

Of All People and Places



*Gendered and local asymmetries in
media constructed networks in Finnish Newspapers
1820–1910*

**Agata Dominowska, Alan Granadino, Antti Kanner, Joni Kokko, Mikko Laitinen, Petri Leskinen,
Jani Marjanen, Andrei Odintcov, Ilona Pikkanen, Jaakko Raunamaa, Risto Turunen**

- Named Entity Recognition (NER) is a method for tagging words in digitized corpora as names of entities: people, places, organization and so on.
- The group's aim was to investigate what kind of insights the use of pre-tagged data can yield from three different perspectives

Main/primary data sources

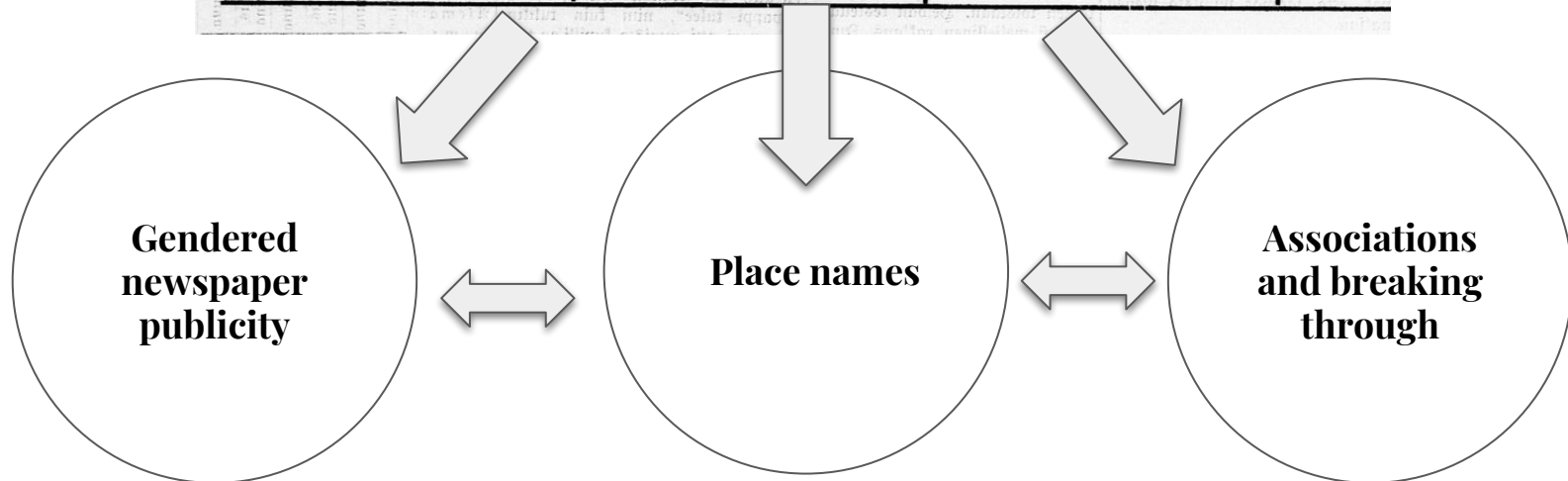
The data for the group's work is the National Library of Finland's digitized newspaper collection, which contains all of the newspapers and periodicals published in Finland from 1770 to 1910.

NER tagged version of the collection is provided by the Language Bank of Finland and shared through the open KORP service.



C. Työväen puolueen vaaliliitto.

7.	10.	11.	
Lahti.	Hausjärvi.	Tammela.	So
Sosialidemokratia.	Sosialidemokratia.	Sosialidemokratia.	So
Sivenius, A., toimittaja, Lahti.	Vuolijoki, Sulo, toimittaja, Hauho.	Nuorteva, Santeri, toimittaja, Tammela.	Jalapehti.
Laine, Maria Sofia, rouva, Helsinki.	Glöckert, Juho Viktor, toimittaja, Hausjärvi.	Holkanen, Evert, mäkätupa-lainen, Urjala.	Lainevarin.
Sumela, J. N., satulaheppä, Lahti.	Holkanen, Evert, mäkätupa-lainen, Urjala.	Kaltonen, Fredrik, torppari, Someroniemi.	Wihe torpp



Connections to external data sources

Proper names have natural connections outside the corpus.

They can be combined to outside sources to enrich the data.

Enriched data sets can be again combined with each other.

• Nimiarkisto
Namnarkivet
Nammaarkiiva
Nõmmarkiiv
Nommâarkkâdâh



Tools and analysis



Data source
(primary)



Data set collection



Data set clean-up
and slicing

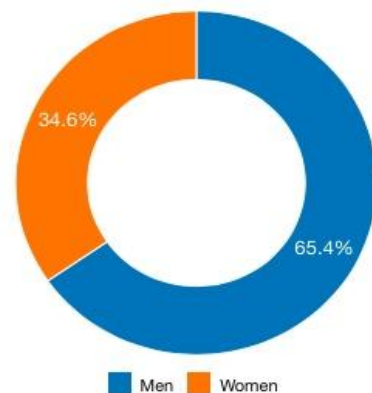
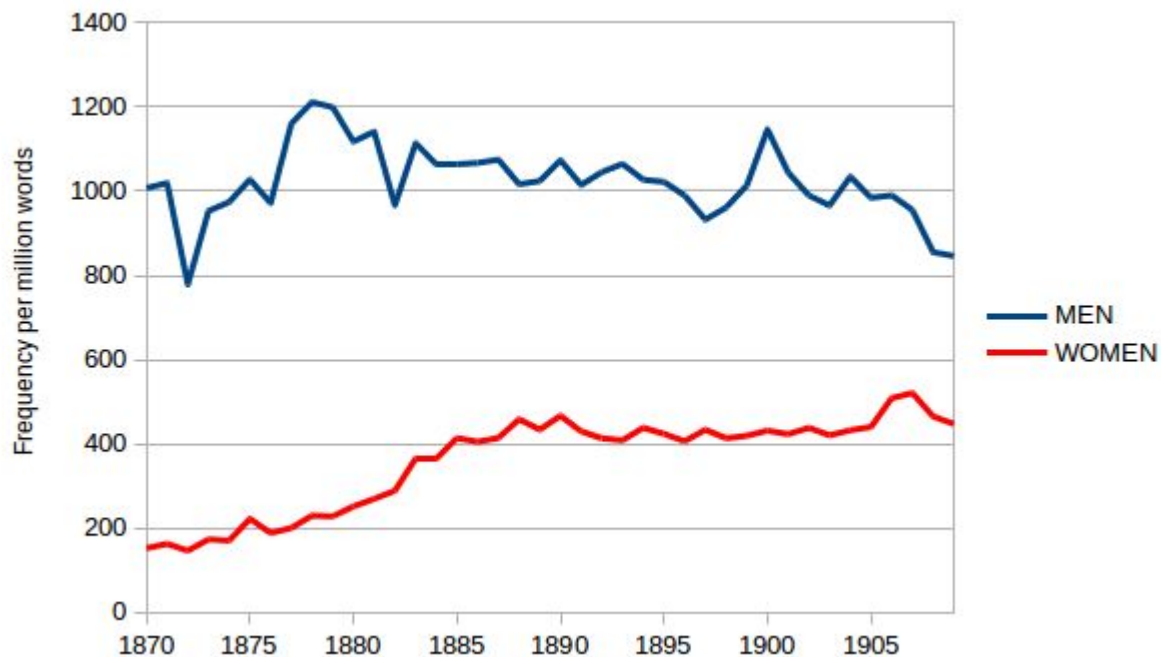


