Friedrich-Alexander-Universität
Erlangen-Nürnberg

FAU

Discoursemes and mutlilingual embeddings as a technical basis for comparative discourse analysis

Stephanie Evert & Philipp Heinrich

Chair of Computational Corpus Linguistics
Friedrich-Alexander-Universität Erlangen-Nürnberg

July 2, 2025

1. Introduction

2. Case Study: ⟨climate change⟩

2.1 GermaParl

2.2 ParlSpeech-UK

2.3 Comparison

3. Conclusion

CADS

- Corpus-Assisted Discourse Studies (CADS)
  - research agenda: macro-level social critique
  - textual data + quantitative and qualitative CL methods
    - frequency comparisons (i.e. **keyword** and **collocation** analysis) $\rightarrow$ n-best lists $\rightarrow$ categorisation
    - discourse expertise
    - close reading of **concordance lines**

- overall research goal: **retroductable** CADS
  - allow subjective, but reproducible analyses
  - provide research documentation

Linguistic and discursive patterns

- **social critique** (macro level)
  - focus on injustices
  - sociological and political (left-wing) enterprise

- identification of **discursive patterns** in corpora (meso level)
  - discursive (nomination, predication, argumentation) strategies
  - framings, stances
  - genre-specific patterns

- description of (queryable) **linguistic patterns** (micro level)
  - word lists and n-grams
  - POS patterns
  - full-fledged corpus queries (e.g. CQP)

Discoursemes

- methodological harmonisation via **discoursemes** and discourseme constellations
  - discoursemes: minimal units of lexical meaning in the context of a specific discourse
- pairings of **hermeneutic interpretation** and **descriptions in terms of linguistic patterns**
  - ⟨climate change⟩, ⟨migration⟩
  - ⟨threat⟩, ⟨flood⟩
  - ⟨counterfactual implication⟩, ⟨ad hominem attack⟩
- discoursemes usually capture only one aspect of meaning
  - e.g. ⟨climate change⟩, ⟨nuclear energy⟩
  - also conflation of several aspects on the micro level possible (*Klimakrise*, *Atomkraftwerk*)
- discoursemes are usually **ambiguous**
  - ⟨ad-hominem⟩: *irrational*, *beschränkt*
  - ⟨threat⟩: *Flut*, *Welle*

Discourseme constellations

- discourseme constellations: **co-occurrence** of several discoursemes
  - ⟨climate change⟩ + ⟨nuclear energy⟩
  - ⟨migration⟩ + ⟨flood⟩
  - ⟨climate change⟩ + ⟨denial⟩ + ⟨anthropogenic⟩

- discourseme constellations: **approximations of discursive patterns**
  - linked to corpus material via queries
  - free discourseme-disambiguation

Semantic Maps

Capturing semantic relatedness

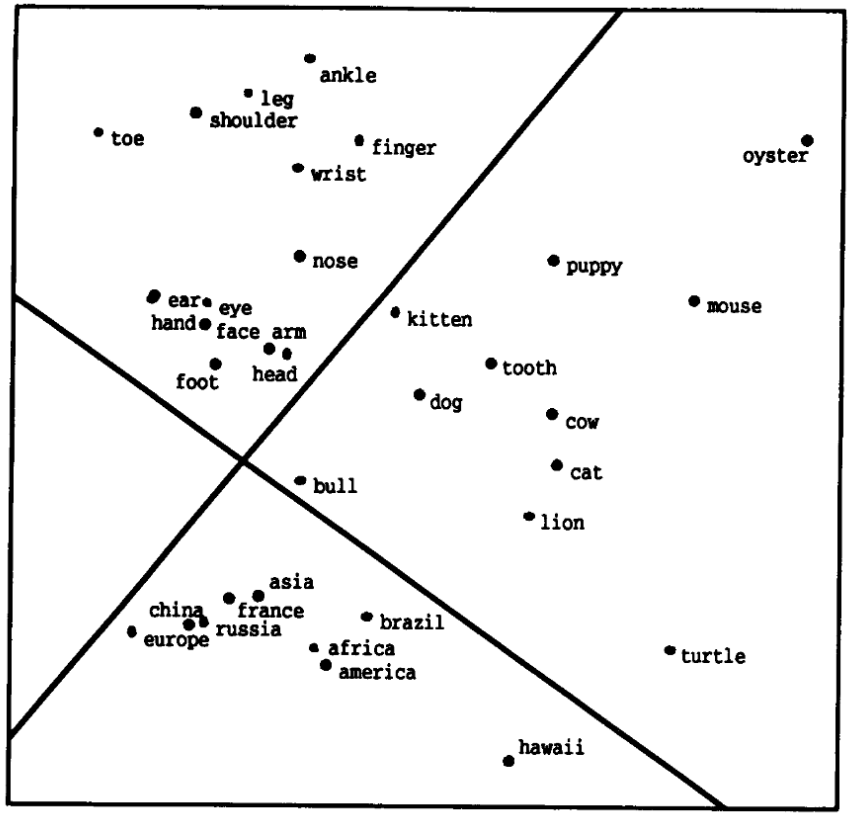


Figure: Lund & Burgess (1996)

Semantic Maps

Capturing semantic relatedness

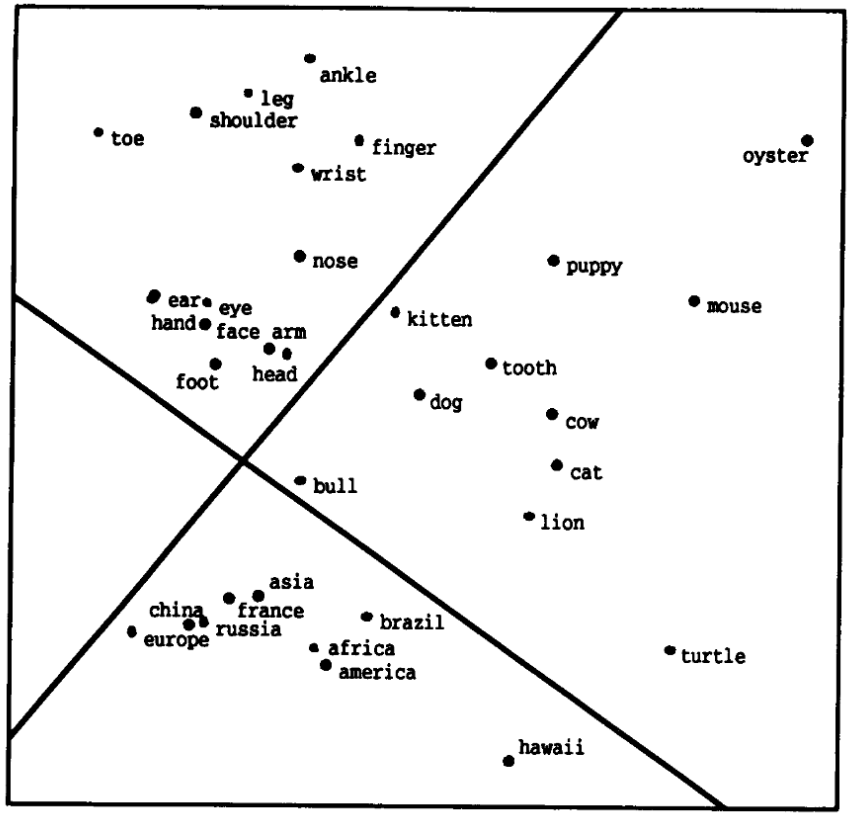


Figure: Lund & Burgess (1996)

- embeddings for capturing semantic relatedness
  - out-of-corpus creation
  - efficient storage (nearest neighbours)
- projection in two dimensions
  - t-SNE / UMAP / ...
  - update (projecting new types)
- combination with AM
  - size = importance / salience
- interactive categorisation

Multilingual CADS with discoursemes

- challenge
  - different framings and conceptualisations
  - different linguistic realisations of discursive patterns
  - lack of language proficiency
- opportunity
  - transfer knowledge across corpora
  - “modular” discourse analysis
- methodology
  - create discourseme descriptions for separate corpora
  - map discoursemes onto one another

