

# Of All People and Places

## Gendered and local asymmetries in media constructed networks in Finnish Newspapers 1820-1910

**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.



The data for the group's work is the National Library of Finland's digitized newspaper collection, which contains almost all of the newspapers and periodicals published in Finland from 1770 to 1910. A NER tagged version of the collection is provided by the Language Bank of Finland and shared through the open KORP service.

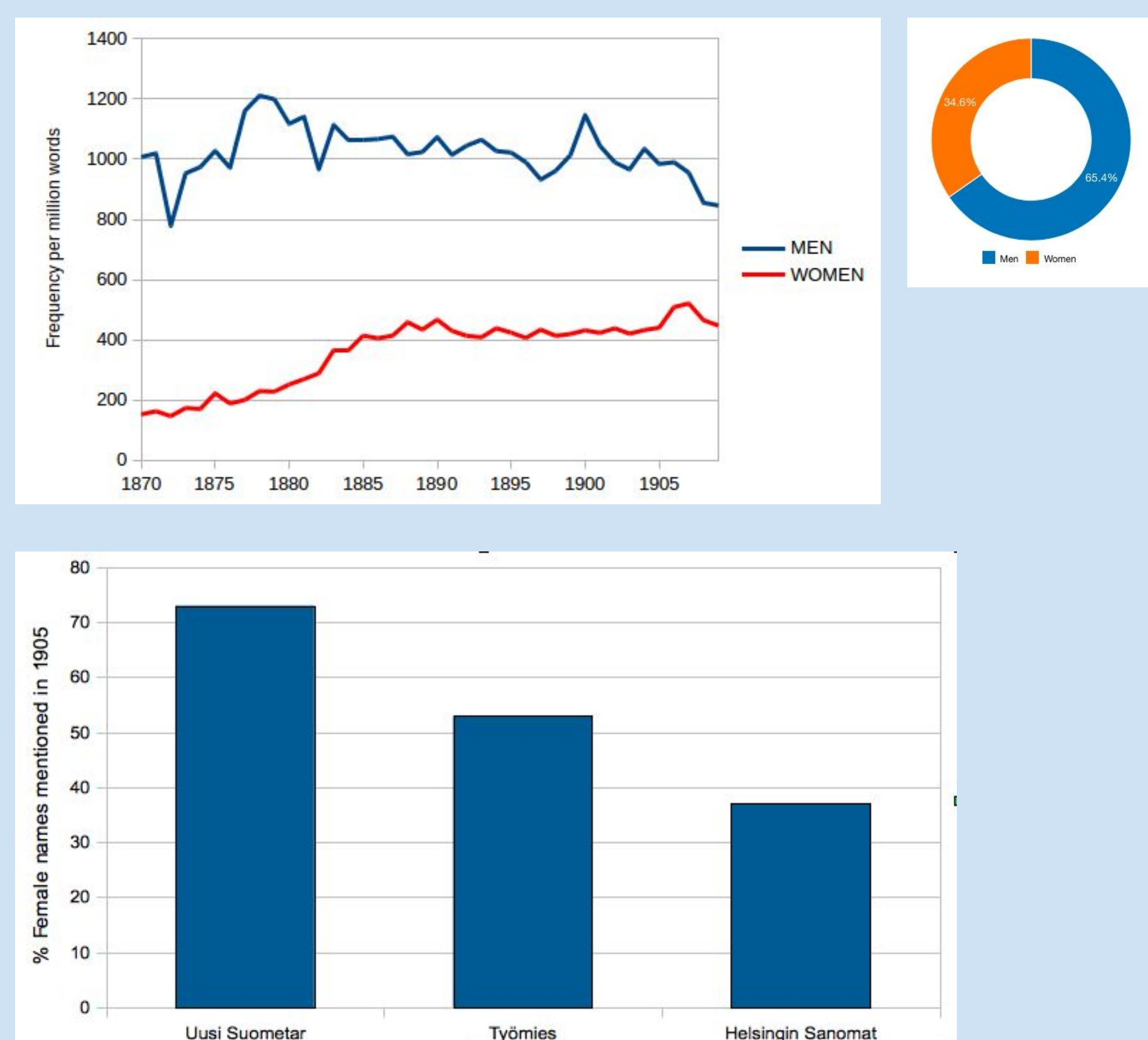
### Gendered newspaper publicity

Were there gender asymmetries in three biggest Finnish-language newspapers (liberal, conservative, and socialist) before and immediately after the establishment of universal suffrage in Finland (1870–1910)?