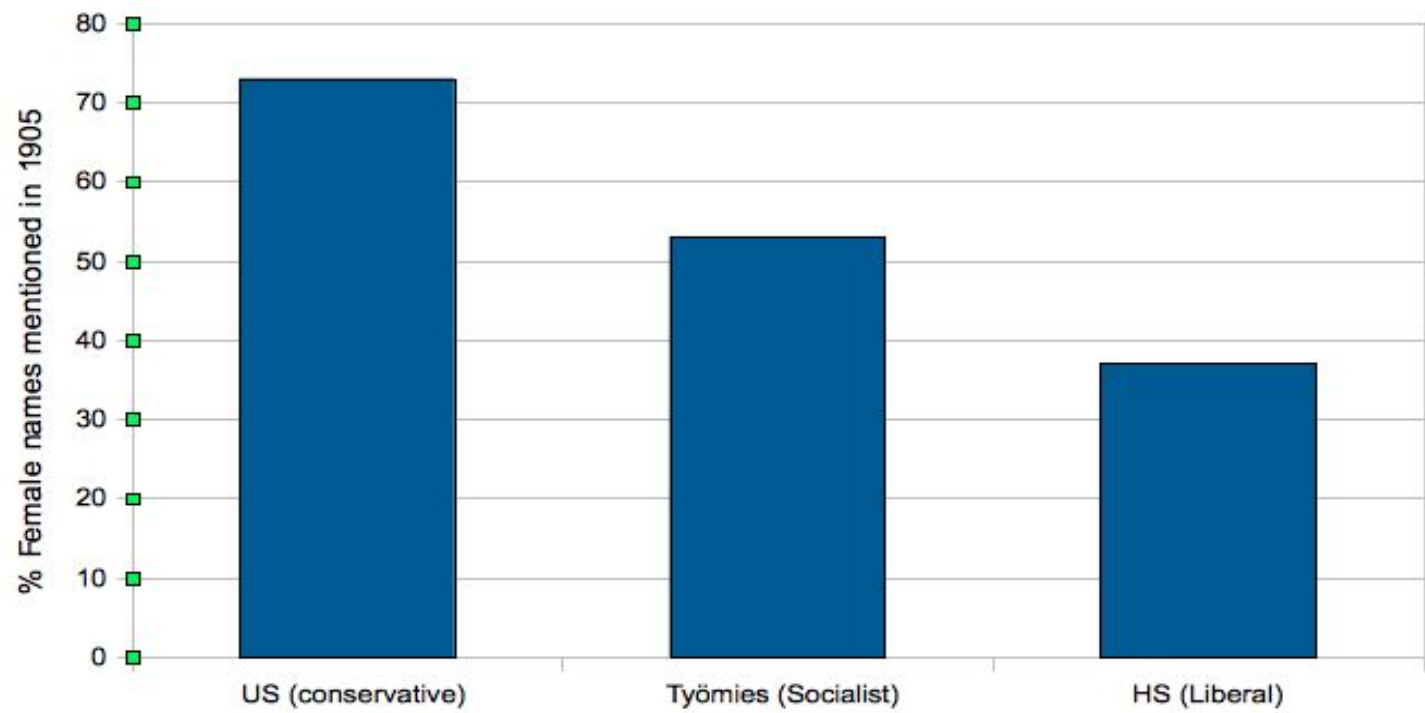
Calculations | analysis | visualization

Naisten osuus HS.fin jutuissa

Luku kertoo, kuinka suuri osuus mainituista henkilöistä on ollut naisia edellisen 24 tunnin aikana julkaistuissa jutuissa.

