Inducing Salient Information Structures from Climate Change Texts

Andrew Salway

Uni Research, Bergen

Why automated text analysis?

- Analysing textual materials is important for investigating the human side of climate change, e.g.:
  - what views do different actors express?
  - what voices are heard?
  - how are statements framed?
  - how does the discourse change over time?
  - how does the discourse vary between media?
- There are very large quantities of text available for analysis, e.g. newspaper articles, blog posts, scientific papers, policy reports, social media, etc.

It's all about counting...

- We count instances of linguistic forms that have meaning with respect to a conceptual framework and non-textual phenomena, e.g.:
  - Assume that the linguistic forms "climate change", "global warming" and "greenhouse gas" reflect the topic of climate change → count these in newspapers to analyse media coverage over time and between newspapers
  - Assume that the linguistic forms "a sustainable future", "a green future" and "a bleak future" reflect positive/negative views of the future → count these in accepting and sceptical blogs to compare their visions of the future

- **But, how do we know what to count?**

The danger of "seek and ye shall find"

- Top-down hypothesis-driven approaches assume that we know in advance what should be counted, i.e. we know what the interesting phenomena are, and we know how the relevant concepts map to countable linguistic forms

- Climate change discourse is complex – multiple facets and perspectives, multiple languages, changing over time → perhaps we shouldn't assume that we know all the relevant concepts and linguistic forms in advance

Data-driven content analysis

- Thus, we favour a data-driven approach, i.e. we use automated techniques to identify potentially "interesting" linguistic forms in our material, and then manually interpret these in order to develop a conceptual framework

  ➔ A more rigorous and reliable mapping between concepts and countable linguistic forms (?)

Investigating representations of the future in climate change blogs

- The NTAP corpus of blogs related to climate change:
  - about 3000 blogs, 1.4m blog posts, 400m words
  - focused on 330,000 sentences containing either "climate change" or "global warming"

Fløttum, K., Gjerstad, Ø., Gjesdal, A.M., Koteyko, N. and Salway, A. (2014). Representations of the future in English language blogs on climate change. *Global Environmental Change* 29, 213–222.

Salway, A., Fløttum, K. and Elgesem, D. (2015). Representations of the future in "accepting" and "sceptical" climate change blogs. To appear: *Procs. Corpus Linguistics 2015*, Lancaster University.

Example of a data-driven method (in brief)

- We identified frequent linguistic forms that could be related to representations of the future using:
  - Frequency lists
  - Word clusters
  - Sorted concordances
- Then, with some close reading, the frequent linguistic forms were interpreted to propose nine categories of meaning representations.

Example of a data-driven method (in brief)

- **Frequency lists:** we took the 1500 most frequent words and selected 11 that could be part of future representations, e.g. "future', "risks", "opportunities"; the 11 selected words had 30,000 instances in total

- **Word clusters and sorted concordances:** these gave a more condensed view of the co-texts around the identified words → manually analysed these to induce 42 patterns

