(u|o)nd(s){0,1}
treaty on (the functioning of the){0,1}european union		european ([a-z]*){0,1}programme(s){0,1}
		european ([a-z]*){0,1}regulation(s){0,1}
		european ([a-z]*){0,1}strateg(y|ies)
		european (case-law|jurisprudence|legal)
		european (single |internal)market{0,1}
		european [a-z]* union
		european currenc(y|ies)
		european mandate(s){0,1}
		police and judicial cooperation in criminal matters
		single currency
		stability and growth pact

5
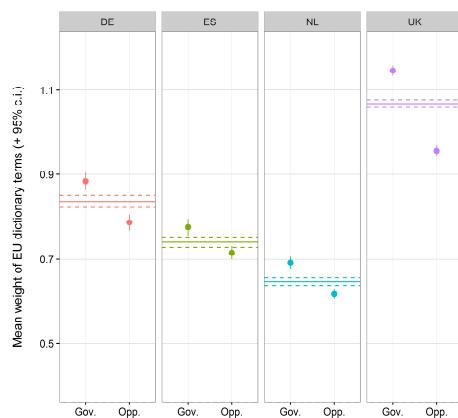
Rauh & De Wilde (2018): Dictionary
(English version, other languages more complex)

|	EU policy
	(c|e)sdp
	(common|european) foreign and security polic(y|ies)
	(common|european) security and defen(s|c)e polic(y|ies)
ervant(s){0,1})	eurozone|euro zone|euro area
	cfsp
	european ([a-z]*){0,1}polic(y|ies)
,1}	european ([a-z]*){0,1}(act(s){0,1}|bill(s){0,1}|law(s){0,1}|legislation(s){0,1}|statute(s){0,1}
	european ([a-z]*){0,1}(aim(s){0,1}|goal(s){0,1}|target(s){0,1}
	european ([a-z]*){0,1}decision(s){0,1}
	european ([a-z]*){0,1}directive(s){0,1}
	european ([a-z]*){0,1}engagement(s){0,1}
	european ([a-z]*){0,1}guideline(s){0,1}
	european ([a-z]*){0,1}(measure(s){0,1}|action(s){0,1})
	european ([a-z]*){0,1}(provision(s){0,1}|prescription(s){0,1}
	european ([a-z]*){0,1}(requirement(s){0,1}|allowance(s){0,1}
	european ([a-z]*){0,1}(standard(s){0,1}|norm(s){0,1}
	european ([a-z]*){0,1}agenda(s){0,1}
	european ([a-z]*){0,1}budget(s){0,1}
	european ([a-z]*){0,1}f(u|o)nd(s){0,1}
	european ([a-z]*){0,1}programme(s){0,1}
	european ([a-z]*){0,1}regulation(s){0,1}
	european ([a-z]*){0,1}strateg(y|ies)

6

3
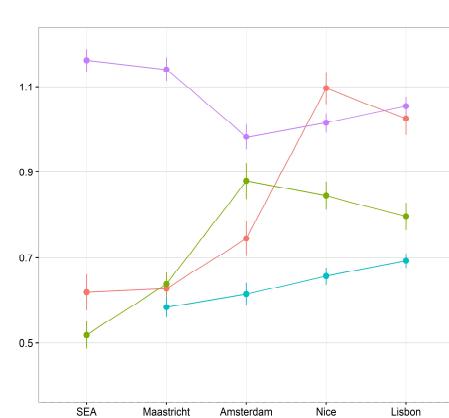
Rauh & De Wilde (2018): Descriptive results