28 %

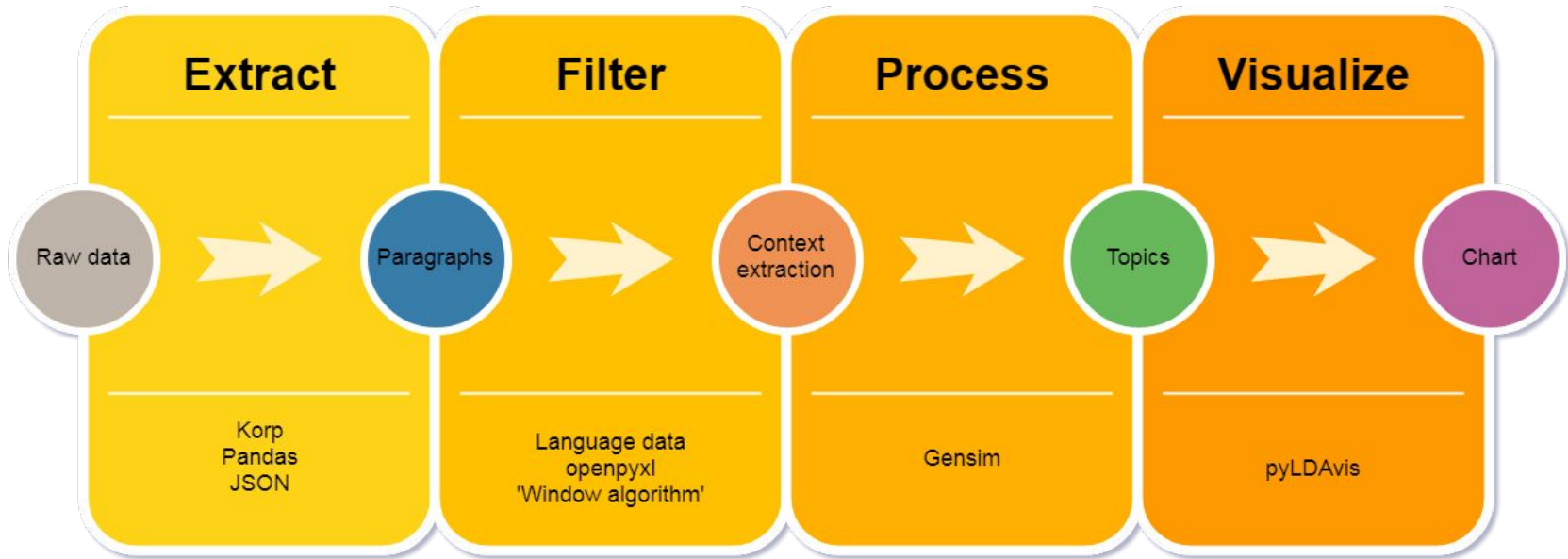




wala, 65 w.; — funnallislautafunnan est-
 mies T. Wahmanen, 60 w.; — ratsutilalli-
 nen R. Iso-Ruikka, 50 w.; — maanwil-
 jelijä E. Bergius, 81 w.; — talollinen J.
 Rantalainen, 58 w.; — maanwiljelijä H.
 Toimiatnen, 54 w.; — maanwiljelijä W.
 Danielshaga, 64 w.; — agronomi G.
 Wellingt, 29 w.; — ent. tilanomistaja A.
 Pesonen, 94 w.; — ent. tilallinen R. Hei-
 nikkala, 97 w.; — maanwiljelijä S. Koli,
 78 w.; — tilanomistaja, hovinneuvos D.
 Hedlund, 62 w.; — tilanomistaja A. Finne,
 45 w.

Naisia: kauppaneuvoksen leski Anna
 Sinebryhoff, 73 w.; — neiti Julia Wil-
 helmína Grenqvist, 74 w.; — tyttökoulun
 ent. johtajatar Albertine Westerstråhle, 86
 w.; — rehtorin leski Juliana Wilhelmina
 Grenman, 72 w.; — leskirouwa Amalia