Part of a sorted concordance

```
46  , men, farmers and pastoralists can have a bright future and never again suffer from famine hopefully th
47  industrialist oleg deripaska said he saw a bright future for nuclear development "because only nuclear c
48  energy that will allow mankind to have a brighter future, and this needs to happen now the only botherson
49  erved climate changes do not portend a calamitous future, global warming alarmism is invading nearly eve
50  ot day proof of global warming and a catastrophic future during their grandparents' early lives (and som
51  hat the world is on the brink of a "catastrophic" future of killer heatwaves, floods and droughts unless
52  ot day proof of global warming and a catastrophic future during times of natural global warming, elevate
53  nly seem to have a strong faith in a catastrophic future has global warming really stopped? has global w
54  e legislation an important step towards a cleaner future for australia but said much more needed to be d
55  e change, guilt, love of nature, wanting a decent future for your children  as an illustration of the li
56   climate change, we might fail to create a decent future – we're pretty close to the edge now  and there
57  limate change under control and preserve a decent future for our grandchildren unless we leave most of t
58  ostlethwaite as an old man living in a devastated future earth, watching archive film of the planet and
59  at it is how we focus collectively on a different future, and in focusing on it, make it happen "a power
60  te chapter of climate wars described a different  future scenario, exploring how climate change would af
61  ng itself to the inevitability of a discontinuous future, with our institutions and life support systems
62  ing, and if they're right, the state has a dismal future if nothing is done to stop it  a group of envir
63  ta and projecting perceived trends into a distant future that is difficult to grasp so much of the publi
64  ta and projecting perceived trends into a distant future that is difficult to grasp so much water is ext
65  d one to think so climate change is not a distant future climate change is not a forever problem climate
66  e " and far from being a threat only in a distant future, "climate change is happening now " and if ther
67  das of governments  and it is not just a distant "future" climate change that threatens us  and it is no
68  e change and climate model projections of a drier future across the south-east * iucn press release, dec
69   argue that it won't be a crisis in a foreseeable future either  neither howard nor rudd have committed
70  and when the main actor in that movie is a former future president, the rules of the game suddenly under
71   baird, wikimedia commons)  re-imagining a global future through dialogue and action tippingpointaustral
72  walk out of any presentation that showed a gloomy future; how people in her church would immediately dis
73  lm the age of stupid projects forward to a gloomy future climate change is a real phenomenon climate cha
```

Three of the induced patterns

Pattern	Unique fillers	Total instances	Number of instances for the five most frequent fillers
a|an WORD future	97	239	sustainable (34); low-carbon (19); better (15); uncertain (12); greener (7)
risk(s)|danger(s)| threat(s) facing WORD	30	142	the (43); our (26); humanity (25); mankind (10); humankind (5)
opportunity(ies) to WORD	325	843	make (39); address (18); put (16); build (16); take (15)

1) Political/environmental meanings

'a|an WORD future': sustainable (f=34); low-carbon (f=19); low-emissions (f=6); carbon-neutral (f=4)

```
any one technology is unlikely to solve the looming climate change and peak oil
problems, but working together within the larger system, they could tilt the globe
away from catastrophe and towards a sustainable future
```

'a|an WORD WORD future': clean energy (f=46); low carbon (f=10); more sustainable (f=10); safe climate (f=6); sustainable energy (f=6); cleaner energy (f=4)

```
recognizing that energy security, food security, climate change are interlinked, and
that eliminating poverty and ensuring sustainable development and a clean energy
future are among the foremost global objectives, the two leaders agreed to enter
into a green partnership to address these global challenges
```

'future of WORD': energy (f=11), coal (f=10), nuclear (f=7), renewable (f=6)

```
as concerns about climate change cast a shadow over the future of coal, a new energy
economy is emerging in the united states
```

'future of WORD WORD WORD' : the kyoto protocol (f=16)

```
the future of the kyoto protocol has become central to efforts to negotiate
reductions of carbon emissions under the un's framework convention on climate change
```

2) Value-laden positive meanings

'a|an WORD future': better (f=15); greener (f=7); green (f=5); livable (f=5); safer (f=5); strong (f=4); decent (f=3); bright (f=3)

'WORD opportunity(ies)': unique (f=48); good (f=12); golden (f=10); perfect (f=9); best (f=8); excellent (f=6); fantastic (f=4); better (f=4)

3) Value-laden negative meanings

'a|an WORD future': uncertain (f=12); bleak (f=6); catastrophic (f=3); apocalyptic (f=3)

'WORD opportunity(ies)': missed (f=20); lost (f=12)

4) Temporal meanings

'in the WORD (WORD) future' : near (f=152); foreseeable (f=31); distant' (f=30) ; 'far distant' (f=23); very near (f=11); immediate (f=9); not-too-distant (f=8); coming (f=4)

