

# SHICO: SHIFTING CONCEPTS OVER TIME

Tracing Concepts in Dutch Newspaper Discourse using  
Sequential Word Vector Spaces

Melvin Wevers



Translantis Project  
Digital Humanities Approaches to Reference  
Cultures: The Emergence of the United States  
in Dutch Public Discourse 1890-1990



# SHICO: SHIFTING CONCEPTS OVER TIME

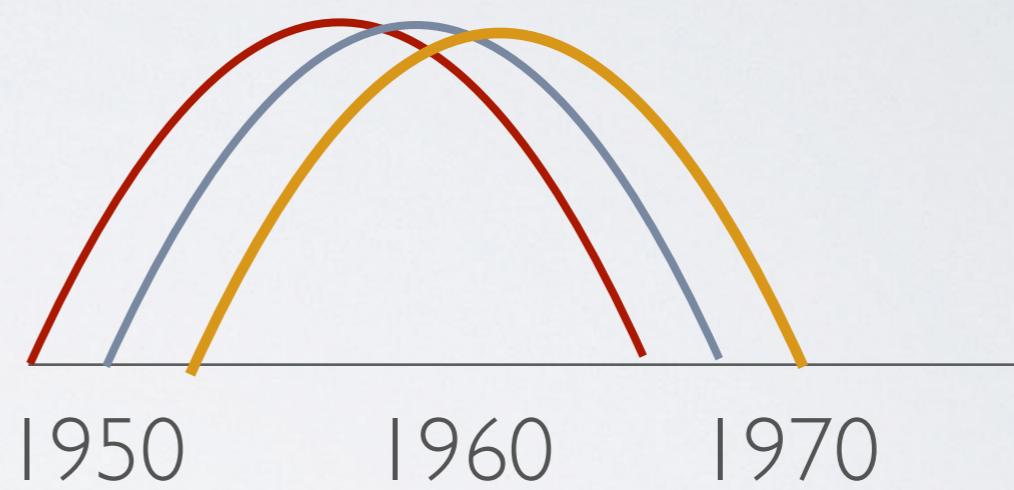
- The emergence and transformation of concepts, ideas, and thoughts
- How can we study concepts in large digitized corpora?
  - What words are used to discuss particular concepts?
  - How do these words relate to each other?
  - Can we trace the historical development of concepts?
- **How to map the genealogy of a concept within a digitized corpus?**

# EXISTING METHODS

- Use of predefined list of words (N-gram viewers / Full-text search)
- Top-down approaches (NER, word classification lists/sentiment analysis) make use pre-established models that are often a-historic
- Topic modeling is useful but quite static and abstract