Case Study

Climate change in German and UK parliaments

- discourse around **focus discourseme** ⟨climate change⟩ in GermaParl and ParlSpeech-UK
  - German: *Klimawandel, Klimakatastrophe, Klimakrise, Klimanotstand, Klimaveränderung, Klimaänderung*
  - English: *climate change, climatic change*
  - also *global[e] Erwärmung* / *global warming*

Case Study

Climate change in German and UK parliaments

- discourse around **focus discourseme** ⟨climate change⟩ in GermaParl and ParlSpeech-UK
  - German: *Klimawandel, Klimakatastrophe, Klimakrise, Klimanotstand, Klimaveränderung, Klimaänderung*
  - English: *climate change, climatic change*
  - also *global[e] Erwärmung* / *global warming*

- expectations
  - "controversial" topic: different parties with different stances
  - negotiation of basic understanding, most pressing questions
    - ⟨causes⟩ & ⟨consequences⟩
    - are human beings a decisive factor, i.e. is there significant ⟨man-made⟩ climate change?
    - ⟨denial⟩ or scepticism
    - do we have to expect negative ⟨consequences⟩ (for some or all human beings)?
    - can we influence the course of climate change, especially to ⟨adapt⟩ or mitigate (negative) ⟨consequences⟩?
    - how do we ⟨combat⟩ it?
    - if there are positive consequences, how do we settle the balance between individuals?
    - how do we reach an ⟨international agreement⟩

UK House of Commons

Country	LP	Election Date	First Meeting	Dissolution Date	Duration
⋮	⋮	⋮	⋮	⋮	⋮
UK	39	23 February 1950	1 March 1950	5 October 1951	1 year, 219 days
UK	40	25 October 1951	31 October 1951	6 May 1955	3 years, 188 days
UK	41	26 May 1955	7 June 1955	18 September 1959	4 years, 104 days
UK	42	8 October 1959	20 October 1959	25 September 1964	4 years, 342 days
UK	43	15 October 1964	27 October 1964	10 March 1966	1 year, 135 days
UK	44	31 March 1966	18 April 1966	29 May 1970	4 years, 42 days
UK	45	18 June 1970	29 June 1970	8 February 1974	3 years, 225 days
UK	46	28 February 1974	6 March 1974	20 September 1974	199 days
UK	47	10 October 1974	22 October 1974	7 April 1979	4 years, 168 days
UK	48	3 May 1979	9 May 1979	13 May 1983	4 years, 5 days
UK	49	9 June 1983	15 June 1983	18 May 1987	3 years, 338 days
UK	50	11 June 1987	17 June 1987	16 March 1992	4 years, 274 days
UK	51	9 April 1992	27 April 1992	8 April 1997	4 years, 347 days
UK	52	1 May 1997	7 May 1997	14 May 2001	4 years, 8 days
UK	53	7 June 2001	13 June 2001	11 April 2005	3 years, 303 days
UK	54	5 May 2005	11 May 2005	12 April 2010	4 years, 337 days
UK	55	6 May 2010	25 May 2010	30 March 2015	4 years, 310 days
UK	56	7 May 2015	27 May 2015	3 May 2017	1 year, 342 days
UK	57	8 June 2017	21 June 2017	6 November 2019	2 years, 139 days
UK	58	12 December 2019	17 December 2019	30 May 2024	4 years, 166 days
UK	59	4 July 2024	9 July 2024	ongoing	330 days +

Parties in UK House of Commons

LP57

Parliamentary Party	Seats
Conservative Party	344
Labour Party	205
Scottish National Party (SNP)	43
Liberal Democrats	15
Democratic Unionist Party (DUP)	8
Sinn Féin	7
Plaid Cymru	3
Green Party of England and Wales	1
Social Democratic and Labour Party	2
Alliance Party of Northern Ireland	1
Alba Party	2
Reform UK	1
Independent	14
Speaker	1
Total	*647*

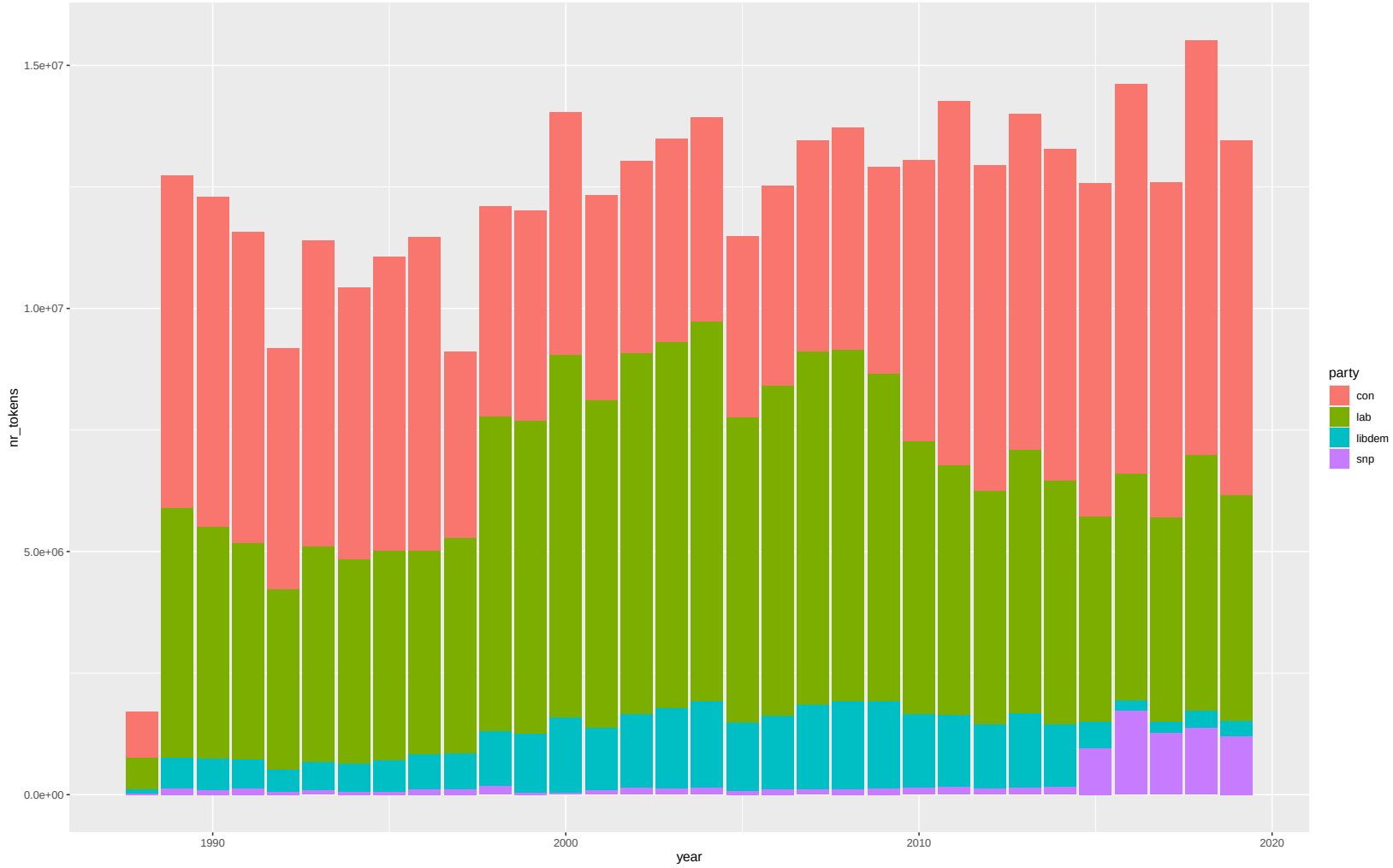
ParlSpeech-UK

- source: Rauh, Christian; Schwalbach, Jan, 2020, “The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies”, https://doi.org/10.7910/DVN/L4OAKN, Harvard Dataverse, V1