'a|an WORD future': distant (f=4); long-term (f=3)

5) Meanings related to the future of people, humanity and the planet

'future of WORD (WORD) (WORD)': the planet (f=41); humanity (f=30); our planet (f=18); the human species (f=10); our children (f=8); my generation (f=6); the world (f=5); humanity today (f=5); mankind (f=5); civilization (f=4); the earth (f=4); mankind on earth (f=4)

Also, some examples like "future for people and nature", "future for brazilian farmers", "dangers facing humans", "dangers to human health"

6) Meanings related to the future for regions and countries

Comparatively few examples, like "future for australia", "future of new zealand", "opportunity for the uk"

7) Meanings related to the future for nature and the environment'

'risk(s)|danger(s)|threat(s) to WORD WORD': polar bears (f=23); the environment f= (21); coral reefs (f=19)

'future of WORD WORD': coral reefs (f=10); himalayan glaciers (f=4); the oceans (f=3)

8) Meanings related to the future for business, industry and the economy

'WORD opportunity(ies)': economic (f=73); business (f=71); investment (f=25); development (f=11); employment (f=10); market (f=9); growth (f=6); commercial (f=4)

'opportunity(ies) for WORD': innovation (f=6); business (f=5); economic (f=5); investment (f=5); growth (f=4); companies (f=4)

'risk(s)|danger(s)|threat(s) to WORD (WORD)': the economy (f=9); their business (f=9); businesses (f=3)

9) Meanings related to the future for security

'risk(s)|danger(s)|threat(s) to WORD WORD': national security (f=31); international peace (f=24); our security (f=12); global security (f=12)

Discussion

- The mapping that we established between the induced linguistic patterns and meaning categories facilitates quantitative analyses, and the identification of samples for close reading.
  - For example, quantitative analysis to test the hypotheses that accepting climate change blogs are more concerned with the future than sceptical blogs.
- Automated techniques made it possible to identify patterning in a large volume of material but the method still required quite a lot of manual effort, and it was somewhat ad hoc

Local grammar induction

- We are developing an automatic technique to induce salient information structures from large text corpora without manual intervention (the researcher still interprets the structures!)

- Adapted and applied a grammar induction algorithm from computational linguistics – ADIOS, Automatic Distillation of Structure (Solan et al. 2005)

Salway, A. and Touileb, S. (2014). Applying Grammar Induction to Text Mining. Procs. ACL 2014.