#### Visualized results

Women's presence in the new media with respect to men is consistently lower. Our results compared with existing research on other times and places hint towards an iron law of 30%.

Almost all data supports this claim, but, however 1905, one year before women gained full suffrage in Finland, seems to be an exception:



### Current shortcomings

Our results show a structural under representation of women in the 19th century newspapers that, surprisingly, is very similar to the present day. At the same time, they show a radical variation in 1905, which difficult to explain.

#### Next steps

We need to think creatively, in order to make sense out of the complex phenomenon of gender representation. Using complementary approaches (analyzing both personal names and common nouns related to gender) seems like a reasonable starting point.

### Conclusions

Examining names is very viable way to investigate different kinds of phenomena in the newspaper data, and can be used to enrich the data by building eg. regional profiles of individual newspapers, as was shown by the place name strand.

Unlike many other linguistic features, names can be used to easily combine different data sources. This was evident in all strands, as all of them utilized combinations of the corpus NER entries and external data sources.

Names also contain much of content related information, as, for example, different people occur in different contexts. Compared to often vague or ambiguous results derived from investigation of word use, people seem to provide quite a robust image of themes and events discussed in texts.

### Place names in the newspapers

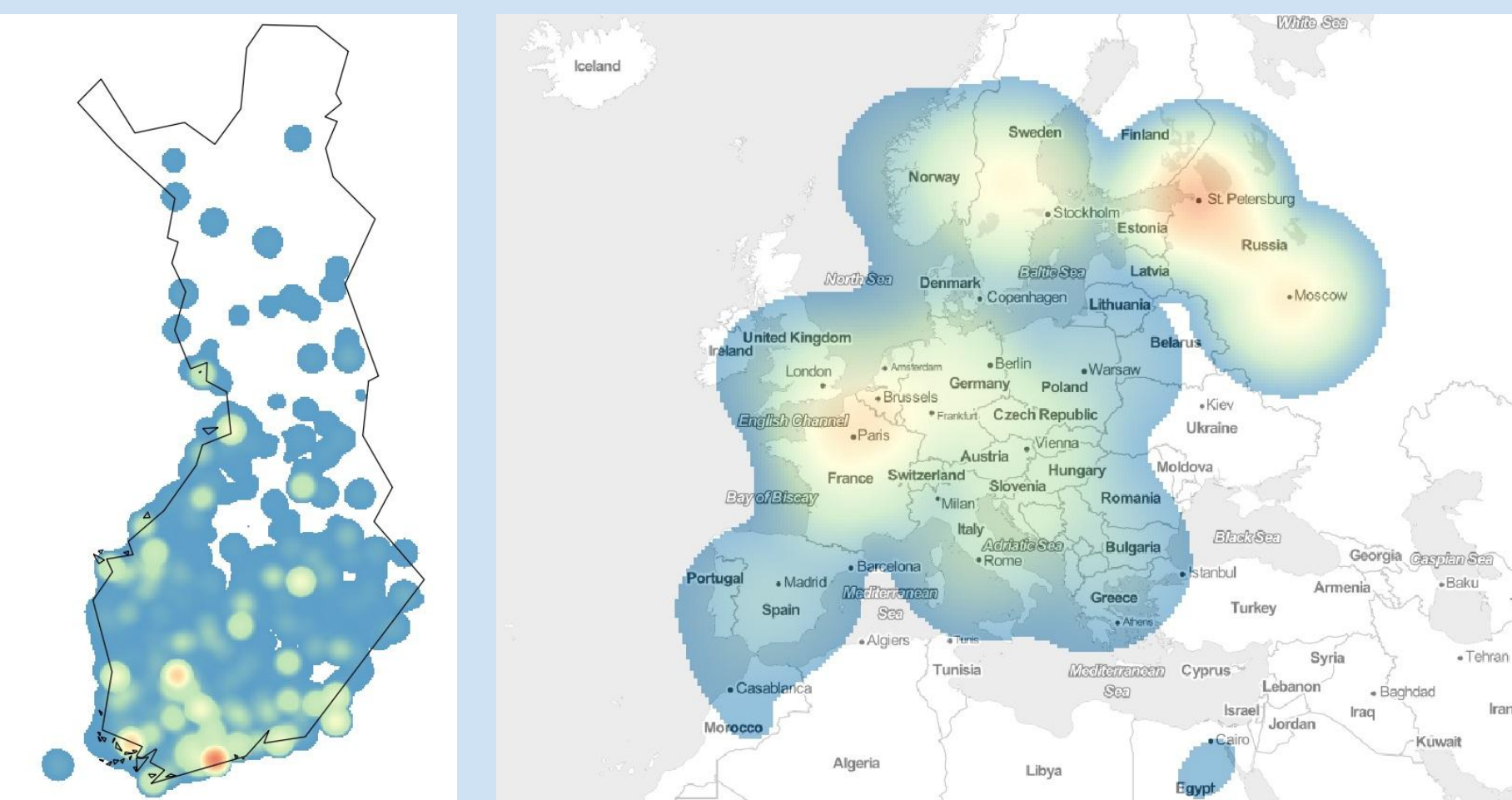
What kind of places are mentioned in different newspapers? How nationally or internationally oriented different newspapers have been? Do regional profiles of papers emerge from the place name data?