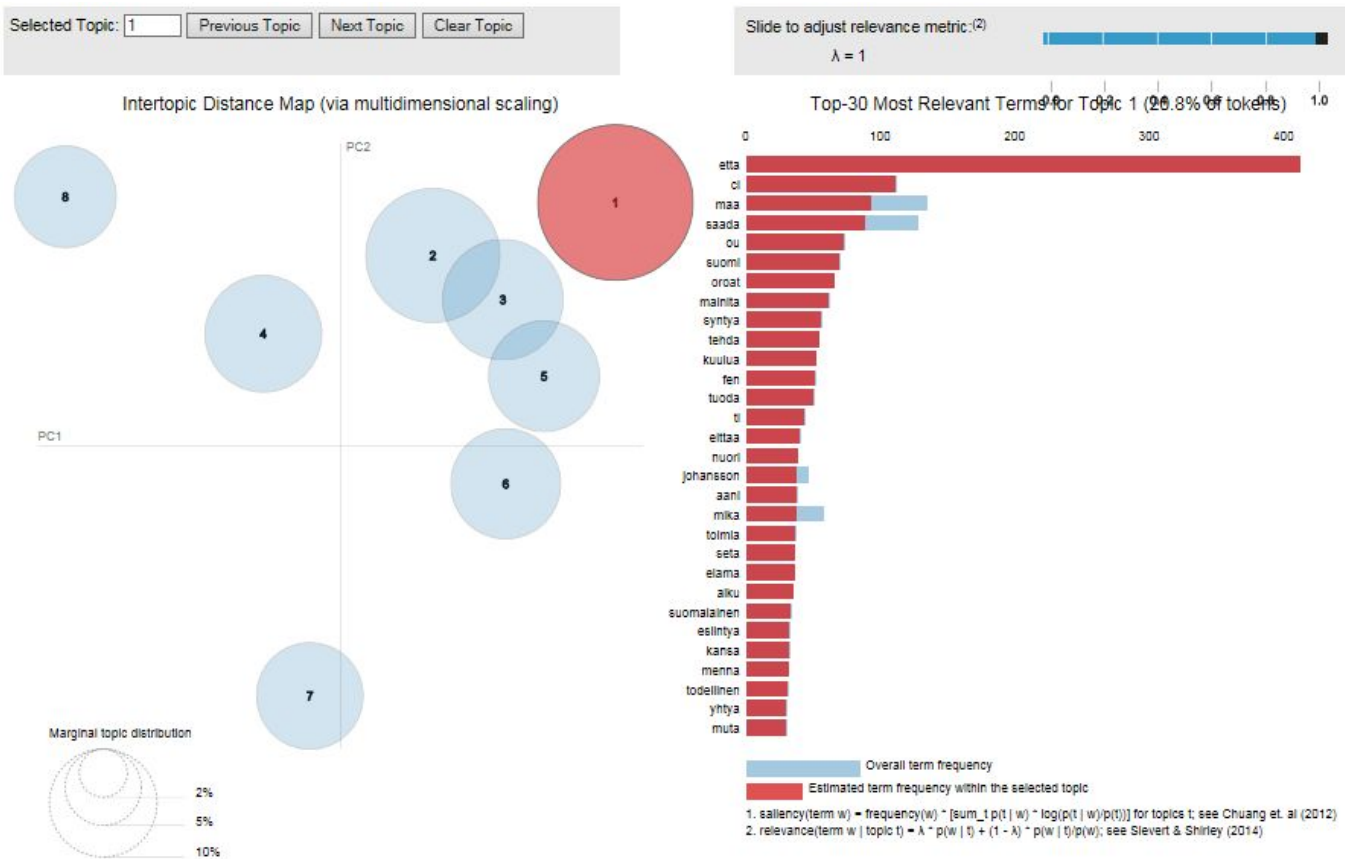
Topic modelling



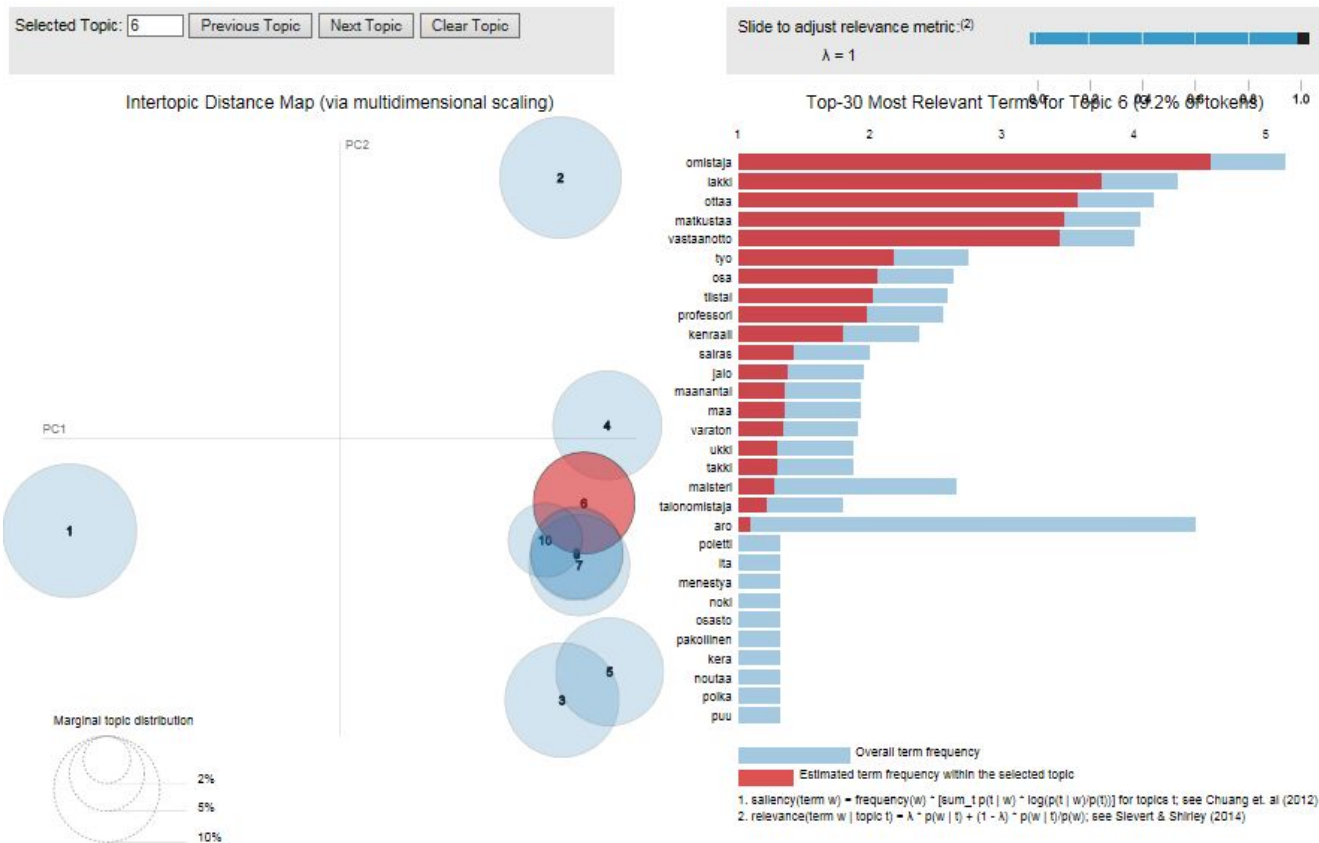
word word word word word word word word NAME word word word word word word.

Window size = 11

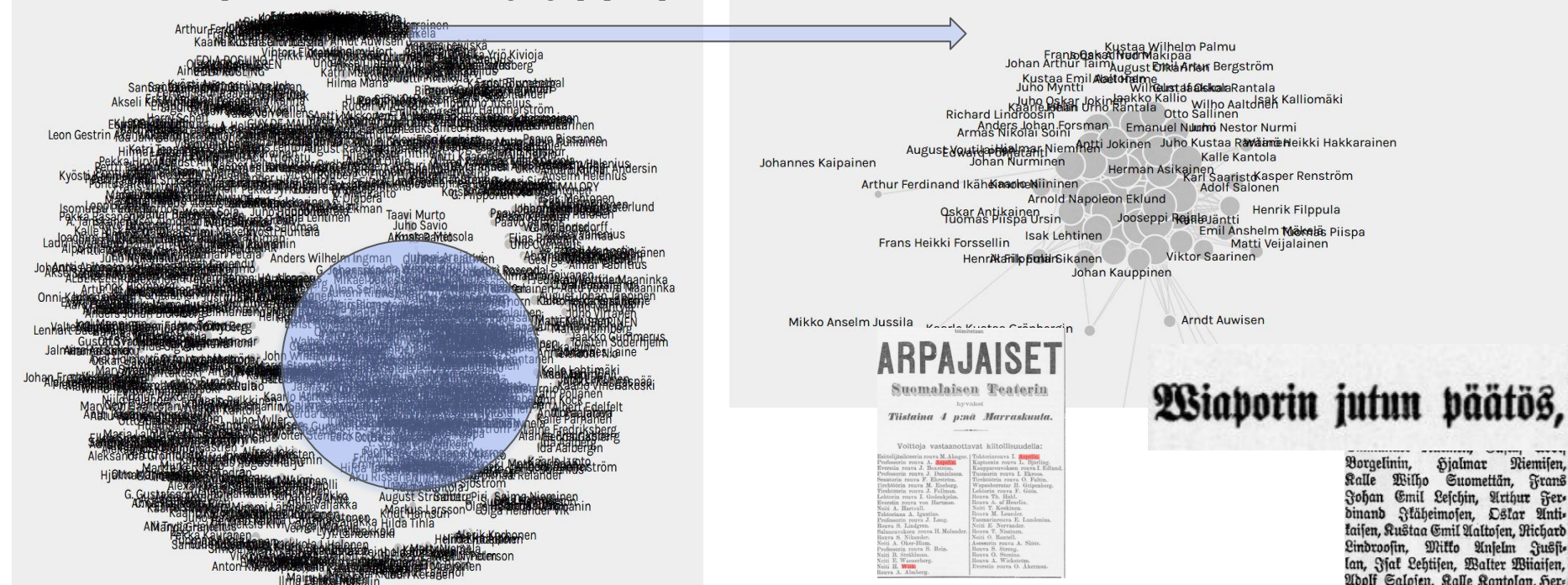
Results: topic modelling for female names 1905