# OUR APPROACH

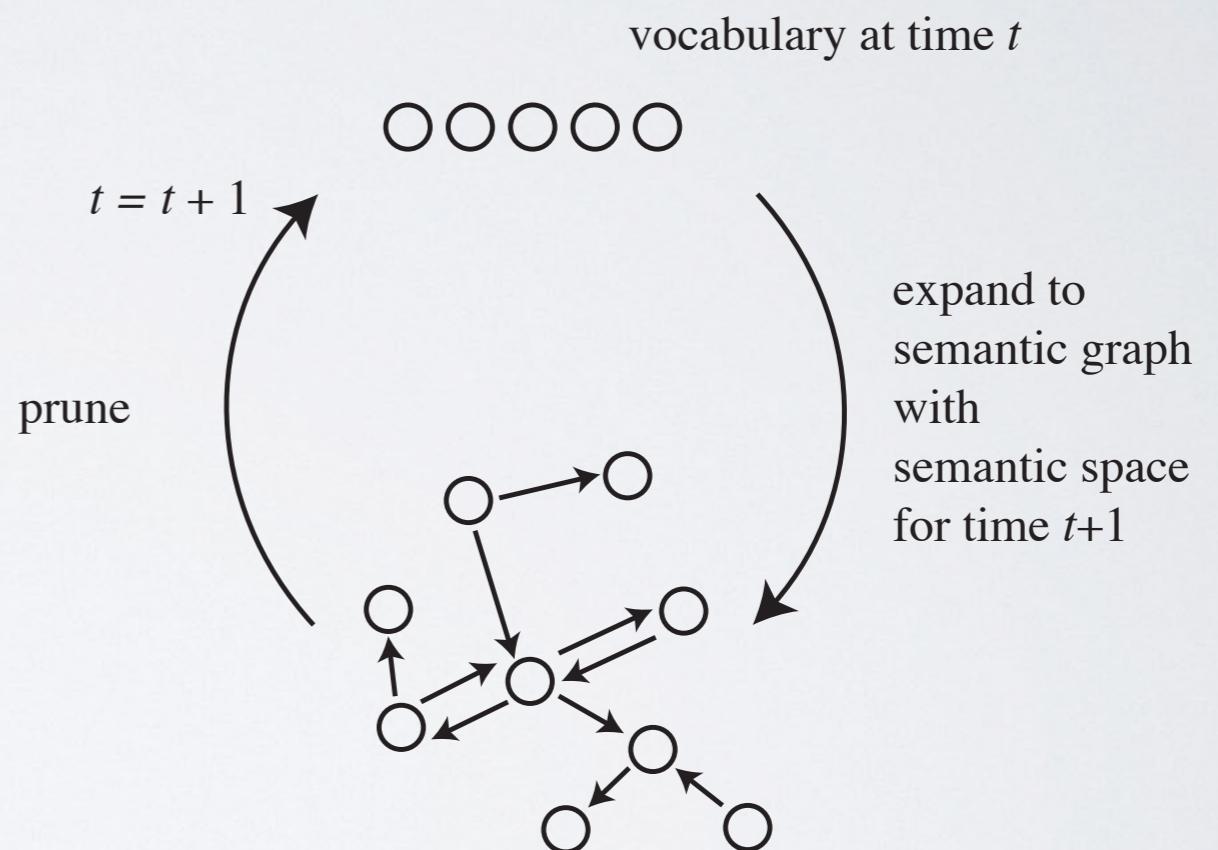
- **Data:** >600.000 digitized newspaper issues from the Dutch National Library
- **Multi-dimensional word-vector space** using Google's word2vec (neural language model)
- **Semantic** and **syntactic** information representation by **geometry** (Baroni & Kruszwesi, 2014; Wijaya & Yeniterzi, 2011)



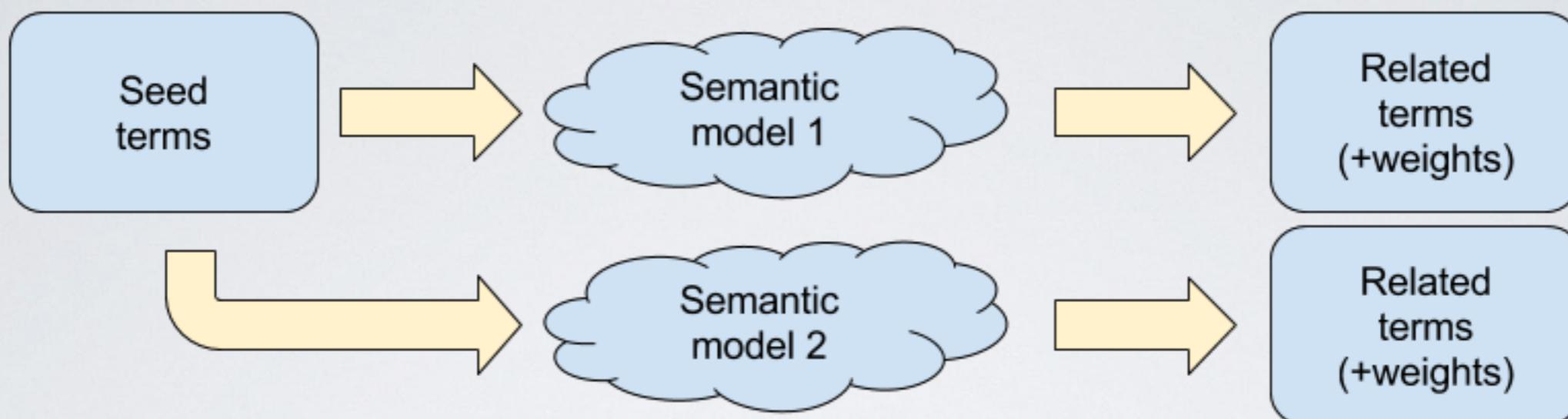
1 model = 10 years  
40 models for period  
between 1950-1990