Salway, A., Touileb, S. and Tvinnereim, E.(2014). Inducing Information Structures for Data-driven Text Analysis. Procs. ACL workshop on Language Technologies and Computational Social Science.
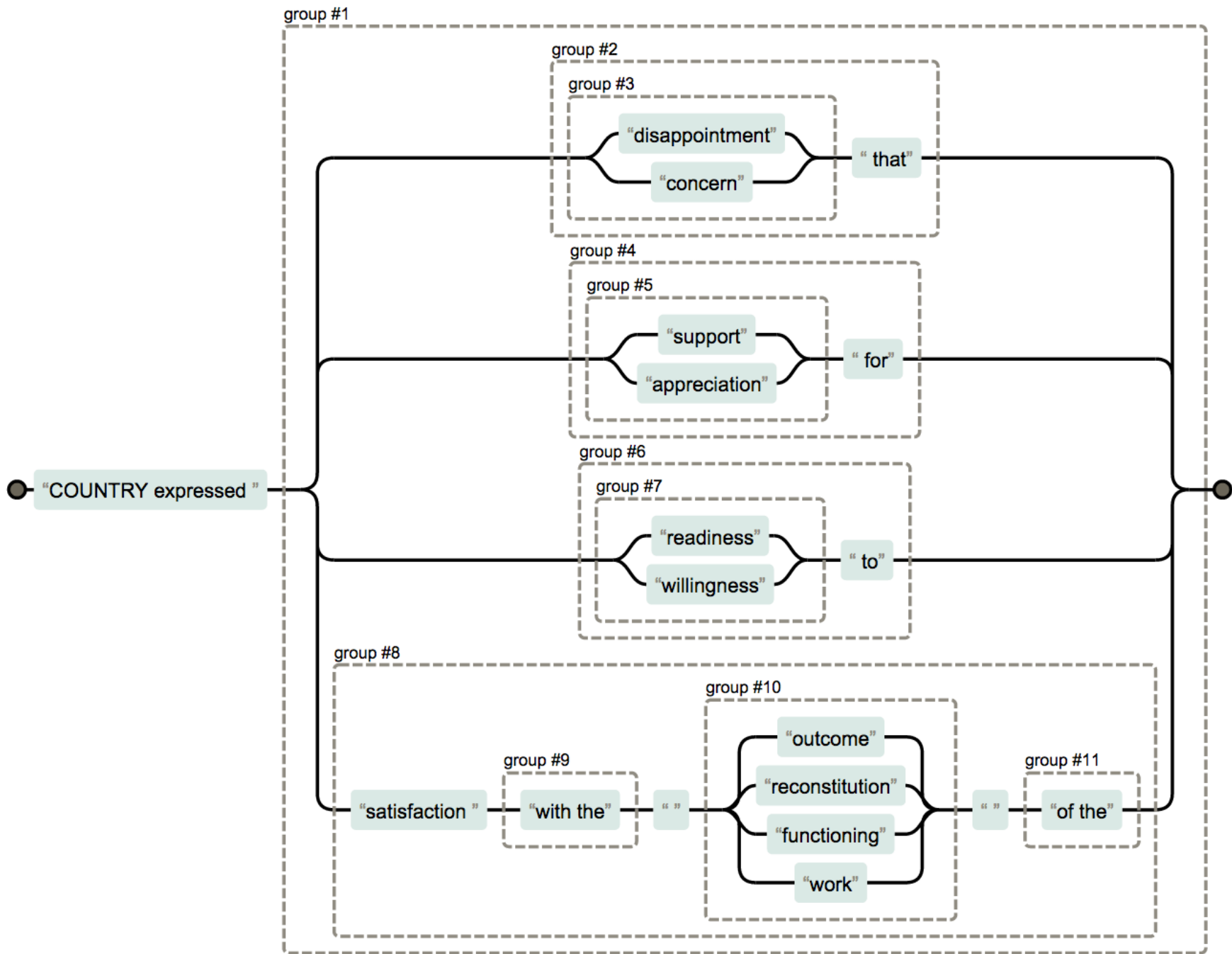
Example of induced structures
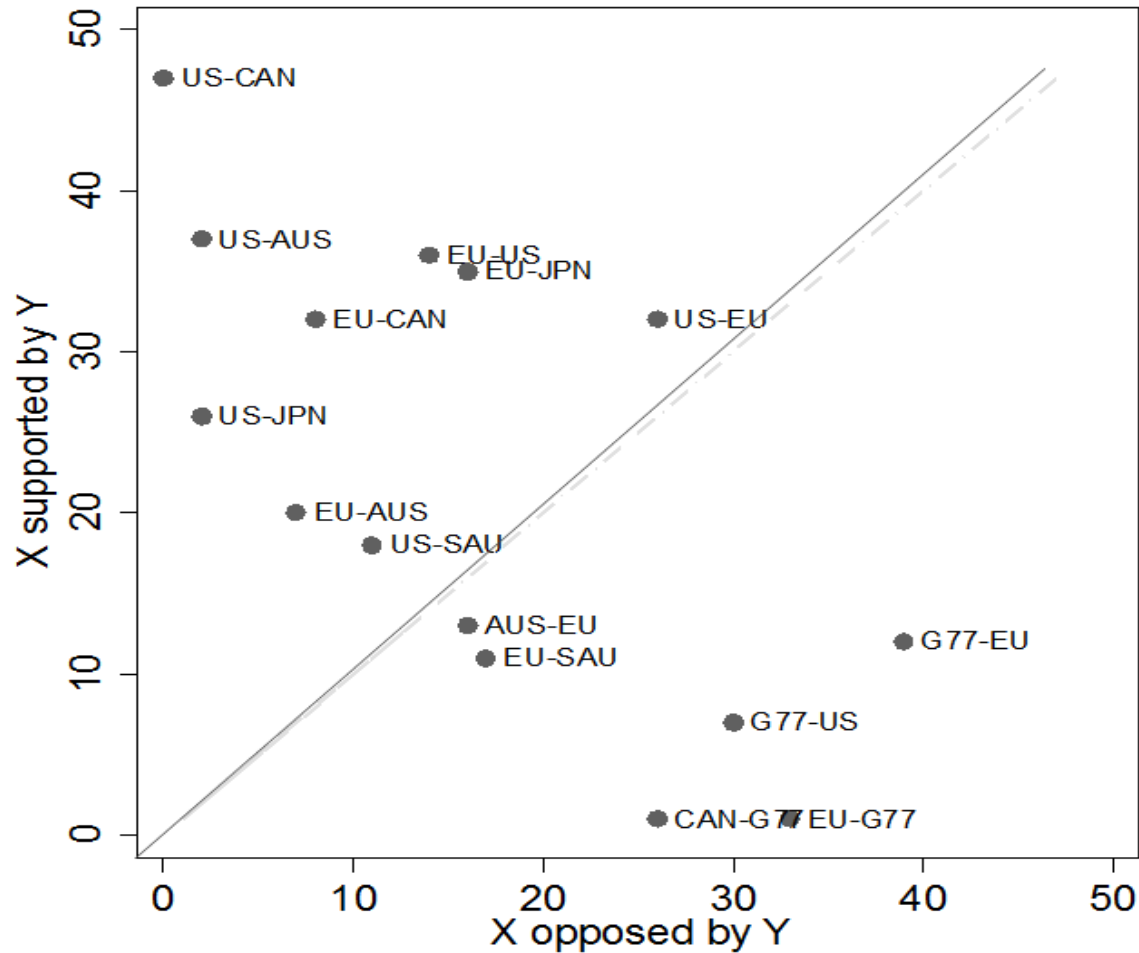
(COUNTRY ((supported|opposed) by) COUNTRY)