Results: topic modelling for male names 1905

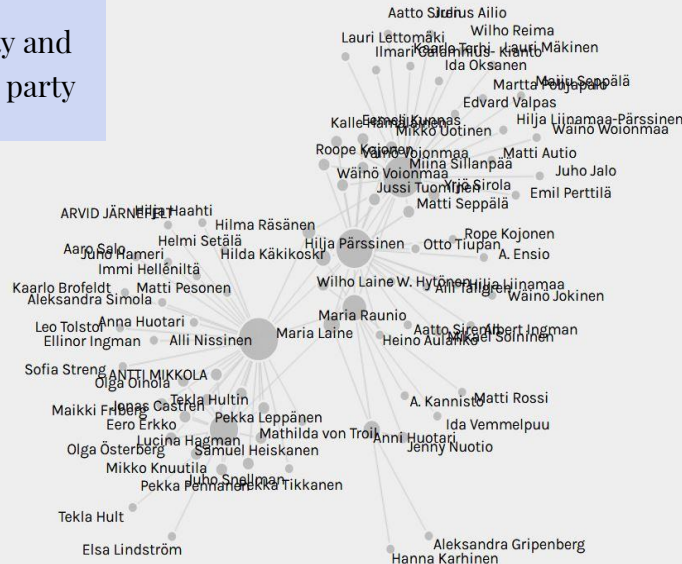


Co-occurrences of persons in all Finnish-language papers published in 1907



- Aim to to identify patterns in the networks that would provide ideas about the social capital of prominent figures.
- In practice, we learned more about how the structure of newspapers produced connections between people.
- 2,383 associations in toto 1907.
- The inner ring tells us about highly linked people, whereas the outer ring tells us about ephemeral and less complex associative structures.
- Many of the associative clusters were occasional: advertisements, election lists, contemporary politics and media events » pragmatic associations

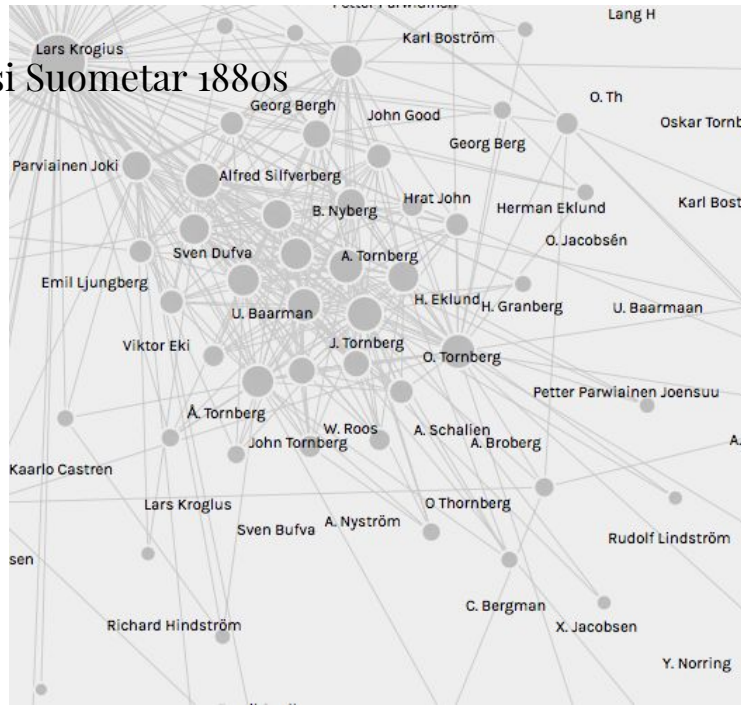
Female MPs from the Nuorsuomalaiset party and the Social democratic party



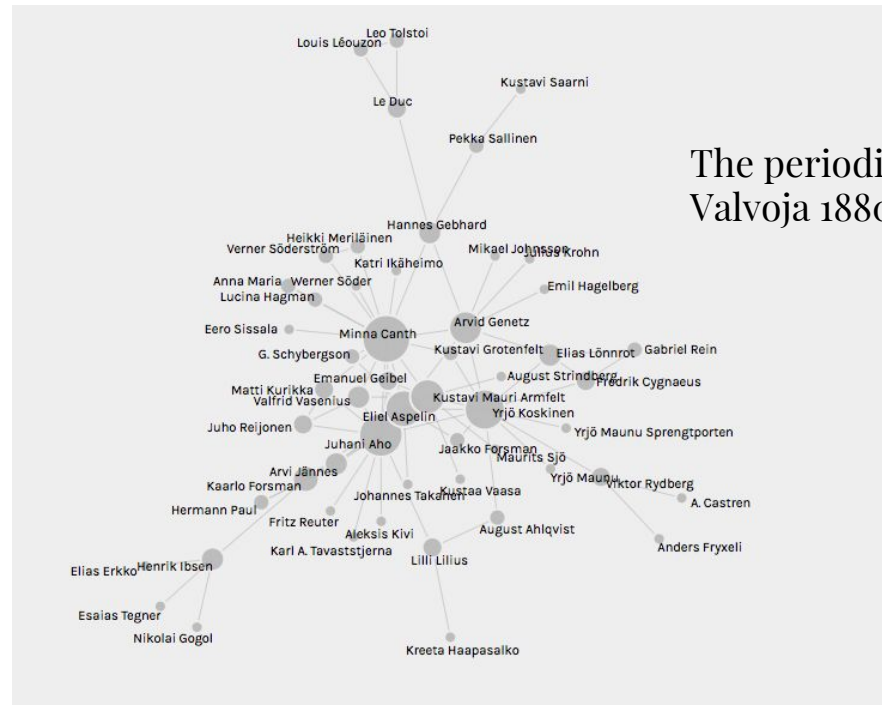
Beyond short-term networks 1 » Politics

- Concentrating on a selection of persons » less “noise”, more meaningful associations.
- Associations in 1907 between female MPs from the first unicameral parliament in 1907
- Female MPs were unsurprisingly more often associated with other women (both MPs and outside the parliament) than was the case with men.
- As compared with male MPs they come across as pivotal figures in their own party as measured by associations to other people.
- Second order affinities suggest that female MPs from the Social Democratic party were more closely associated with MPs from the Finnish party than the Young Finnish party.

Uusi Suometar 1880s



The periodical Valvoja 1880s



Beyond short-term networks 2–3? » Longer perspective (a decade) // From newspapers to periodicals?