- coverage:
  - from 22 November 1988 (period 50)
  - until 19 December 2019 (period 57)

- amount:
  - 1,956,223 texts
  - 18,644,975 sentences – 409,134,520 tokens

ParlSpeech-UK: number of tokens over time



German Bundestag

Country	LP	Election Date	First Meeting	End of Term	Duration
DE	1	14 August 1949	7 September 1949	7 September 1953	4 years
DE	2	6 September 1953	6 October 1953	6 October 1957	4 years
DE	3	15 September 1957	15 October 1957	15 October 1961	4 years
DE	4	17 September 1961	17 October 1961	17 October 1965	4 years
DE	5	19 September 1965	19 October 1965	19 October 1969	4 years
DE	6	28 September 1969	20 October 1969	13 December 1972	3 years, 2 months
DE	7	19 November 1972	13 December 1972	13 December 1976	4 years
DE	8	3 October 1976	14 December 1976	4 November 1980	4 years
DE	9	5 October 1980	4 November 1980	29 March 1983	2 years, 5 months
DE	10	6 March 1983	29 March 1983	18 February 1987	4 years
DE	11	25 January 1987	18 February 1987	20 December 1990	4 years
DE	12	2 December 1990	20 December 1990	10 November 1994	4 years
DE	13	16 October 1994	10 November 1994	26 October 1998	4 years
DE	14	27 September 1998	26 October 1998	17 October 2002	4 years
DE	15	22 September 2002	17 October 2002	18 October 2005	3 years, 1 month
DE	16	18 September 2005	18 October 2005	27 October 2009	4 years
DE	17	27 September 2009	27 October 2009	22 October 2013	4 years
DE	18	22 September 2013	22 October 2013	24 October 2017	4 years
DE	19	24 September 2017	24 October 2017	26 October 2021	4 years
DE	20	26 September 2021	26 October 2021	27 December 2024	3 years, 2 months
DE	21	23 February 2025	25 March 2025	ongoing	135 days +

Parl. groups in German Bundestag

LP19

Parliamentary Group	Seats
Social Democratic Party (SPD)	206
Christian Union (CDU/CSU)	197
Alliance 90/The Greens (GRUENE)	118
Free Democratic Party (FDP)	92
Alternative for Germany (AfD)	78
The Left (LINKE)	28
Non-attached Members (fraktionslos)	6
Total	*725*

GermaParl (v2.2.0)

- source: Blaette, Andreas (2024): GermaParl. Linguistically Annotated and Indexed Corpus of Plenary Protocols of the German Bundestag. CWB corpus version 2.2.0-rc1. https://zenodo.org/records/12795193

- coverage:
  - from 7 September 1949 (LP 1)
  - until 28 June 2024 (LP 20)
  - here focus on LP11-LP18

- amount:
  - 4,524 protocols – 1,018,495 speeches
  - 16,948,210 sentences – 288,054,199 tokens

GermaParl: number of tokens over time

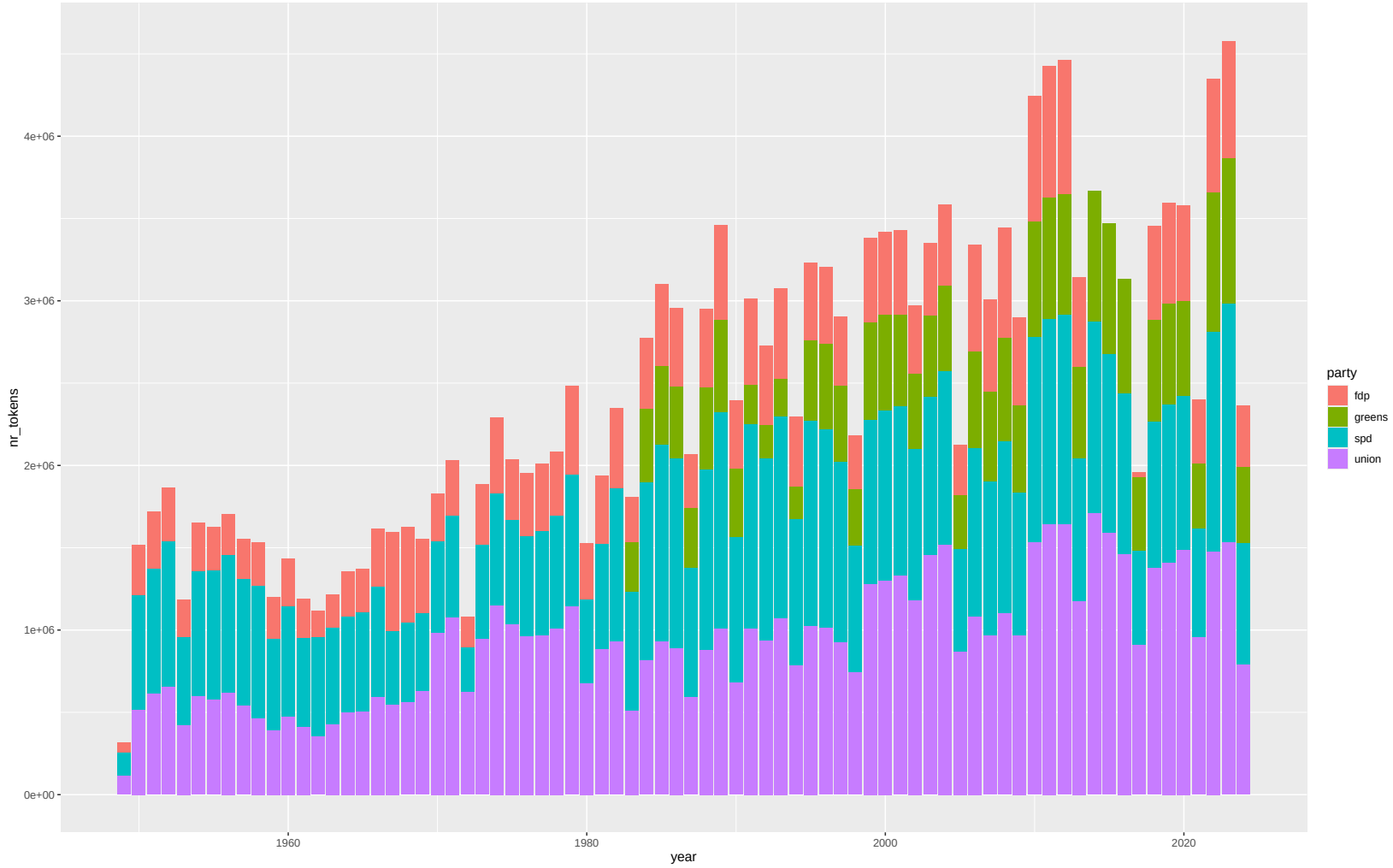
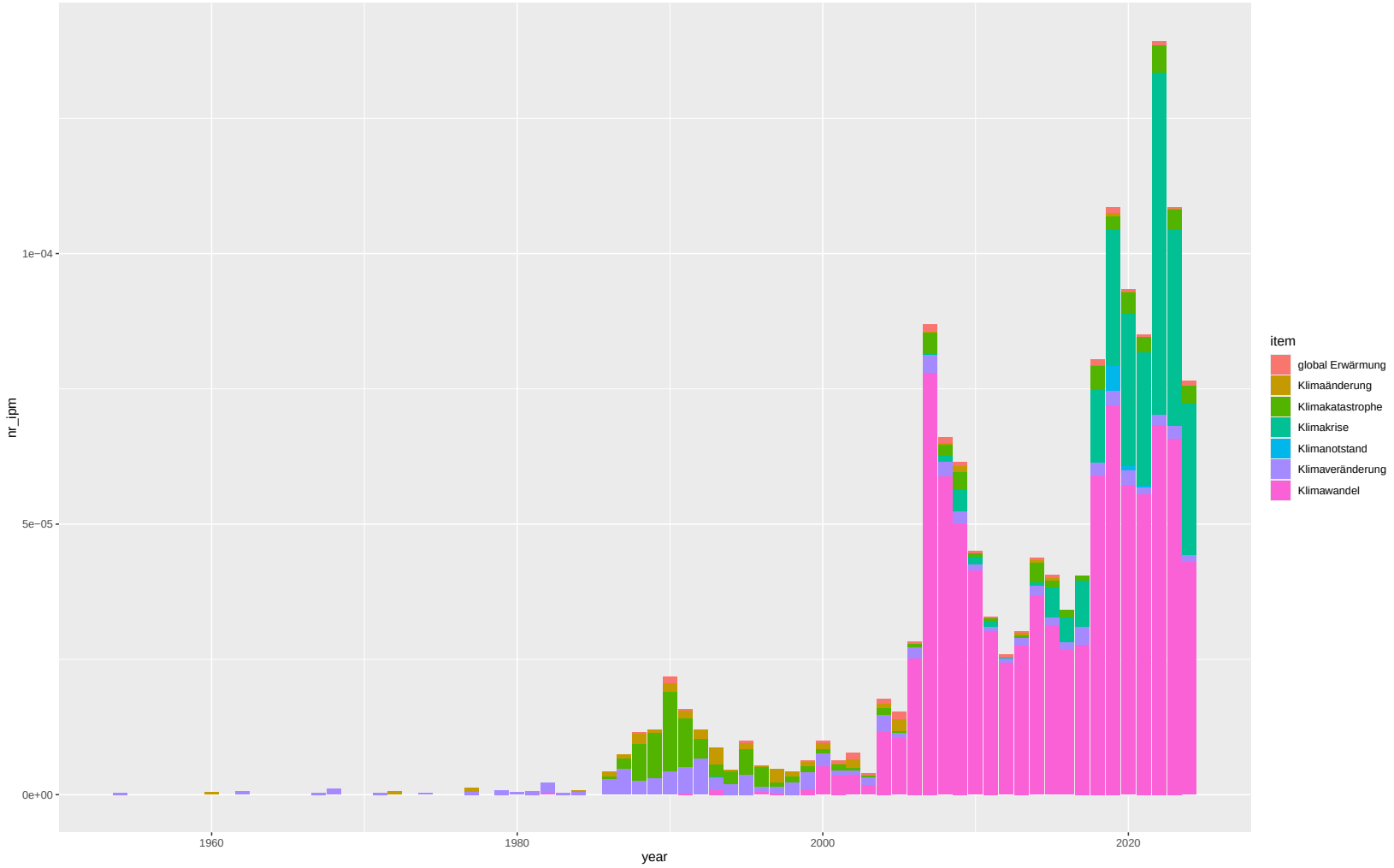