# TRACING CONCEPTS

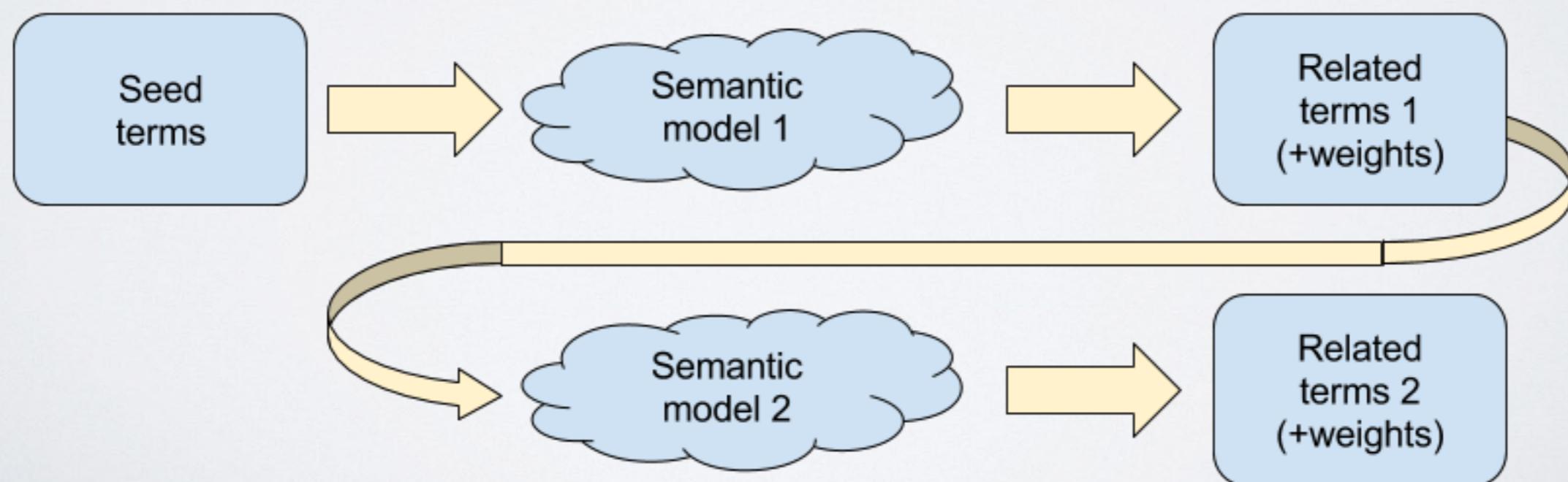
- One or more words as **entry-points** into concept
- Concept is a **network of words**
- Find **closely related words** to seed word based on distance
- **Weighting** based on frequency + sum distance
- **Tracing concepts:** which words remain and disappear from network
- See: Kenter, Wevers, Huijnen, de Rijke (2015)  
**“Ad Hoc Monitoring of Vocabulary Shifts over Time”**