Variation over countries and party types

Variation over treaty in force
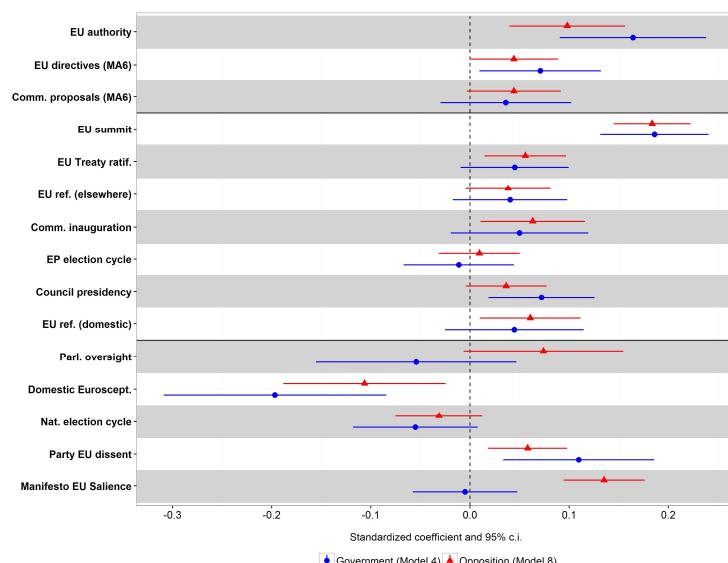
christian-rauh.eu

7

7

Rauh & De Wilde (2018): Multivariate results

EU salience in plenary speeches (ES, DE, NL, UK)

EU authority
EU directives (MA6)
Comm. proposals (MA6)
EU summit
EU Treaty ratif.
EU ref. (elsewhere)
Comm. inauguration
EP election cycle
Council presidency
EU ref. (domestic)
Parl. oversight
Domestic Euroscept.
Nat. election cycle
Party EU dissent
Manifesto EU Salience

Standardized coefficient and 95% c.i.

Government (Model 4)   Opposition (Model 8)

8

4

WZB

Example II: NGOs in the public discourse on the WTO (Rauh & Bödeker 2014 mimeo)
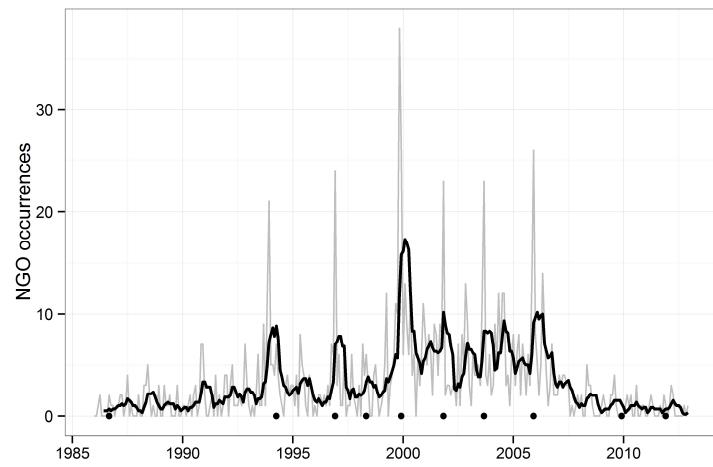
- **Question:** Which (type of) non-governmental actors participate in the public discourse on the World Trade Organization?

- Approach
    - o Financial Times (UK), Straits Times (S'pore), New York Times (US)
    - o Download, parse and clean all 11.388 articles from LexisNexis that mention the WTO in headline or lead (1985–2012)
    - o Generate encompassing list of all transnational NGOs (Sources: WTO stakeholder directory and UN ECOSOC database)
    - o Tag and count the occurrence of each of these NGOs in the articles
    - o 'Manually' classify tagged NGOs as 'business' or 'public interest'
    - o Aggregate and analyse the data

9

9

WZB

Example II (cont.)
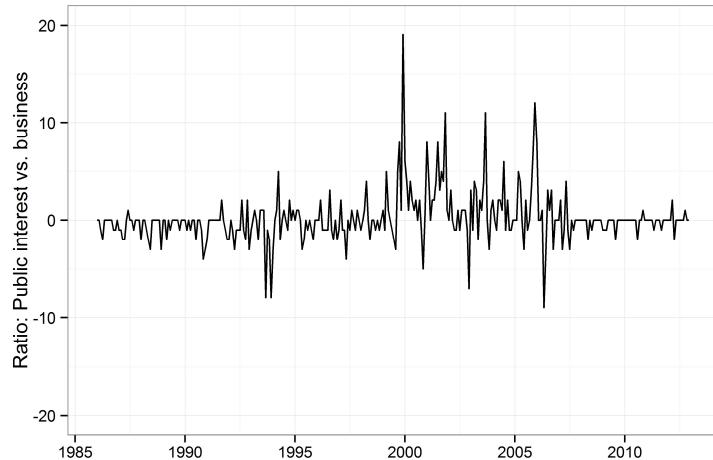

10

10

WZB

Example II (cont.)