(COUNTRY (said|noted|recommended|explained|responded|stressed|questioned|addressed|reiterated|reported|urged|amended|invited…))

(COUNTRY ((clarified|urged|reported) that)

(COUNTRY ((presented|demanded|outlined|favored (the|a))

"COUNTRY expressed"

group #1

group #2

group #3
- "disappointment"
- "concern"

" that"

group #4

group #5
- "support"
- "appreciation"

" for"

group #6

group #7
- "readiness"
- "willingness"

" to"

group #8

"satisfaction"

group #9
"with the"

" "

group #10
- "outcome"
- "reconstitution"
- "functioning"
- "work"

" "

group #11
"of the"

Dyads of support and opposition

Scale of climate change statements

Austria (-2.38), Belgium, Germany, the UK, Switzerland, the US, Canada, Australia, Norway, France, Russia, New Zealand, Japan (-.62) ,Papua New Guinea (-.26), Tuvalu, Peru, Mexico, Brazil, Argentina, Malaysia, South Korea, Colombia, Saudi Arabia, Chile, Kuwait, Nigeria, Grenada, Uganda, Bangladesh, China, Egypt, the Philippines, South Africa, Indonesia, Venezuela, Iran, Bolivia, Barbados, India, Algeria (1.44)

Closing remarks

- A data-driven approach alleviates the problem of "seek and ye shall find"
- Automated techniques can elucidate interesting linguistic patterning
- They do depend on sufficient patterning in the corpus – better for large corpora, more stylised language?