## Non-adaptive Model



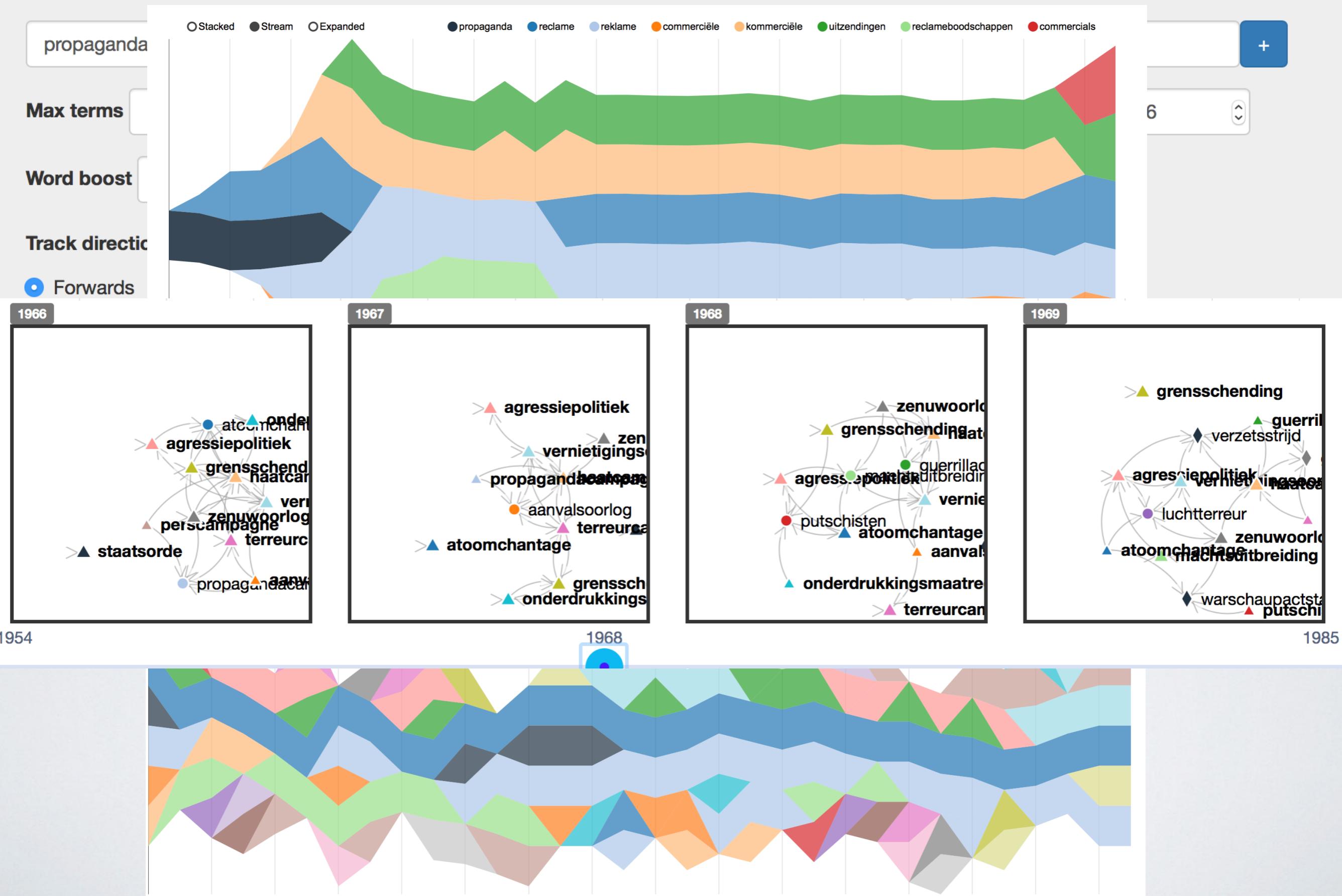
## Adaptive Model