11

**WZB**

Example II (cont.)
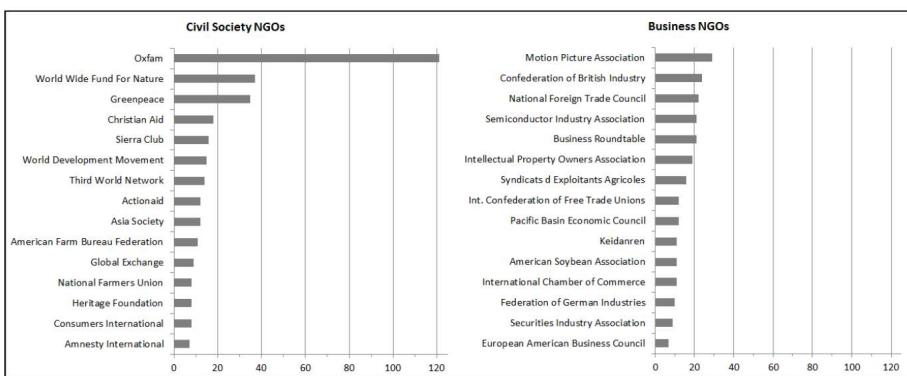
Figure 5: The 15 most present NGOs from the civil society and the business sector

12

12

WZB

Sentiment analysis

– Typical dictionary application

   o Do the analysed messages convey information positively or negatively (tone)?

   o Sentiment dictionary: List of terms with individual tone scores; usually ranging between –1 (negative) and 1 (positive)

   o Sentiment at document level: rate at which positively or negatively connoted words occur
(often: relative to the overall number of terms in document)

13

13

WZB

Exemplary sentiment dictionary (Young and Soroka 2011)

Exemplary terms (stemmed) from the
Lexicoder Sentiment dictionary

Positive connotation	Negative connotation
ALLEVIAT*	ABSURD*
BENEFIT*	BELLIGEREN*
COOPERAT*	CONFRONT*
DESERV*	CONTAGIOUS*
EXCITE*	FALTER*
FAIR*	HELPLESS*
OUTSTAND*	IDEOLOGUE*
PERFECT*	LOSE*
RESOLV*	NEGLECT*
USEFUL*	SCANDAL*
All in all, the LSD has 4,567 unique entries	

14

14

WZB

Sentiment analysis:
our mock example from yesterday

Fictitious speech
Expressed evaluations towards China (pos/neut/neg) ?

"China is a powerful neighbour.
Clearly, the Chinese political system hardly lives up
to our ideals about popular democracy.
But China has proven to be a reliable and
trustworthy trading partner.
We are concerned with its military strength, though
bilateral talks have improved significantly.
Nevertheless, China's decision as of yesterday is
absolutely unacceptable to us."
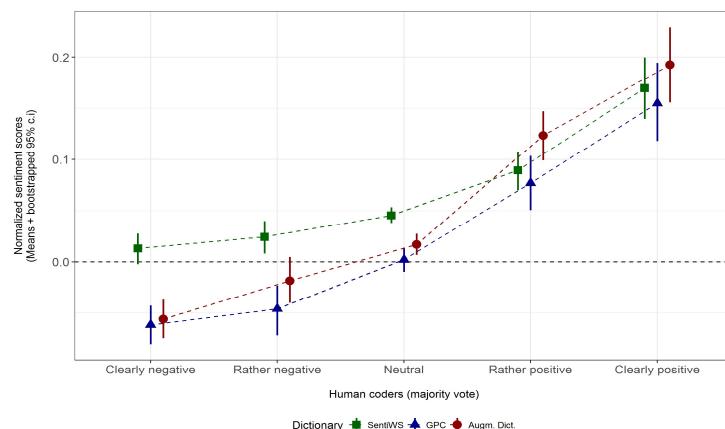
15