- A longer perspective shows us more persistent links, but visualizing change was more difficult.
- Periodicals contain less advertisements » less “noise”, more meaningful associations & possibility to identify authors and how they produced networks
- They provide a possibility to look at social capital in the making by focussing especially on foreign canonical authors and how Finnish authors were associated with them
- With Valvoja, we could also try to identify authors to get a more fine-grained picture

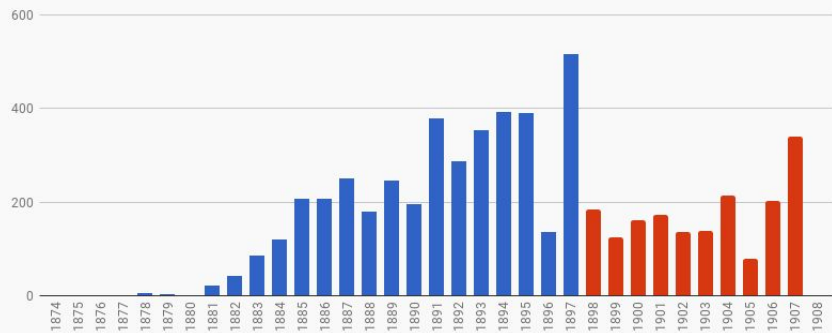
Newspaper mentions of people in National Biography

- Extracted 307 people from Biographies who lived in 1830–1910
 - with known years of birth and death
 - with 47 different types of profession
- Querying the Newspaper database for person mentions
 - All newspaper material between 1850–1909
 - the query used a person's full name; name variants, e.g. with changed family name, or initials, were not concerned
 - produced histograms based on number of mentions by year and age
 - histograms for individual person and grouped by professions

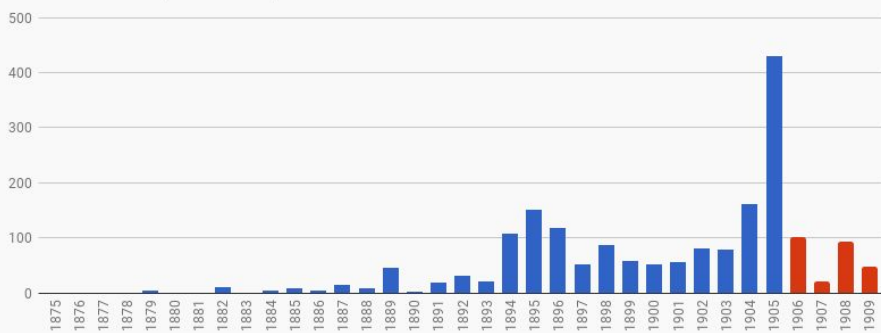
Newspaper mentions of people in National Biography

Individual person with a timespan of her/his lifetime and ten years after

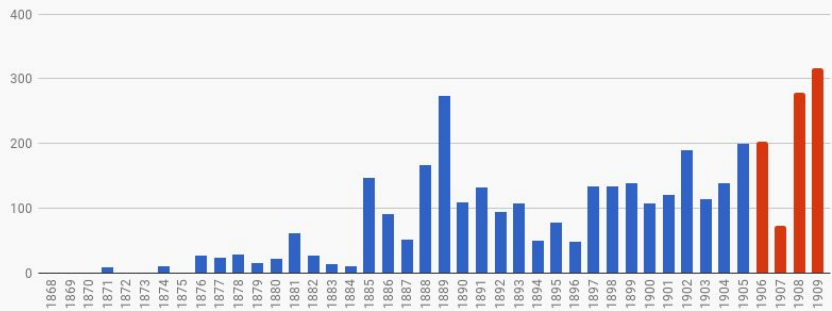
Canth Minna (1844-1897)



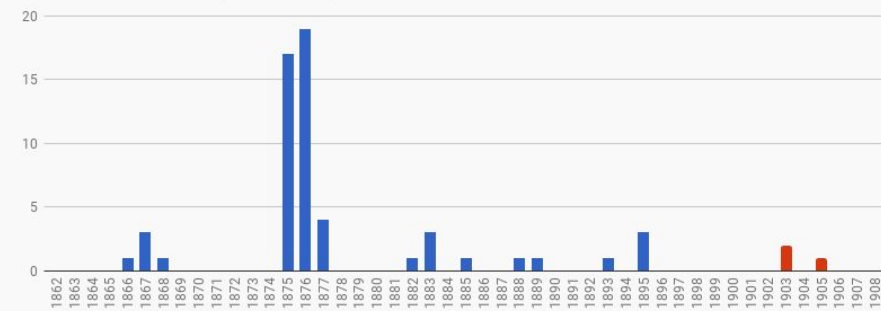
Edelfelt Albert (1854-1905)



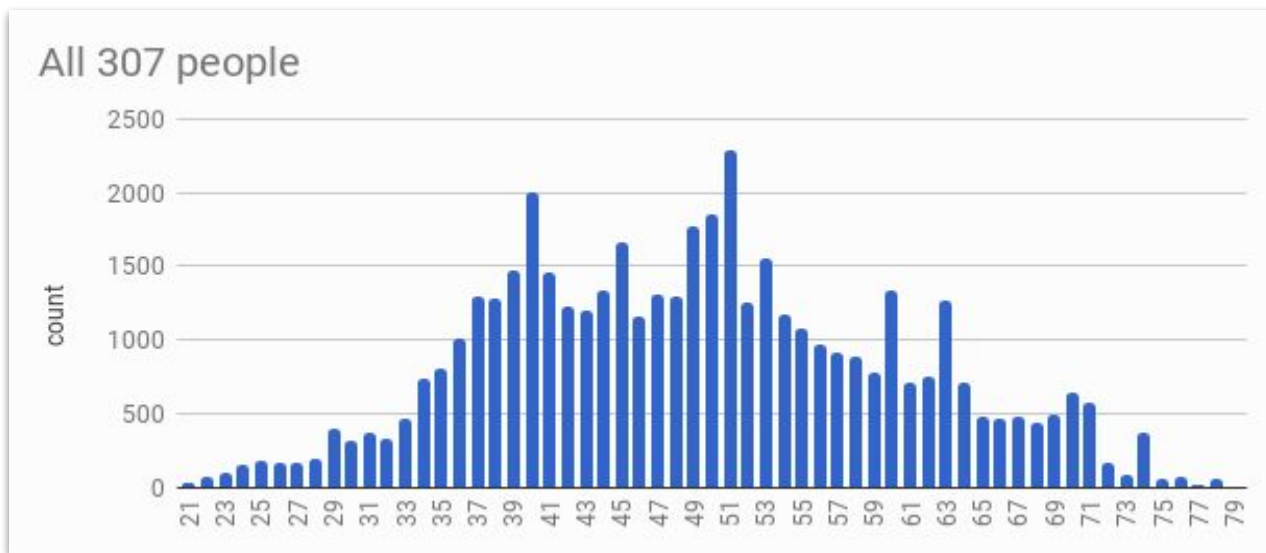
Forselius Victor (1838-1905)