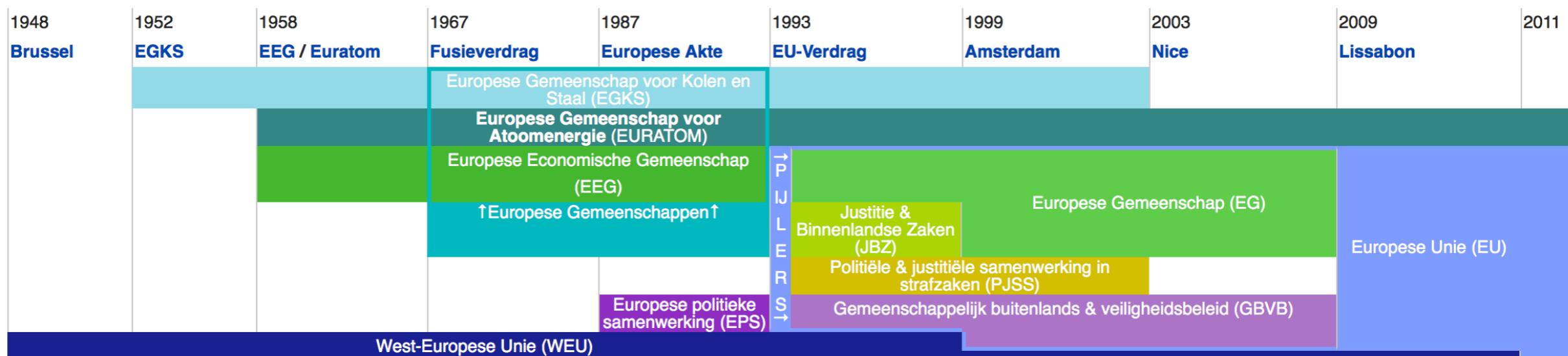
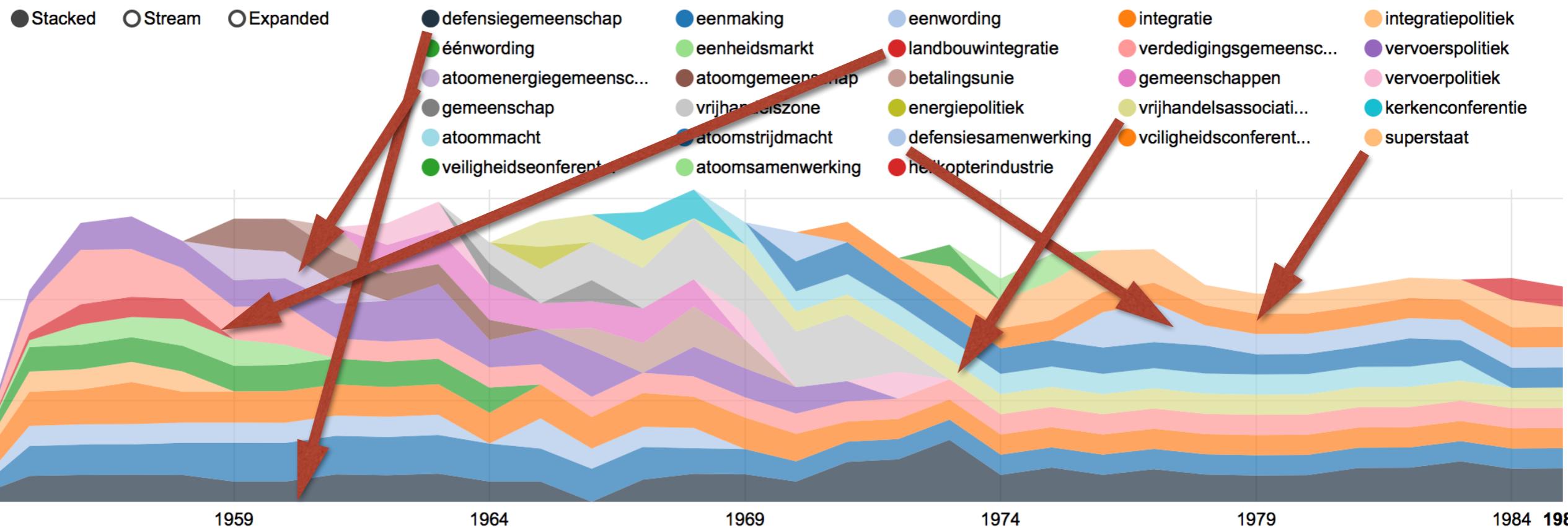
# RAW OUTPUT