15

WZB

Sentiment dictionary validation
(example from Rauh 2018, JTIP)

Automated sentiment scores of 1,500 randomly sampled
sentences from speeches in the German Bundestag compared to
judgement by three human coders
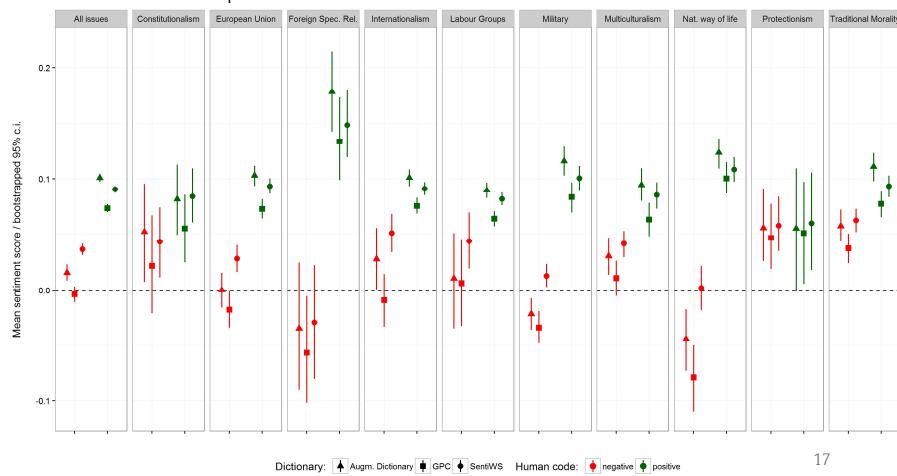
16

WZB

Sentiment dictionary validation
(example from Rauh 2018, JTIP)

Sentiment scores compared against human coding of directional categories in Manifesto project (14,008 quasi sentences from AT, CH, DE manifestos published 1998–2017)

17
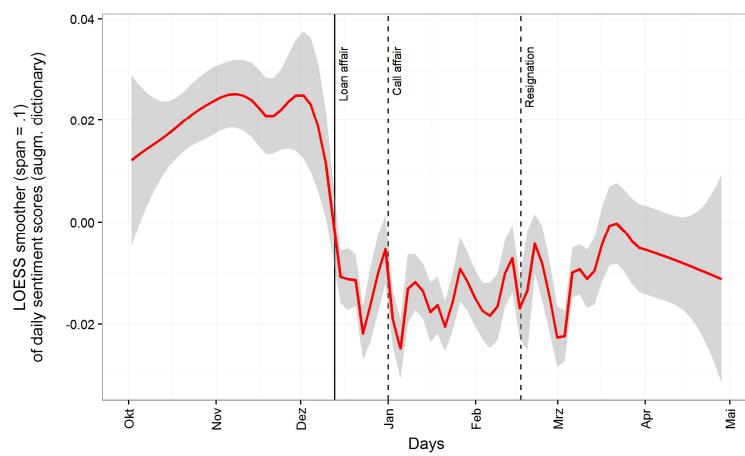
17

WZB

Sentiment dictionary validation
(example from Rauh 2018, JTIP)

Media sentiment around the scandal and resignation of German President Wulff in 2011/12 (4,038 articles with 'Wulff' in the headline from 256 German outlets covered in LexisNexis)

18

9

WZB

A sentiment analysis
# applied to our running example

- How positively or negatively do national delegates in the United Nations General assembly speak about climate change?
  ➢ And: Does this meaningfully capture expressed political *positions* on climate change issues?