Gestrin Emil Teodor (1831-1895)



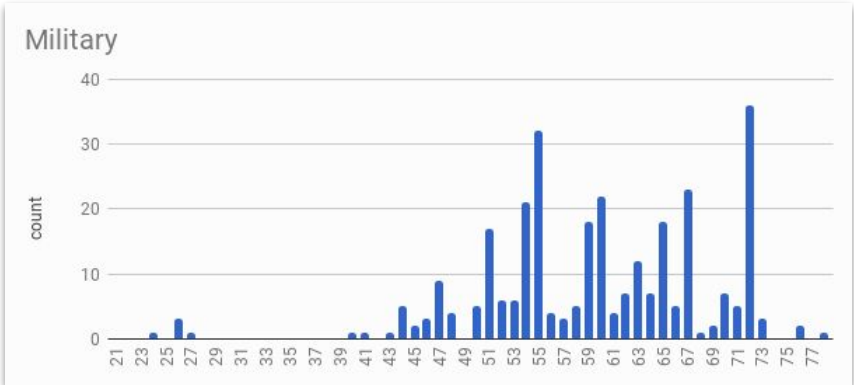
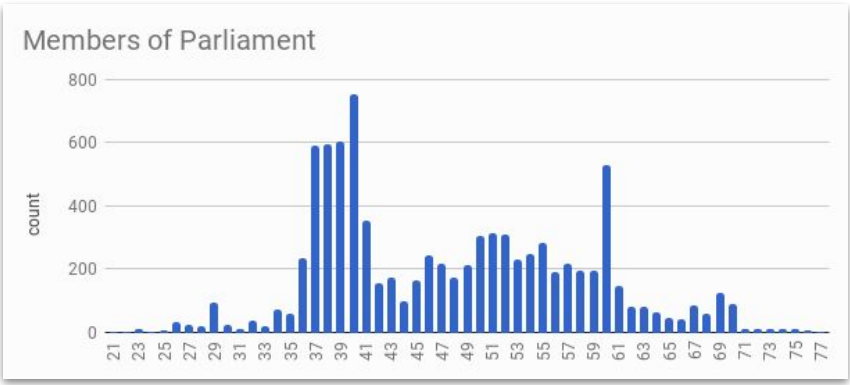
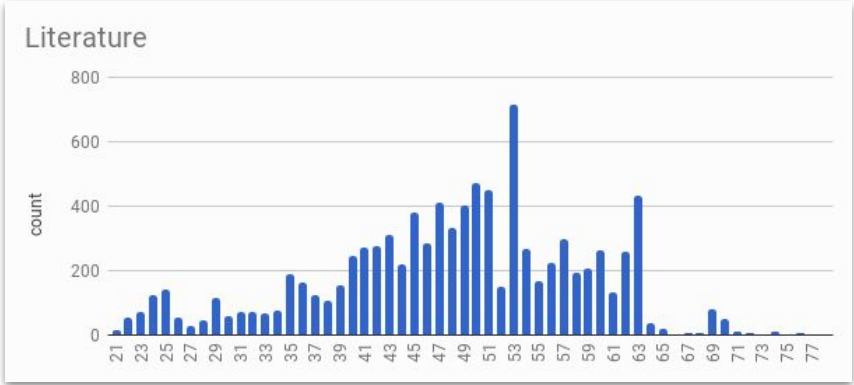
Newspaper mentions of people in National Biography



5th percentile	mean	95th percentile
30 years	47 years	66 years

Newspaper mentions of people in National Biography

Grouped by professions, 3 examples:



profession/age	5th	mean	95th
Members of Parliament	26	36	47
Literature	26	49	63
Military	42	46	60

Places and newspapers



Querying NERitized
place name lemmas.
The most common
name "li". -> ☹

NER-material was
compared to multiple
Finnish place name
sources including
coordinates.

```
Created on 28.5.2018

@author: Johnneagle
...

import json, sys, os

def getLocations(scVPath):
    scv = {}

    with open(scVPath, "r", encoding="utf-8") as f:
        for line in f:
            parse = line.split(',')
            cords = [parse[2], parse[3]]
            name = parse[0].split(',')[1]

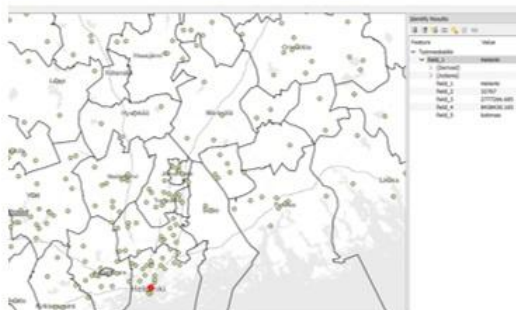
            if name not in scv:
                scv[name] = cords

    return scv

def makeMap(givenJson, givenCsv, filePath):
    path = givenJson

    with open(path, "r", encoding="utf-8") as f:
        data = json.load(f)

    res = {}
```



	A	B	C	D	E
1	Name	Hits	X	Y	Finland/Abroad
2	Helsinki	117383	2777267	8438430	kotimaa
3	Suomi	103659	2894307	9349764	kotimaa
4	Turku	80521	2478736	8500996	kotimaa
5	Tampere	63735	2645095	8741019	kotimaa
6	Kuopio	46003	3082318	9074058	kotimaa
7	Oulu	43498	2835541	9612003	kotimaa
8	Pietari	40234	3374836	8388614	ulkomaa
9	Tukholma	36573	2011389	8251936	ulkomaa

Results

The results shows which place names were used in the newspapers and where these places are. Basically, many kind of conclusion can be made based on the place name data produced.