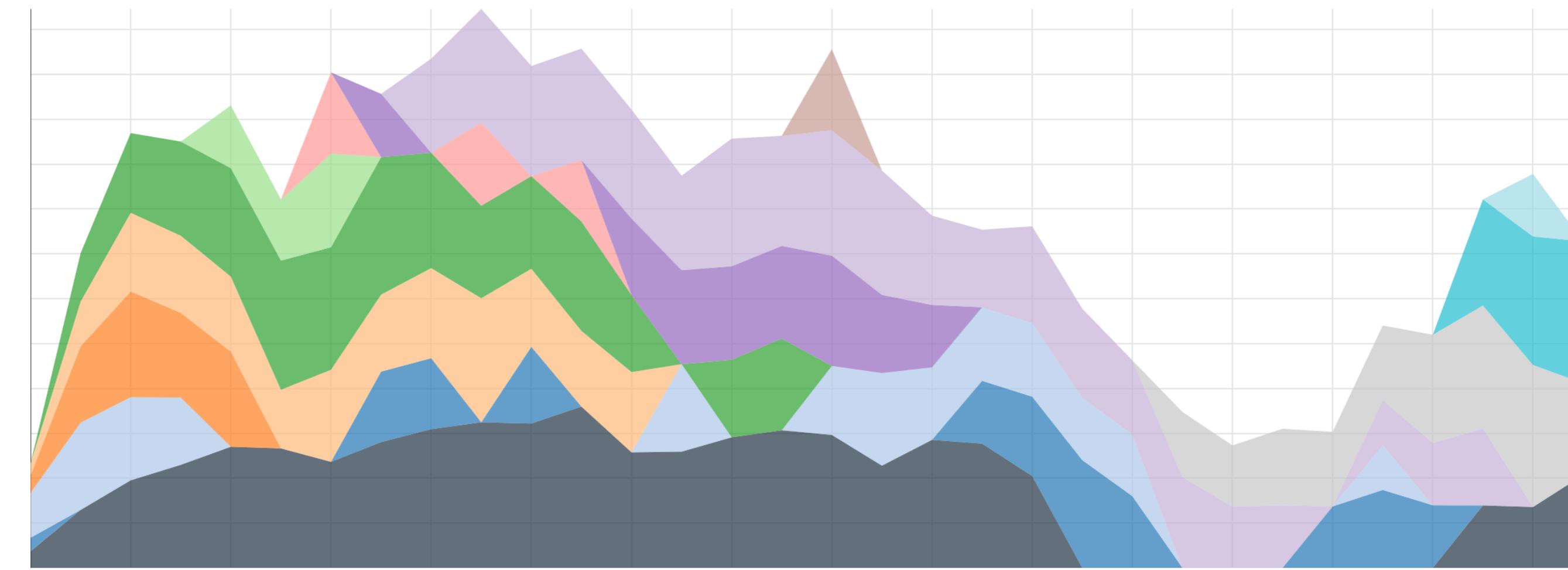
```
>>> tc.trackWord(dModels, 'buitenlanders')
1950_1959: vreemdelingen (0.76), nederlanders (0.69), indonesiërs (0.65), toeristen (0.62), europeanen (0.61), vacantiegangers (0.58), mensen (0.57), vakantiegangers (0.56),
duitsers (0.54), dagjesmensen (0.54)
1951_1960: vreemdelingen (0.76), nederlanders (0.74), toeristen (0.64), indonesiërs (0.64), europeanen (0.64), bezoekers (0.59), immigranten (0.58), duitsers (0.57), mensen
(0.57), kampeerders (0.57)
1952_1961: vreemdelingen (0.74), toeristen (0.69), nederlanders (0.68), indonesiërs (0.61), dagjesmensen (0.61), bezoekers (0.61), kampeerders (0.60), europeanen (0.59),
vakantiegangers (0.59), duitsers (0.57)
1953_1962: vreemdelingen (0.74), toeristen (0.70), bezoekers (0.64), nederlanders (0.63), vacantiegangers (0.62), kampeerders (0.59), vakantiegangers (0.59), dagjesmensen
(0.57), mensen (0.57), automobilisten (0.55)
1954_1963: toeristen (0.69), vreemdelingen (0.68), nederlanders (0.66), bezoekers (0.62), vakantiegangers (0.60), kampeerders (0.59), vacantiegangers (0.58), immigranten (0.56),
jongelui (0.55), jongeren (0.55)
1955_1964: toeristen (0.70), vreemdelingen (0.70), nederlanders (0.64), vacantiegangers (0.64), bezoekers (0.63), kampeerders (0.63), vacantiegangers (0.59), mensen (0.59),
dagjesmensen (0.56), jongelui (0.55)
1956_1965: vreemdelingen (0.71), toeristen (0.70), vakantiegangers (0.64), kampeerders (0.63), nederlanders (0.62), bezoekers (0.62), mensen (0.61), duitsers (0.57),
vakantiegangers (0.56), gezinnen (0.56)
1957_1966: vreemdelingen (0.68), toeristen (0.68), nederlanders (0.63), kampeerders (0.62), vakantiegangers (0.60), mensen (0.59), bezoekers (0.58), duitsers (0.57), sportvissers
(0.56), vacantiegangers (0.55)
1958_1967: toeristen (0.71), vreemdelingen (0.71), nederlanders (0.68), vakantiegangers (0.64), kampeerders (0.63), bezoekers (0.60), marokkanen (0.59), duitsers (0.58),
dagjesmensen (0.58), mensen (0.57)
1959_1968: toeristen (0.69), nederlanders (0.68), vreemdelingen (0.66), kampeerders (0.62), bezoekers (0.61), vacantiegangers (0.61), vakantiegangers (0.58), sportvissers (0.58),
hotelgasten (0.57), mensen (0.57)
1960_1969: toeristen (0.72), vreemdelingen (0.70), nederlanders (0.68), kampeerders (0.61), vakantiegangers (0.61), zakenmensen (0.59), marokkanen (0.59), mensen (0.59),
zakenlieden (0.58), bezoekers (0.58)
