Dialects of Discord

Analyzing the Debate on Nuclear weapons in the Netherlands 1970-1990

Milan van Lange &
Ralf Futselaar

Overview:

1. We asked the wrong historical question
2. We used the wrong text analysis approach
3. We made it all work nevertheless! (well, sort of)

---

1. The Wrong Question

---

1. The Wrong Question

- Around 1980 strong rise in opposition to nuclear proliferation in Europe
- Political impact seemingly small, activists mostly defeated
- Extensive discussions in parliament

Cruise missiles in the Netherlands

- The Euromissile crisis of 1977-1982
- 1979:NATO Double track decision
- Pershing cruise missiles to be positioned in NL, GB, DE, IT, BE
- Rise of "Hollanditis" (Walter Lacquer)

The failure of Hollanditis

- Dutch government agrees to placement in 1986
- NATO no longer interested, missiles never actually come
- (Sate Secret Disclosure: US nuclear weapons have been stationed on Volkel Air Base since 1961)

The failure of Hollanditis

- Stable political majority in favour of nuclear weapons, voting likewise
- Politicians walk the walk. But do they talk the talk?

The Political Arena

- Several parties oppose NATO and missiles (PSP, CPN)
- Several parties favour NATO and missiles (VVD, SGP)
- Several parties favour NATO, but dislike missiles (PvdA, CDA)

The Political Arena

- Very heated debates in parliament
- Rebellion within the Christian Democrats (CDA): pro
- Social Democrats (PvdA) electorally threatened: contra

1. The Wrong Question

- No change in observed political behaviour (party programmes, voting behaviour)
- Will these viewpoints be reflected in preferred vocabularies?

2. The Wrong Analysis

- Using a Word Embedding Model (Word2vec) in our case

  - Powerful
  - Broadly adopted and tested
  - Not very suitable for historical research

The corpus

- Proceedings of the (bicameral) Dutch parliament (the Estates General)
- Lemmatised by Martin Reynaert (Tilburg Uni)
- Enriched by Maarten Marx, Uni of Amsterdam
- Check out: www.dans.nl

Keep in mind:

- We never leave the universe of the spoken and speakable
- We have no access to nonverbal communication, or tone
- We will not catch any "elephants in the room"

13

Word Embedding Model (WEM)

- WEM = spatial representation of a corpus
- Position of a word in this *discursive space* mapped by vector
- Vector = sequence of numbers
- Close distance in vector-space = similarities in usage and/or context between words = nearest neighbour (NN)

14

Nearest Neighbours (NNs)

- Synonyms very close…

```
> religie <- c("religie")
> vec6575 %>% nearest_to(vec6575[[religie]], 15) %>% round(3)
      religie        godsdienst wereldbeschouwing       christendom
        0.000             0.230             0.245             0.263
```

15

Nearest Neighbours (NNs)

- … but antonyms too!

```
> schoon <- c("schoon")
> vec6575 %>% nearest_to(vec6575[[schoon]], 15) %>% round(3)
      schoon        schone          vuil       schoner
        0.000         0.232         0.233         0.264

> wit <- c("wit")
> vec6575 %>% nearest_to(vec6575[[wit]],
          wit         zwart          geel
        0.000         0.238         0.369
```
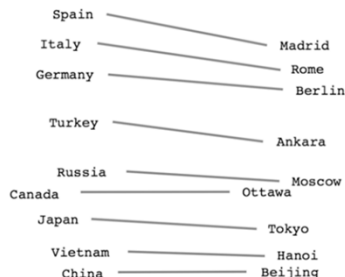
Nearest Neighbours (NNs)

- NNs tell us something about closeness: there is a close semantic relationship
- Nature of this relationship needs interpretation

```
> moeder <- c("moeder")
> vec6575 %>% nearest_to(vec6575[[moeder]], 15) %>% round(3)
       moeder         vader          kind     grootouders      kinderen
        0.000         0.105         0.184         0.214         0.217
```

17

Spain ——— Madrid
Italy ——— Rome
Germany ——— Berlin
Turkey ——— Ankara
Russia ——— Moscow
Canada ——— Ottawa
Japan ——— Tokyo
Vietnam ——— Hanoi
China ——— Beijing

Country-Capital

Using WEMs in practice: first round

- 1. Train a WEM on complete corpus 1970-1990

- 2a. Using NNs to find all the words that represent the concept of 'nuclear weapon'

- 2b. Find the words used to express each of the two viewpoints in this bipolar debate

19

Step 1: Train model

- Train WEM on our complete dataset of parliamentary proceedings 1970-1990
- Using Google Word2Vec
  - Package in R
  - Vectors = 100
  - Min_count = 5

Step 2a: Nuclear weapon vector

- Looked at the 200 NNs of '<u>nuclear weapon</u>' in WEM based on complete corpus
- Selected the synonyms of 'nuclear weapon' (= 'nuclear weapons' + 'atomic bomb' etc.)
- Create 'combined vector' for these words

```
# Create combined vector with meaning 'nuclear weapon'
comb_nuc_words <- c("nucleair", "nucleaire", "atoomwapen", "atoomwapens", "kernwapen", "kernwapens
comb_vec_nuc_words <- vec_corp_tot %>% nearest_to(vec_corp_tot[[comb_nuc_words]], 25) %>% names
```

Step 2b: Viewpoint vectors

- Looked at the 200 NNs of '<u>nuclear weapon</u>' in WEM based on complete corpus (again)
- Manually classify words associated with the two viewpoints: pro- and anti-proliferation

- List of +/- 20 words per viewpoint. Examples: 'defense' or 'strategic' versus 'disarmament'

```
13  non_prol_vec <- c("atoomvrij", "ontwapening")
```

Using WEMs in *historical* research: round 2

- How can we use the *discursive space* of *words* to learn something about **changes through time?**
  - 1. Model change through time by training models on (overlapping) series of corpora: 1970-75, 1972-77, etc.