Table of Contents

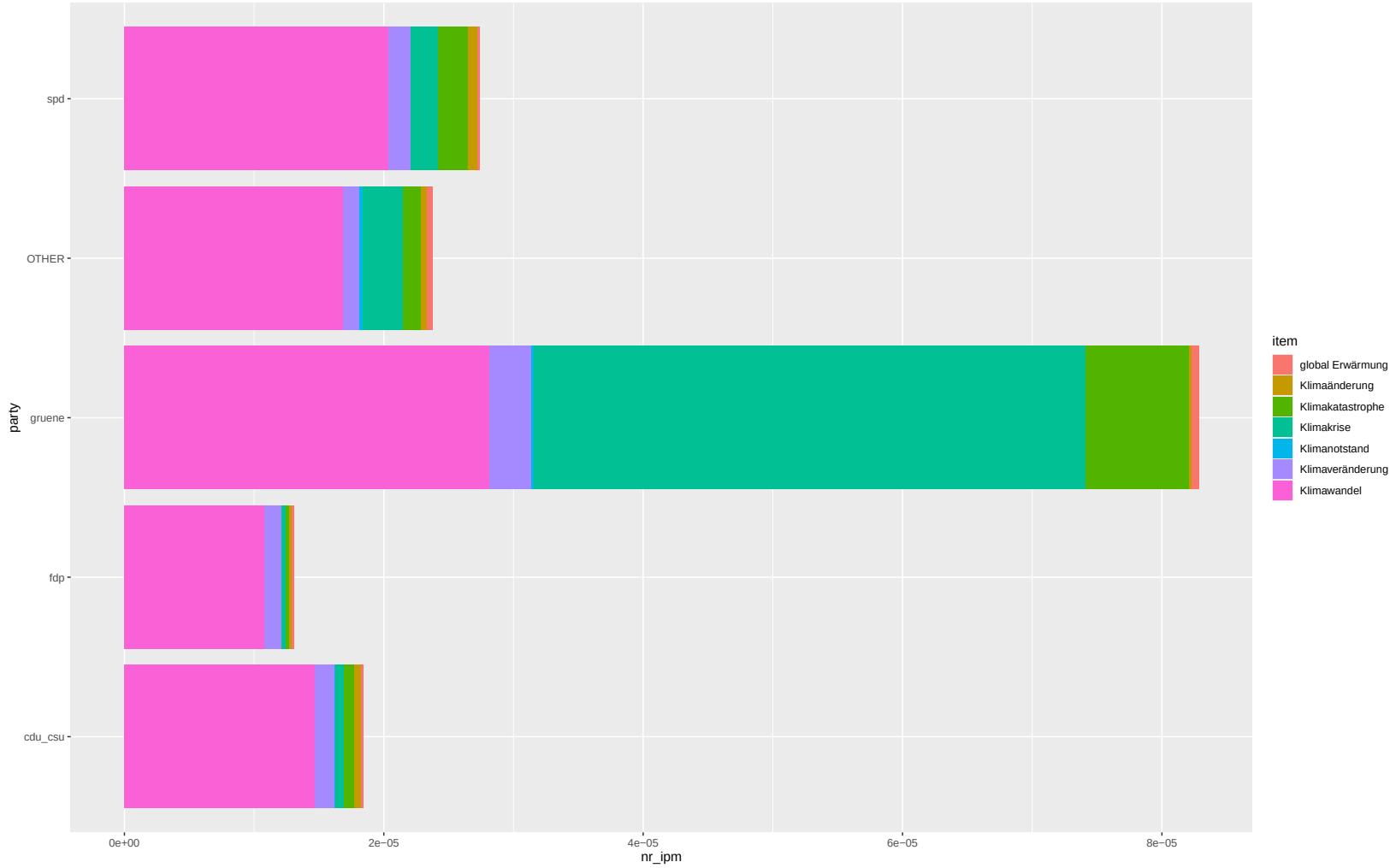
GermaParl: Focus discourseme breakdown

Across years



GermaParl: Focus discourseme breakdown

Across parties



GermaParl: collocation analysis

10L-10R-S-collocates of *Klimawandel* in GermaParl

rank	lemma	LRC
1	menschengemacht	11.74
2	Artensterben	9.20
3	Ressourcenknappheit	7.76
4	Anpassungsstrategie	7.74
5	Menschheitsherausforderung	6.85
6	Erderwärmung	6.37
7	Dürre	6.19
8	drohend	5.98
9	Wetterextrem	5.89
10	Hauptverursacher	5.66
11	Wassermangel	5.62
12	Artenvielfalt	5.48
13	Bevölkerungswachstum	5.37
14	Anpassungsmaßnahme	5.34
15	Biodiversität	5.27
16	Herausforderung	5.24
17	Naturkatastrophe	5.23
18	Ernährungskrise	5.22
19	Treibhauseffekt	5.16
20	leugnen	5.08

rank	lemma	LLR
1	menschengemacht	2406.84
2	Herausforderung	1848.52
3	die	1425.40
4	Folge	1400.30
5	global	1099.30
6	Auswirkung	854.99
7	Anpassung	850.94
8	Bekämpfung	826.61
9	Kampf	790.89
10	drohend	703.85
11	gegen	652.04
12	Artensterben	616.13
13	bekämpfen	545.56
14	leugnen	526.50
15	,	492.06
16	dass	482.90
17	Digitalisierung	461.30
18	aufhalten	402.87
19	Klimaschutz	389.67
20	und	376.72