To investigate these questions, we used place name data regarding Finnish-language newspapers in KORP. This material was compared to multiple Finnish place name sources for filtering and enriching the data with eg. locations.

#### Visualized results

The results shows which place names were used in the newspapers and where these places are.

Name	Hits	X	Y	Finland/Abroad
Helsinki	117383	2777267	8438430	kotimaa
Suomi	103659	2894307	9349764	kotimaa
Turku	80521	2478736	8500996	kotimaa
Tampere	63735	2645095	8741019	kotimaa
Kuopio	46003	3082318	9074058	kotimaa
Oulu	43498	2835541	9612003	kotimaa
Pietari	40234	3374836	8388614	ulkomaa
Tukholma	36573	2011389	8251936	ulkomaa



Heat Map 1:  
Place names mentioned in the paper Työmies (1870–1910). Most of the mentions are from bigger towns: Helsinki, Tampere, Turku, Viipuri and Kuopio.

Heat Map 2:  
Mentions of countries and capital cities in Työmies (1870–1910).

### Current shortcomings

This study produced large amount of useful data. The biggest challenge encountered was how to visualize so much data in understandable form.

#### Next steps

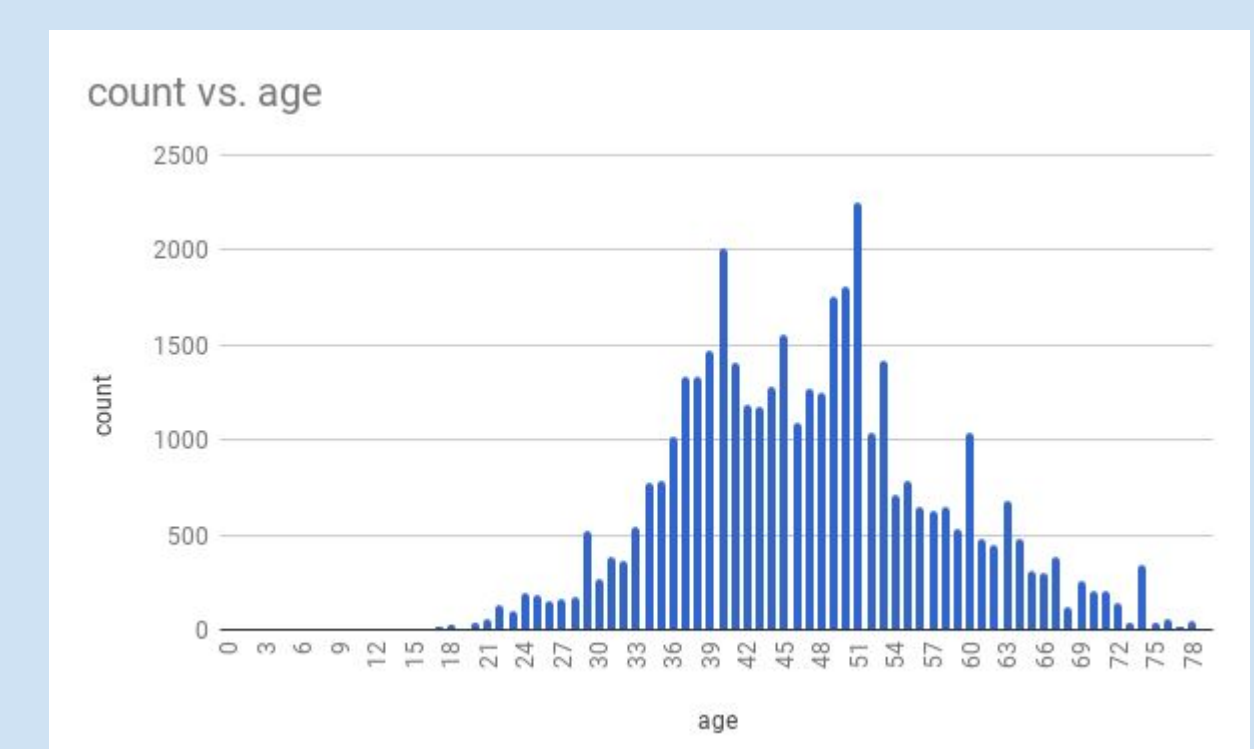
The data could be used, for example, to study the area of distribution that different old newspapers had.