Examples:

- The amount of Finnish or foreign places mentioned in newspapers

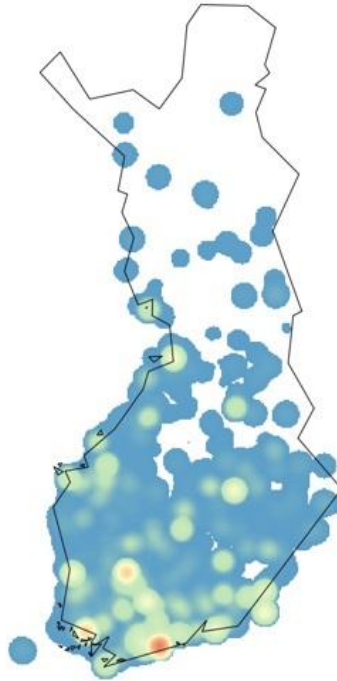
	prosent	
27300	0.728135919	Finland
10193	0.271864081	Abroad

Työmies (1870-1910)

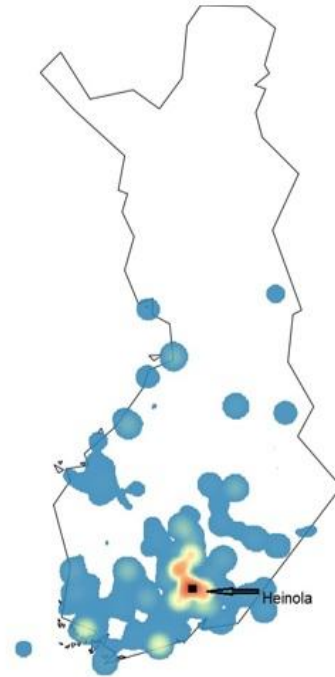
hits	prosent	
7205	0.930998837	Finland
534	0.069001163	Abroad

Heinolan Sanomat (1908-1909)

Examples:



*Heat Map 1:
Place names mentioned in the paper
Työmies (1895–1910).
Most of the mentions are from bigger
towns: Helsinki, Tampere, Turku, Viipuri
and Kuopio.*

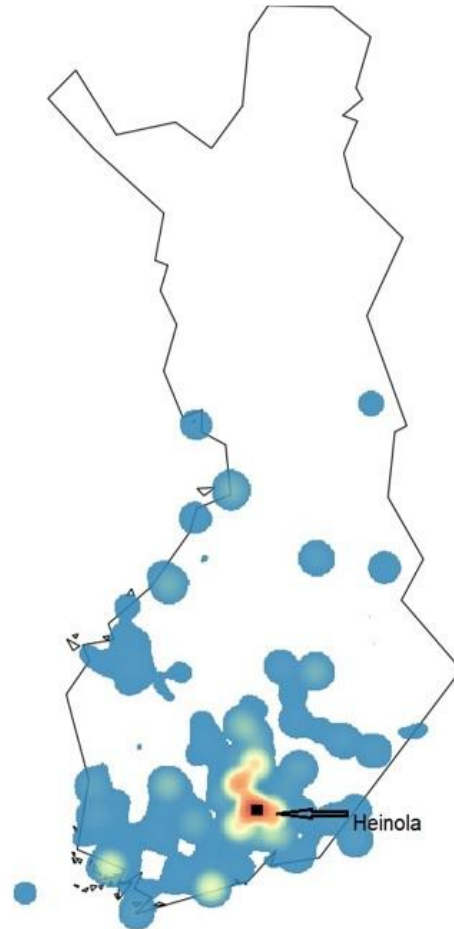


*Heat Map 2: Place names mentioned in the
paper Heinolan Sanomat (1908-1909).*

Further research possibilities

The data could be used, for example, to study the area of distribution (to whom the paper was aimed) that different old newspapers had.

Heinolan Sanomat, here as an example



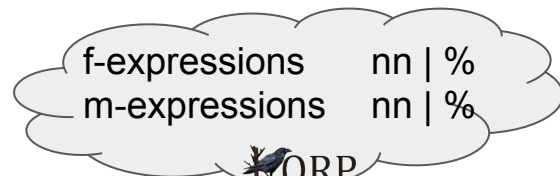
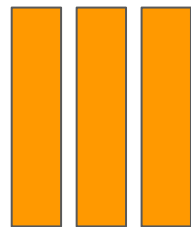
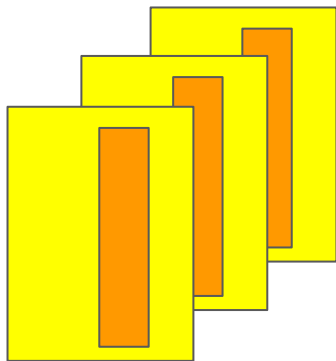
Conclusions & Further work

Proper names provide a viable access point to newspaper data from many different perspectives

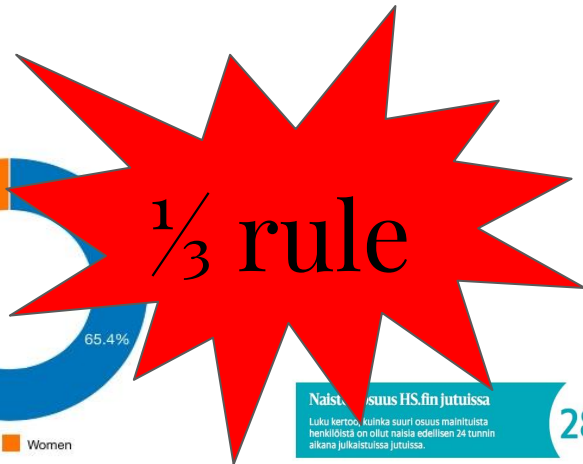
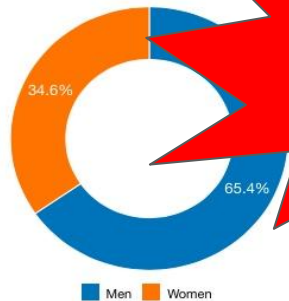
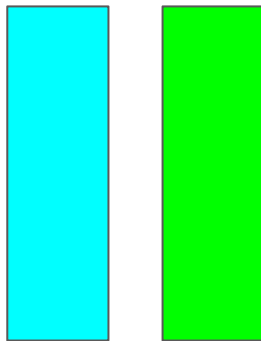
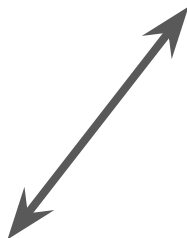
- Density of information
- Links to outside information sources
- Centrality in newspaper reporting

The group was able to set up workflows and make interesting exploratory observations. Interpreting and examining them further would follow next.

Thank you!



f-names	nn		%
m-names	nn		%
u-names	nn		%



Naisten osuus HS.fi:n jutuissa
Luku kertoo, kuinka suuri osuus mainituista henkilöistä on ollut naisia edellisen 24 tunnin aikana julkaistussa jutussa.

28%