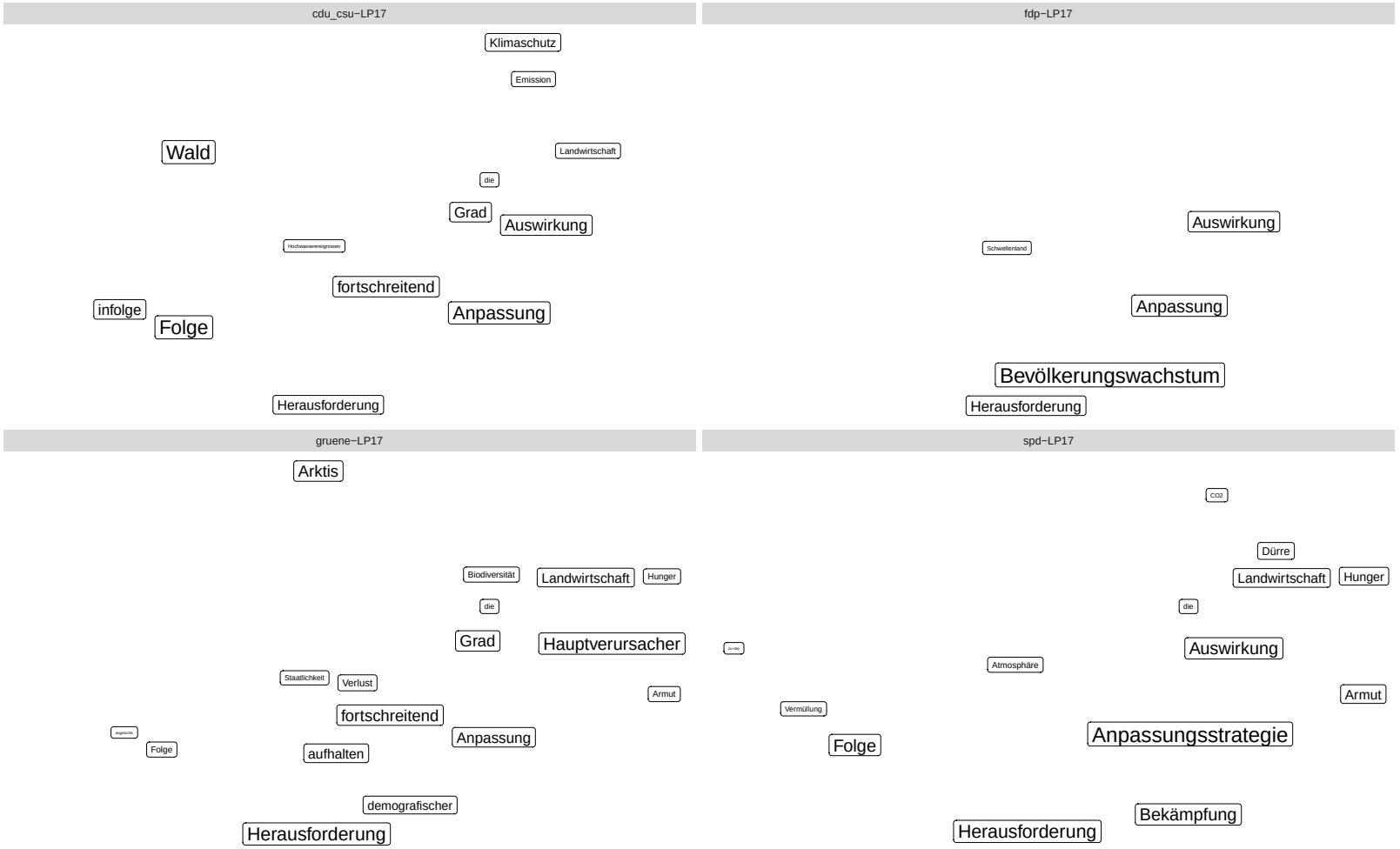
rank	lemma	MI
1	menschengemacht	3.38
2	Menschheitsherausforderung	3.08
3	Artensterben	3.04
4	Ressourcenknappheit	3.03
5	Extremwetterlage	3.00
6	Ernährungskrise	2.99
7	Biodiversitätsverlust	2.97
8	Anpassungsstrategie	2.89
9	Leugner	2.85
10	Wetterextrem	2.83
11	Nicholas	2.70
12	Wassermangel	2.68
13	Extremwetterereignis	2.65
14	Eisbär	2.64
15	Klimaflüchtling	2.62
16	Hungerkrise	2.58
17	Dürreperiode	2.56
18	Menschheitsfrage	2.56
19	Wüstenbildung	2.52
20	Erderwärmung	2.50

GermaParl: Collocation analysis

Semantic map

GermaParl: Collocation analysis

Parl. groups in LP 17



Interpretation of semantic map

- analysis layer: unigrams (single positional attribute or a tuple thereof)
- types that are close to one another occur in similar contexts; they're not necessarily similar
  - extreme case: synonymy (*carbon dioxide* and *CO2*)
  - also different but related concepts (*draught* + *biodiversity*)

- type n-grams (on the displayed analysis layer) are best visualised by their surface (on the analysis layer)
- complex linguistic patterns are best visualised by a name (e.g. query for argument pattern)
  - such queries can have various surface realisations on the analysis layer
  - conflation is not necessarily visible on the surface
  - for two identical surface realisations, one can be included in the discourseme and one not
- but cases of "looks like it, but is not" become increasingly infrequent with growing n
  - on unigrams: yes, WSD almost necessary (lemma_pos is a good start)
  - on bi- and trigrams rather infrequent (*cold shoulder* depends on the context, but besides pathology is always read idiomatically; *kick the bucket* is almost always used idiomatically)
  - *pull the plug on*, *at the end of the day*, *I don't want to talk about*

GermaParl: Collocation analysis

Discourseme description (1/2)

adaptation Anpassungsstrategie, Anpassungsmaßnahme, Anpassung, anpassen, Energiesicherheit, Potential

causes menschengemacht, Hauptverursacher, Treibhauseffekt, Bevölkerungswachstum, verursacht, menschgemacht, befeuern, Treibhausgas, menschenverursacht, verursachen, anthropogen, beitragen, hervorgerufen, Mitverursacher, Landwirtschaft, CO2, Verursacher, Raubbau, Ursache, Ausstoß, fossil, Klimagas, Umweltverschmutzung, Emission, Atmosphäre, Vermüllung

combat Rahmenübereinkommen, aufhalten, Kampf, Bekämpfung, eindämmen, entgegenwirken, entgegentreten, bekämpfen, abwenden, stoppen, bewältigen, verschärfen, wirksam, angehen, Bewältigung, verlangsamen, gegen, Abwendung, 12/4489, Urwaldschutz, Klimavorsorge, begrenzen, Marktversagen, Klimaschutz, Energiewende, Forschungsrahmenprogramm, Ressourcenschutz, Nachhaltigkeit, Abschwächung, Eindämmung

consequences Artensterben, Erderwärmung, Dürre, Wassermangel, Wetterextrem, Artenvielfalt, Naturkatastrophe, global, Biodiversität, Ernährungskrise, Überschwemmung, Wüstenbildung, Wasserknappheit, Meeresspiegel, Folge, Auswirkung, Arktis, leiden, Umweltzerstörung, infolge, Extremwetterlage, Hunger, Migration, Extremwetterereignis, Wetterveränderung, Wirtschaftskrise, Ökosystem, Wald, Vielfalt, Verlust, Hungerkrise, Lebensgrundlage, Verteilungskonflikt, Fluchtursache, Wetterereignis, Dürreperiode, Trockenheit, Erwärmung, Polkappe, Meer, Sturm, Eisbär, Bodenerosion, Grad, Borkenkäfer

denial leugnen, Leugner, Dogma, natürlich

GermaParl: Collocation analysis

Discourseme description (2/2)