### Breaking through in the newspapers

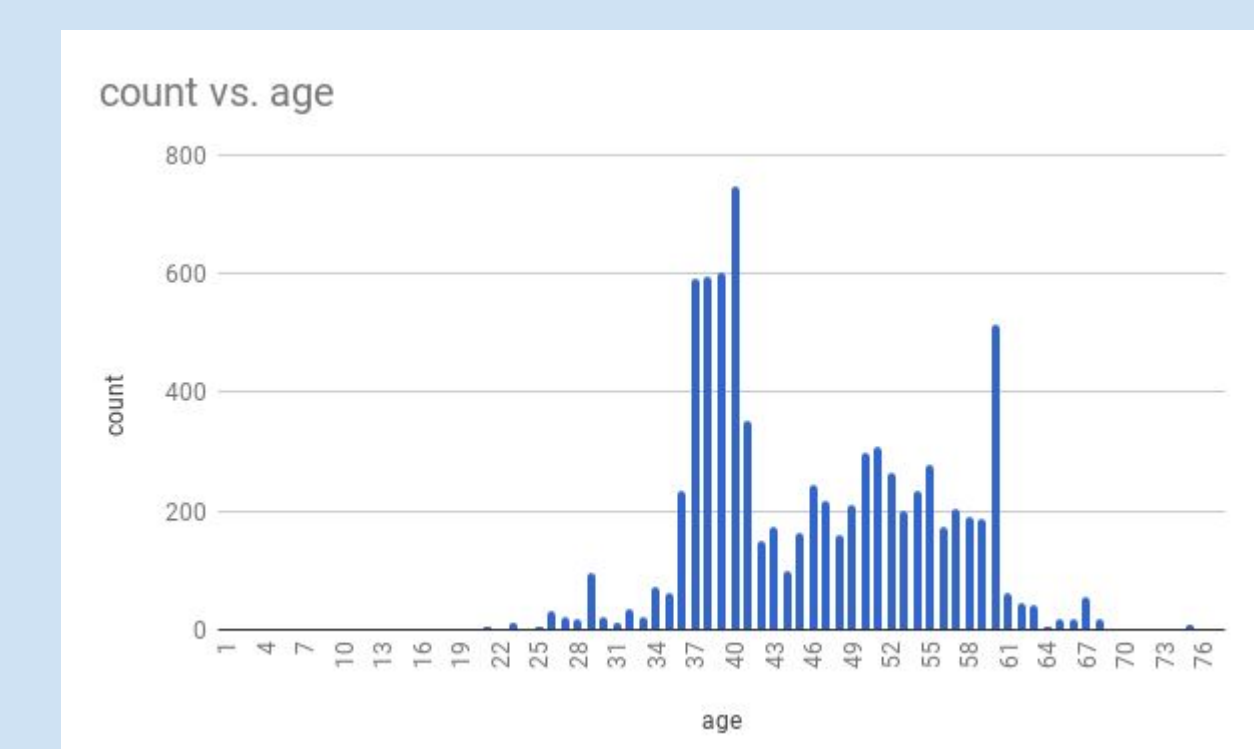
At which age did particular persons emerge as public figures in the newspapers?

We took a selection of people born after 1830 and deceased before 1910 and counted their occurrences in the Finnish-language newspapers and compared the results to the canonized list of persons in Finnish history (the Finnish National Biography, SKS). We also compared overall results to certain occupational groups, and ran individual case studies (e.g. Minna Canth).