- Fundamental weakness of WEMs: there is no comparability of vectors between models
  - 2. Therefore using cosine similarity of vectors for comparison

23

t
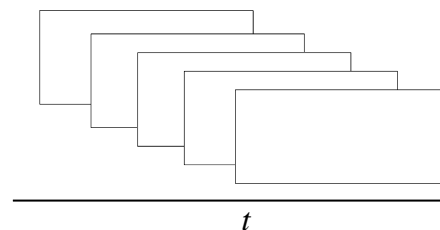
24

Step 3: Diachronic models

- Train models for each time slot (1970-75, 1972-77, etc.)…
- …on each of the party-specific corpora
- Result: 32 different sub-models/corpora

Step 4: Nuclear weapon discourse

- Use combined nuclear weapon vector to define **party- and time-specific** nuclear weapon discourse: 25 NNs of nuclear weapon in vector space…

- …for every party…
- …and for every time slot!
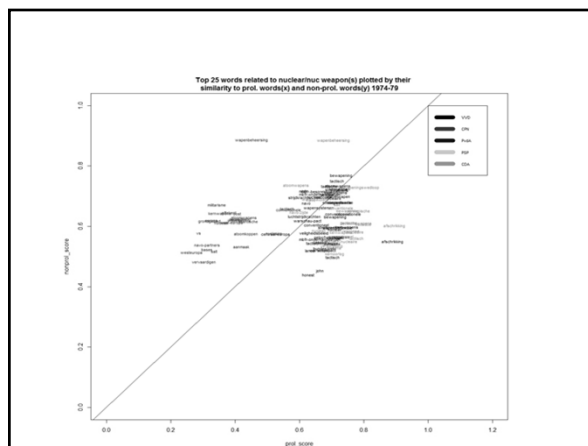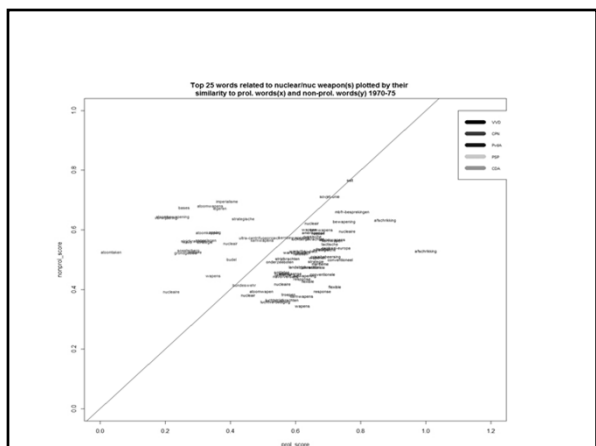
Step 5: Calculate similarity with viewpoints

- Calculate distance between vector of **each NN of nuclear weapon**…
- … and pro- and anti-proliferation vectors

- Second part of our solution: we calculated distance/angle by using <u>cosine similarity</u>
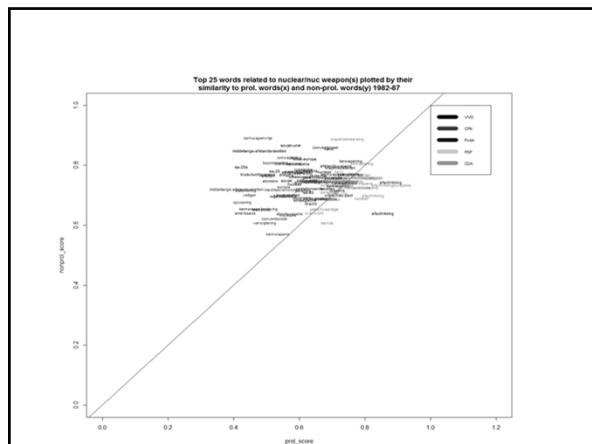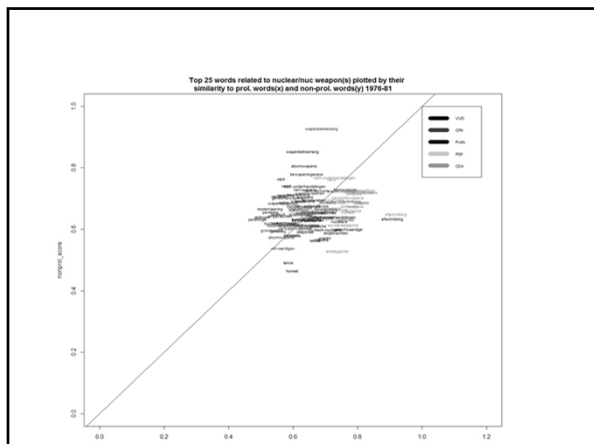- Comparison of closeness of vectors from different WEMs is now possible

So, eh, what did we do?

- For every party- and time-specific model:
  - Nuclear weapon (combined) vector
  - 25 NNs with their vectors
  - Measuring closeness in discursive space by calculating cosine similarity between the NNs vectors and viewpoint vectors

6 nonprol_score_CPN = comb_vec_kernwap %>% cosinesimilarity(veccpn[[c("verwijdering", "atoomvrij",

Top 25 words related to nuclear/nuc weapon(s) plotted by their similarity to prol. words(x) and non-prol. words(y) 1976-81

Top 25 words related to nuclear/nuc weapon(s) plotted by their similarity to prol. words(x) and non-prol. words(y) 1982-87

What have we learned?

1. Eventually, everybody caught Hollanditis
2. It just did not help at all
3. Walk the walk ≠ talk the talk
4. We can use word embeddings to investigate changes through time

Many thanks for your attention!

(milanvanl.github.io)