economic Ressourcenknappheit, Ressourcenverknappung, Entwicklungsland, Ressourcenkonkurrenz, Afrika, Armut, Ressource, Bangladesch, Industriestaat, Süden, Investitionsschwäche, Schwellenland, Industrienation, Rohstoff, Knappheit

international Kyoto, IPCC, Kioto

related Ozonloch, Digitalisierung, Ozonschicht, demografisch, **demografisch[er] Wandel**, Coronakrise, Demografie, Finanzkrise, Energie, Globalisierung, Terrorismus, Ärmste, Pandemie, vektorübertragen

science Wissenschaftler, These, technologisch

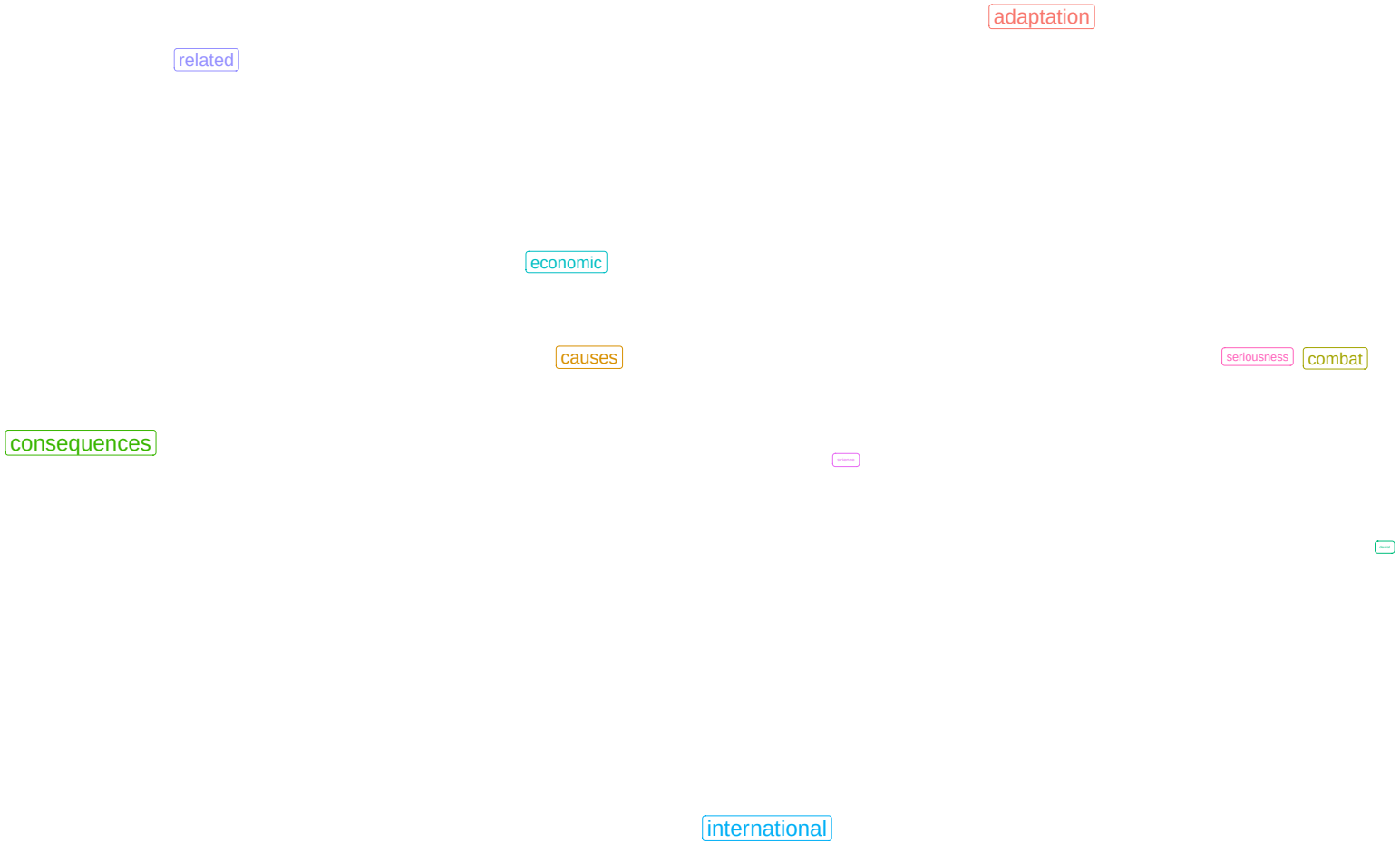
seriousness Menschheitsherausforderung, drohend, Herausforderung, fortschreitend, ungebremst, voranschreitend, schreiten, weltweit, bedrohen, stattfindend, voranschreiten, Fortschreiten, Problem, gravierend, Bedrohung, Menschheit, wachsend, dramatisch, entschlossen, Welt, groß, spürbar, existenziell, Krise, millionenfach, tiefgreifend, Risiko, Erde, Gefahr, Angst, abrupt