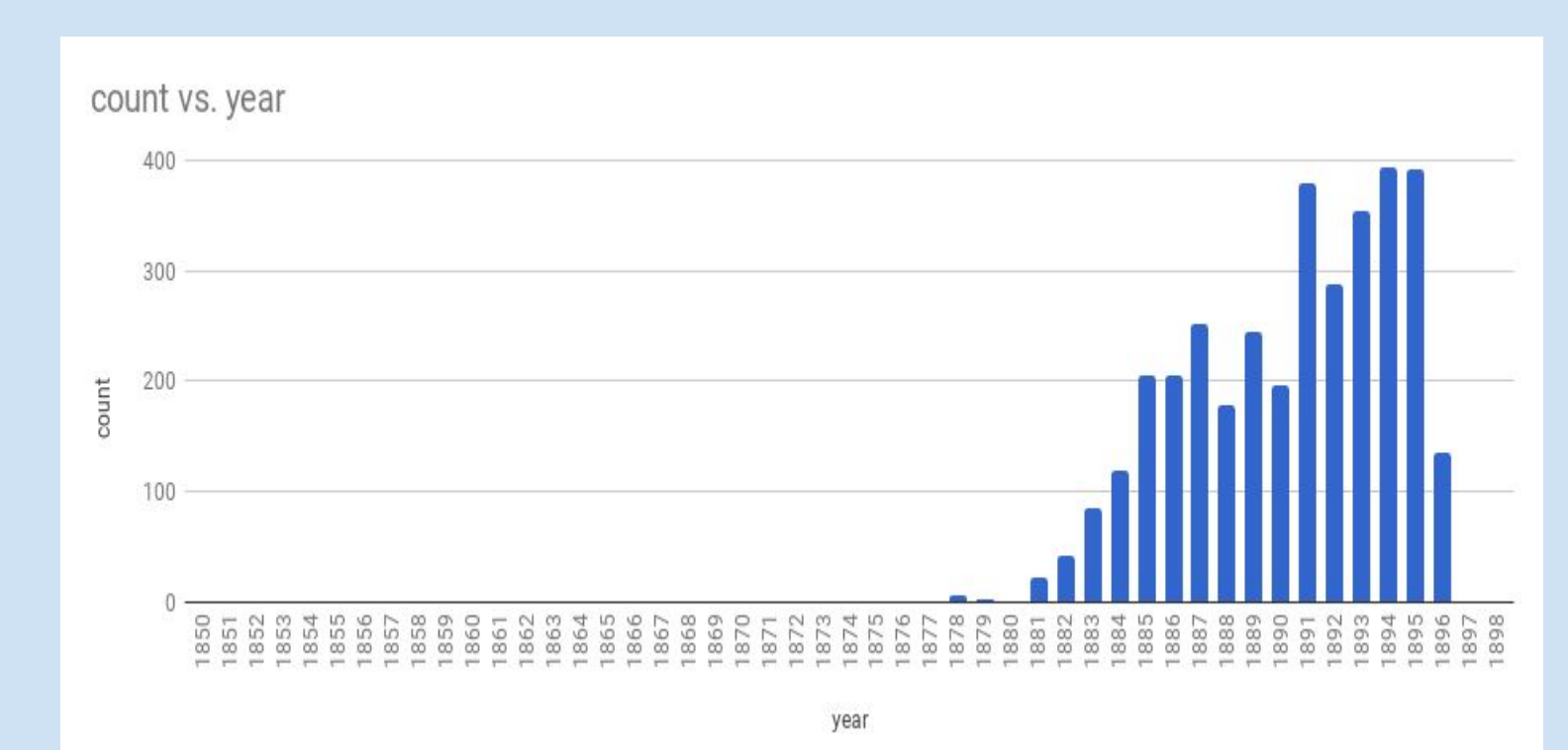
#### Visualized results



Occurrences of all the chosen 307 persons in the newspapers according to their respective ages



Occurrences of names of parliamentarians in the newspapers according to their respective ages



Occurrences of "Minna Canth" in the newspapers according to their respective ages.

### Current shortcomings

- The model may conflate two or more persons with the same name.
- Name changes (married women, ennobled persons) are not combined.
- Afterlives → fame after death?

#### Next steps

Developing a more refined way of measuring the breakthrough of persons as media personalities (e.g. frequency, geographical spread, number of links to other people and qualitative analysis.)