- Approach
  ➢ Apply the Lexicoder sentiment dictionary with the respective functions in the *quanteda* R package to the corpus created above
  ➢ Normalized sentiment score in 100-term window around climate change references
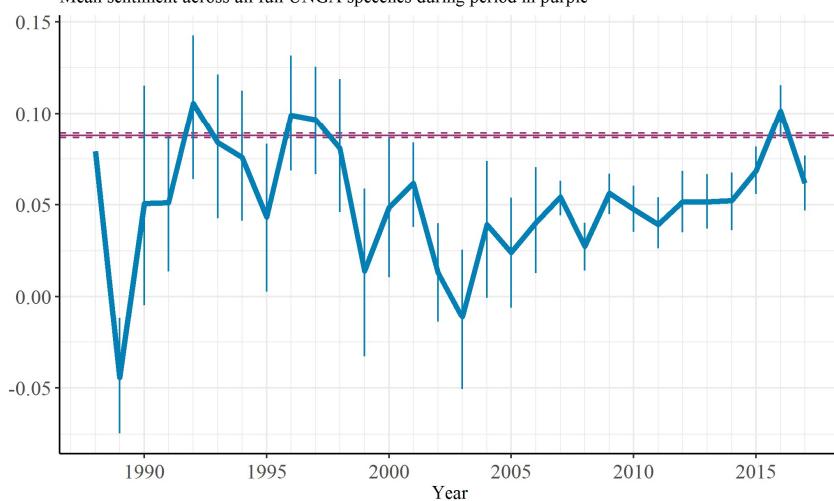
19
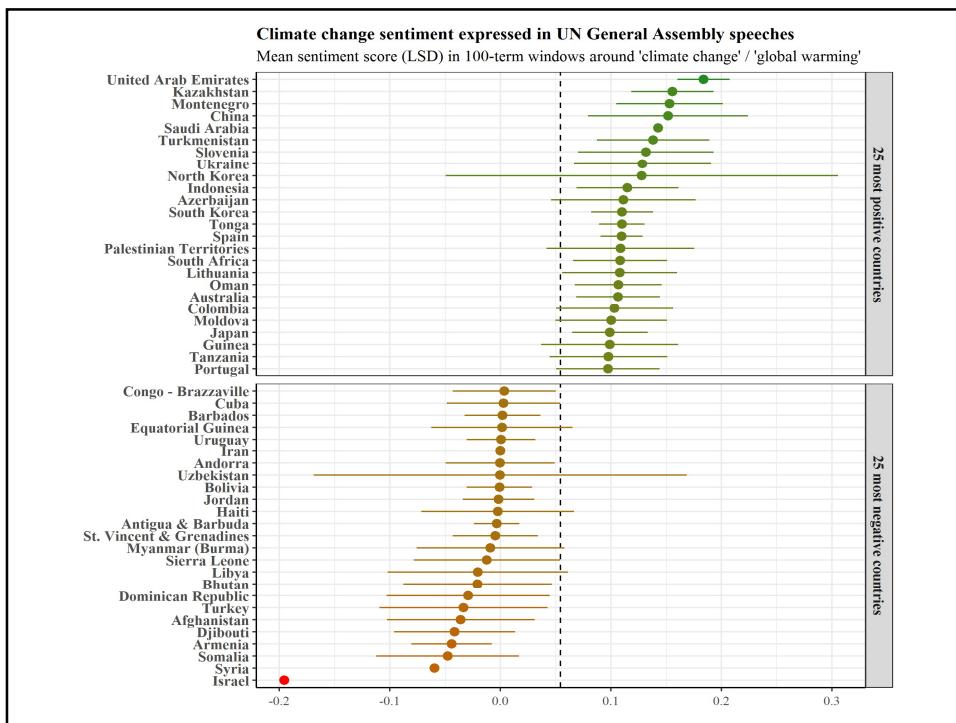
WZB

Climate change sentiment expressed in UN General Assembly speeches
Mean sentiment score (LSD) in 100-term windows around 'climate change' / 'global warming' references
Mean sentiment across all full UNGA speeches during period in purple

20

Climate change sentiment expressed in UN General Assembly speeches
Mean sentiment score (LSD) in 100-term windows around 'climate change' / 'global warming'

21

Text examples with rather extreme sentiment scores

Positive outlier (UAE)