GermaParl: Collocation analysis

Categorisation

GermaParl: Collocation analysis

Conflation



GermaParl: Discourseme associations

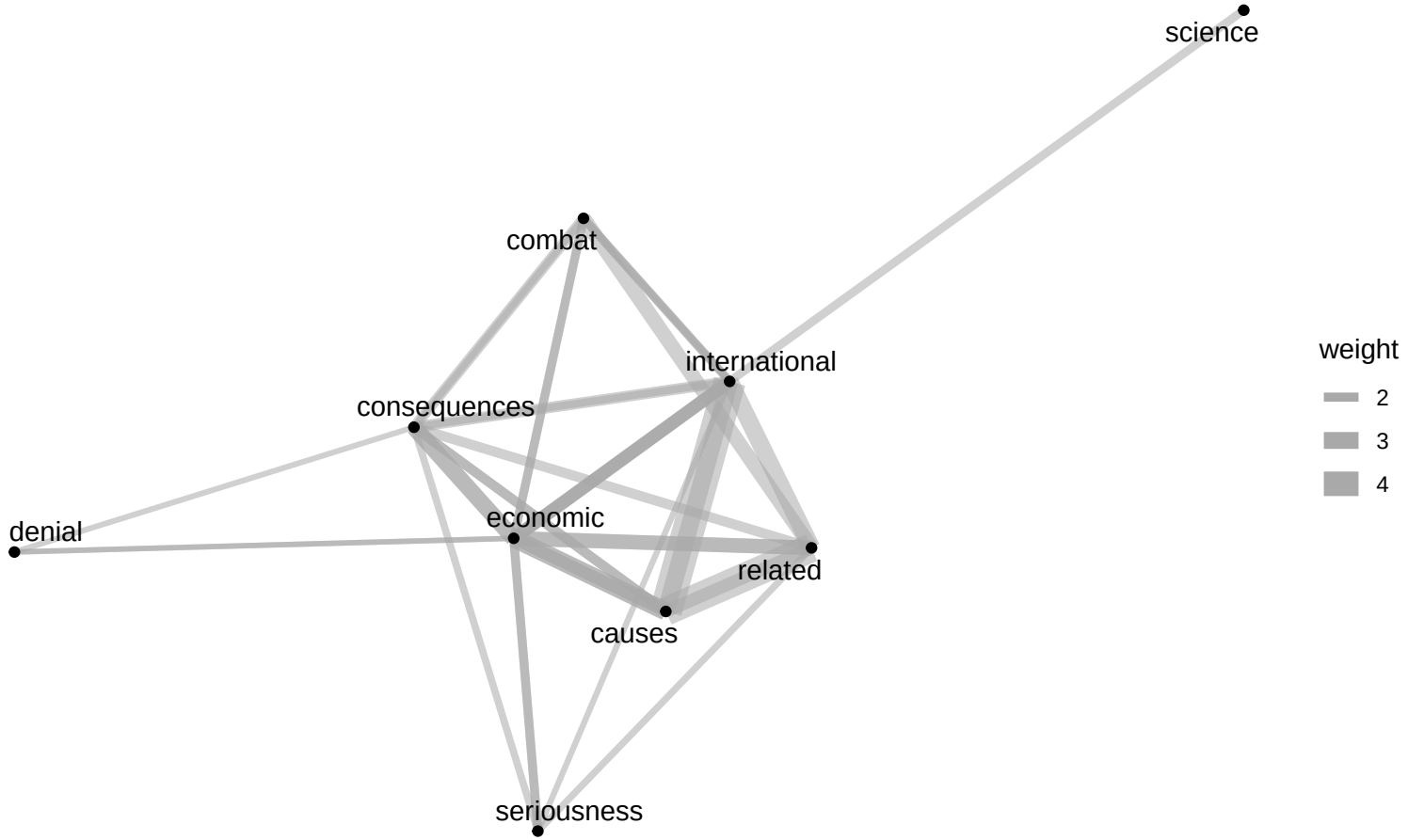
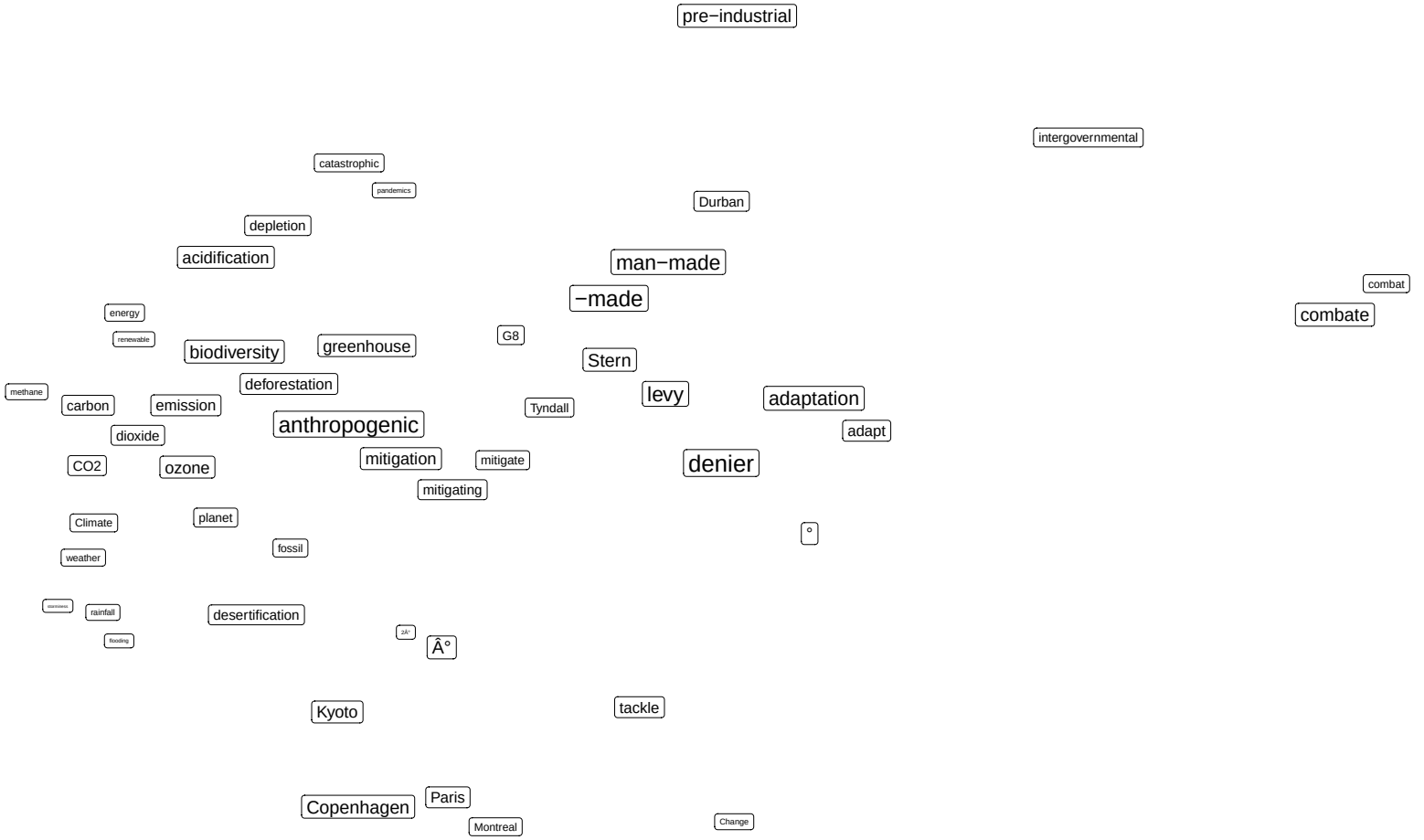


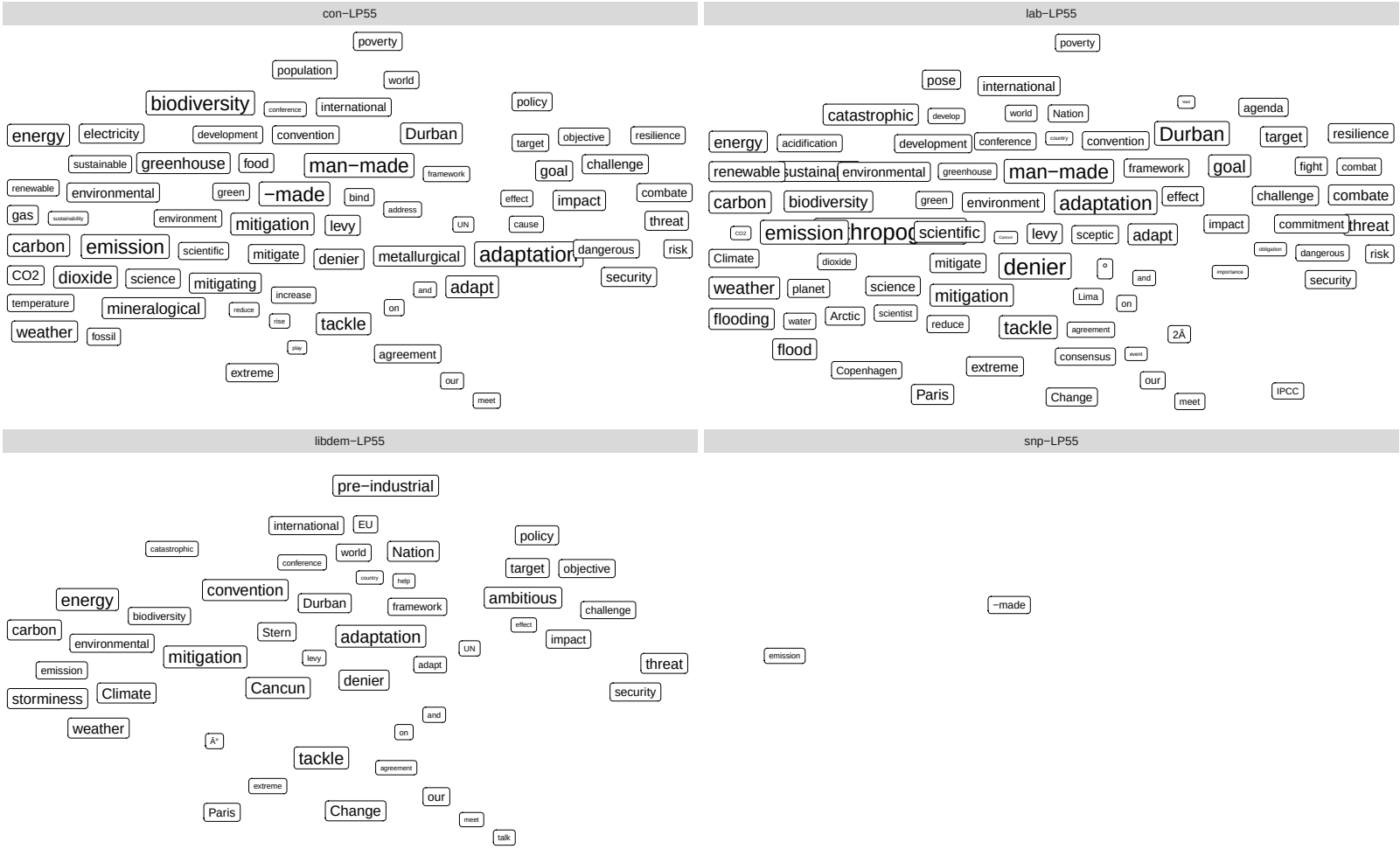
Table of Contents

ParlSpeech-UK: collocation analysis



ParlSpeech-UK: collocation analysis

Parl. groups in LP55



ParlSpeech-UK: collocation analysis

Discourseme description

adaptation adaptation, adapt, resilient, resilience

causes **man -made**, anthropogenic, man-made, greenhouse, **greenhouse gas**, **greenhouse gas emission**, emission, deforestation, CO2, dioxide, **carbon dioxide**, carbon, fossil, **fossil fuel**, **fossil fuel emission**, methane, aviation, **aviation emission**, hfc, pollution, gas, **rain forest**