1961_1970: vreemdelingen (0.71), toeristen (0.68), nederlanders (0.65), kampeerders (0.63), vakantiegangers (0.62), reizigers (0.61), marokkanen (0.59), bezoekers (0.59),
vakantiegangers (0.59), mensen (0.58)
1962_1971: vreemdelingen (0.71), nederlanders (0.68), toeristen (0.67), kampeerders (0.63), indonesiërs (0.59), vakantiegangers (0.59), dagjesmensen (0.59), marokkanen (0.58),
sportvissers (0.57), vakantiegasteren (0.57)
1963_1972: vreemdelingen (0.72), nederlanders (0.71), toeristen (0.68), indonesiërs (0.62), kampeerders (0.61), mensen (0.58), gezinnen (0.58), scandinaviërs (0.58), turken
(0.57), duitsers (0.57)
1964_1973: nederlanders (0.69), vreemdelingen (0.67), toeristen (0.66), surinamers (0.62), indonesiërs (0.62), marokkanen (0.61), sportvissers (0.60), turken (0.59), mensen (0.58),
antillianen (0.57)
1965_1974: nederlanders (0.73), vreemdelingen (0.71), toeristen (0.64), marokkanen (0.62), turken (0.60), kampeerders (0.59), indonesiërs (0.59), surinamers (0.59), spanjaarden
(0.57), duitsers (0.56)
1966_1975: nederlanders (0.70), vreemdelingen (0.69), toeristen (0.68), indonesiërs (0.64), prostituées (0.61), marokkanen (0.60), gezinnen (0.59), mensen (0.59), surinamers
(0.58), kampeerders (0.58)
1967_1976: nederlanders (0.71), toeristen (0.65), indonesiërs (0.63), vreemdelingen (0.63), chilenen (0.57), surinamers (0.57), kampeerders (0.57), gezinnen (0.57), duitsers (0.56),
jongelui (0.55)
1968_1977: nederlanders (0.72), vreemdelingen (0.68), toeristen (0.64), vakantiegangers (0.62), kampeerders (0.62), indonesiërs (0.59), duitsers (0.59), loeristen (0.58), mensen
(0.58), tunesiërs (0.58)
1969_1978: nederlanders (0.73), vreemdelingen (0.72), toeristen (0.66), surinamers (0.63), indonesiërs (0.61), tunesiërs (0.59), guyanezen (0.58), gezinnen (0.58), chilenen (0.58),
```

## Concept search



# INTEGRATION





Mentally disturbed > adaptive model  
 Discourse moves to more euphemistic terms

# IDEAS



- Find better ways to conceptualize networks of words as concepts
- Ways to detect change in concepts?
- Backbone extraction??
- How to improve visualization of conceptual change???

# Thanks to:

- Carlos Martinez-Ortiz (eScience Research Engineer)
- Pim Huijnen (Cultural History, University of Utrecht)
- Tom Kenter (Information and Language Processing Systems, University of Amsterdam)

[https://github.com/c-martinez/ShiCo/  
blob/explanations/docs/howItWorks.md](https://github.com/c-martinez/ShiCo/blob/explanations/docs/howItWorks.md)