"[....] meantime , we welcome the positive results of the West Point conference on combating terrorism , and look forward to the success of the next conference , to be held in the Republic of Korea. As part of our efforts to strengthen international cooperation in addressing the challenges of climate change and in order to assist States most at risk from the adverse effects of that phenomenon, we have initiated a partnership programme with the Pacific small island developing States , and we call upon the international community to support and expand the partnership at the global level . It is our hope that States parties to the Framework Convention on Climate Change will achieve concrete results before the forthcoming Conference in Mexico . We are pleased to see that the accord establishing the International Renewable Energy Agency has entered into force after ratification by the required number of States . In its capacity as host country , the United Arab Emirates continues [...]"

22

22

11

WZB

Text examples with rather extreme sentiment scores

Negative outlier (ISR)

"[....] is not just a problem for my country , it is a problem for the other countries too . If the United Nations spends so much time condemning the only liberal democracy in the Middle East , it has far less time to address war , disease , poverty , climate change and all of the other serious problems that plague the planet. Are the half-million slaughtered Syrians helped by Member States' condemnation of Israel ? The same Israel that has treated thousands of injured Syrians in our hospitals , including a field hospital that I built right along [...]"

➢ Context matters: the message clearly conveys negative sentiment but is not really about climate change (mitigation!?)...

➢ Be aware of such sources of interpretation/measurement error! Always inspect sample texts!
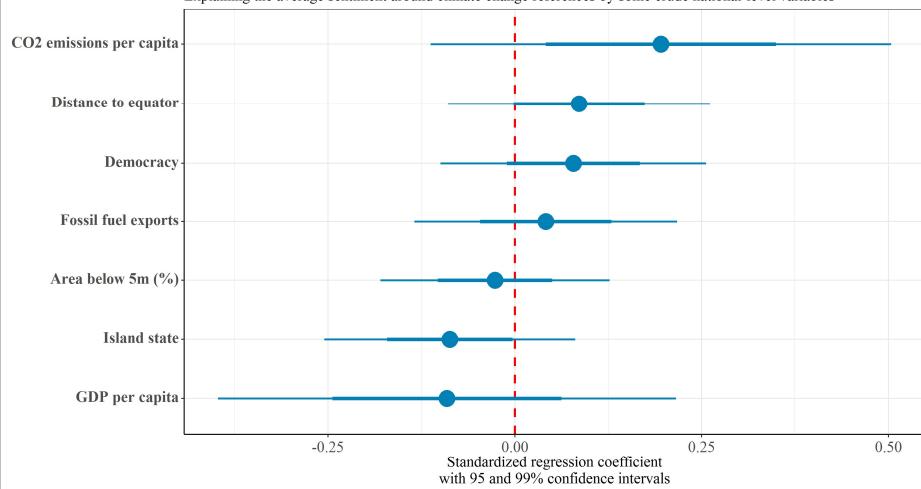
23

23

WZB

What sentiment do states attach to climate change?
Explaining the average sentiment around climate change references by some crude national-level variables

Standardized regression coefficient
with 95 and 99% confidence intervals

Linear Model, n = 191 countries, Adj. R2: .02.

24

24

12

WZB

Pitfalls of dictionary approaches

– Validity of the derived measures not granted
  - o Can the theoretical concepts/objects of interest be captured at term/document level?
  - o Do term level scores closely align with the typical word usage and with your object of inference in the analysed context?

⇒ Use 'off-the-shelf' term lists developed in other contexts only with extreme caution

⇒ Ideally: Develop your own dictionaries tailored to your research question, exploit regular expressions for flexibility

⇒ Apply/calculate context-specific baselines

⇒ In any case: Validate your results!
(e.g. against human coders or external data related to your concepts)

⇒ Validated sentiment dictionary for English political language: Young and Soroka (2011, *PC*)
⇒ Validated sentiment dictionary for German political language: Rauh (2018, *JITP*)
⇒ Validated auto-translated sentiment dictionaries for diff. languages: Proksch et al (2019, *LSQ*)