combat combate, mitigation, tackle, mitigating, mitigate, **Climate Change Bill**, combat, **renewable energy**, 2Â°, CHP, avert, **environmental policy**, **environmental objective**, goal, ambitious, decarbonise, sustainability, agenda, 2050, target, aluminium, sustainable, exemption, chp, horticulture, fight, **climate change levy**

consequences biodiversity, acidification, desertification, **extreme weather**, **changes in weather**, **changing weather**, rainfall, flooding, storminess, degradation, ocean, **environmental impact**, **climatic effect**, **climatic result**, impact, drought, effect, atmospheric, acid rain, flood, sea, **rising sea level**

denial denier, sceptics

international Copenhagen, Kyoto, Paris, Durban, G8, intergovernmental, Montreal, **Intergovernmental Panel on Climate Change**, **Inter - governmental Panel on Climate Change**, **Climate Change Act**, **Climate Change Act 2008**, COP, Gleneagles, convention, **post -2012**, globally, conference, globalisation, Rio, summit, international, panel, Bangladesh, IPCC

related ozone, **ozone depletion**, **the depletion of the ozone layer**, depletion, pandemics

science Stern, **Stern report**, **Stern review**, pre-industrial, Â°, °, Tyndall, temperature, scientist, scientific, science, C

seriousness planet, catastrophic, threat, irreversible, habitat, challenge, runaway, pose, mankind, imperative, dangerous, existential

ParlSpeech-UK: collocation analysis

Categorisation

ParlSpeech-UK: collocation analysis

Parl. groups in LP55, categorised

ParlSpeech-UK: associations

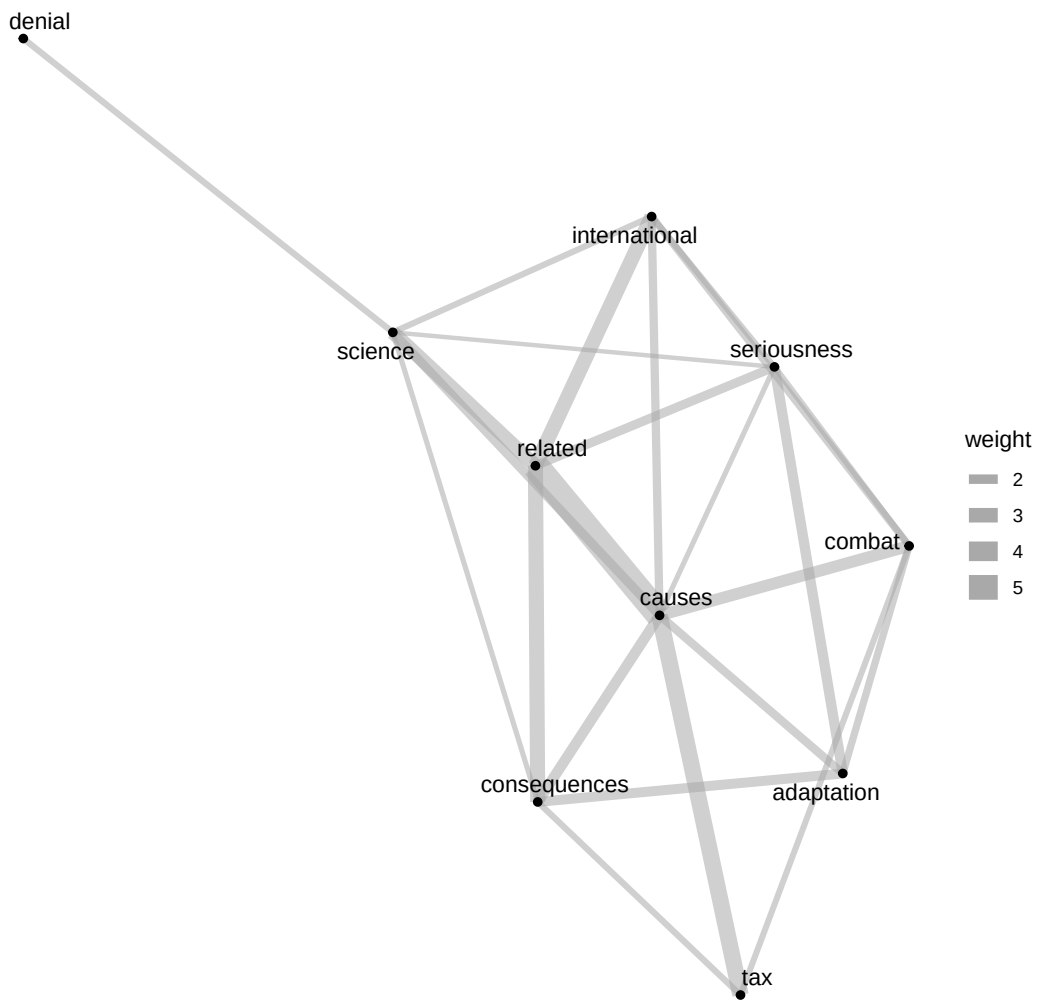


Table of Contents

Discoursemes across corpora

Sorted by association with ⟨climate change⟩

GermaParl

	discourseme	LRC
1	⟨consequences⟩	5.111
2	⟨international collaboration⟩	4.934
3	⟨adaptation⟩	4.611
4	⟨causes⟩	3.834
5	⟨related issues⟩	3.778
6	⟨combat⟩	3.660
7	⟨seriousness⟩	2.646
8	⟨science⟩	2.034
9	⟨denial⟩	2.021

ParlSpeech-UK

	discourseme	LRC
1	⟨denial⟩	6.409
2	⟨related issues⟩	5.612
3	⟨adaptation⟩	5.281
4	⟨causes⟩	4.540
5	⟨combat⟩	3.914
6	⟨science⟩	3.621
7	⟨international collaboration⟩	3.577
8	⟨seriousness⟩	3.452
9	⟨consequences⟩	3.225

Joint semantic map

Joint semantic map (LLR)

Joint semantic map (zoomed in)

Conclusion

- discoursemes as building blocks for CADS
  - set of linguistic patterns
  - clear hermeneutic defintion
- find descriptions across (sub-)corpora
  - concordancing still “single most important tool”
- MWUs:
  - English: define compounds, especially nouns (MWUs)
  - German: noun compounds as single tokens
- discourseme constellations
  - visualisation of association graph
  - automatic discourseme-sense disambiguation in constellations

Interactive discourse analysis

- CADS: highly interactive procedure
  - forming discoursemes
  - n-best lists $\rightarrow$ semantic maps
  - reading concordance lines

- software tools
  - Python library `cwb-ccc` (+ `association-measures`) `https://pypi.org/project/cwb-ccc/`
  - REST API `cwb-cads` `https://github.com/ausgerechnet/cwb-cads/`
  - R API (experimental)
  - react.js frontend `mmda-v2` (beta) `https://corpora.linguistik.uni-erlangen.de/mmda-v2/`

Thanks

⟨Thanks⟩ for your ⟨attention⟩!

Time for ⟨questions⟩.