

Disappearing discourses:

Avoiding anachronisms and teleology with data-driven methods in studying digital newspaper collections

Elaine Zosa, Lidia Pivovarova, Simon Hengchen,
Jani Marjanen & Mikko Tolonen



The NewsEye project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 770299.



Historian's dilemma:
How to be true to the past and relevant for the present?

“... an irretrievable image of the past threatens to disappear if any present does not recognize it as meaningful to itself.”

“... es ist ein unwiederbringliches Bild der Vergangenheit, das mit jeder Gegenwart zu verschwinden droht, die sich nicht als in ihm gemeint erkannte.”

– Walter Benjamin, *Gesammelte Schriften*, I:2, p. 695. (Translation from David Armitage)

“The veto-right of the sources”

– R. Koselleck (2010)

“Seeing things their way”

– Q. Skinner (2002)

- Data-driven methods cannot solve the theoretical problem
- BUT, on some level, something that has been a question of the historian's craft and perspective, now is an issue of method.
- Data-driven text-mining methods allow – *more than ever before* – a historian to disregard the present perspective

⇒ Which discourses in fact did disappear? How can we categorize them?

Preliminary results

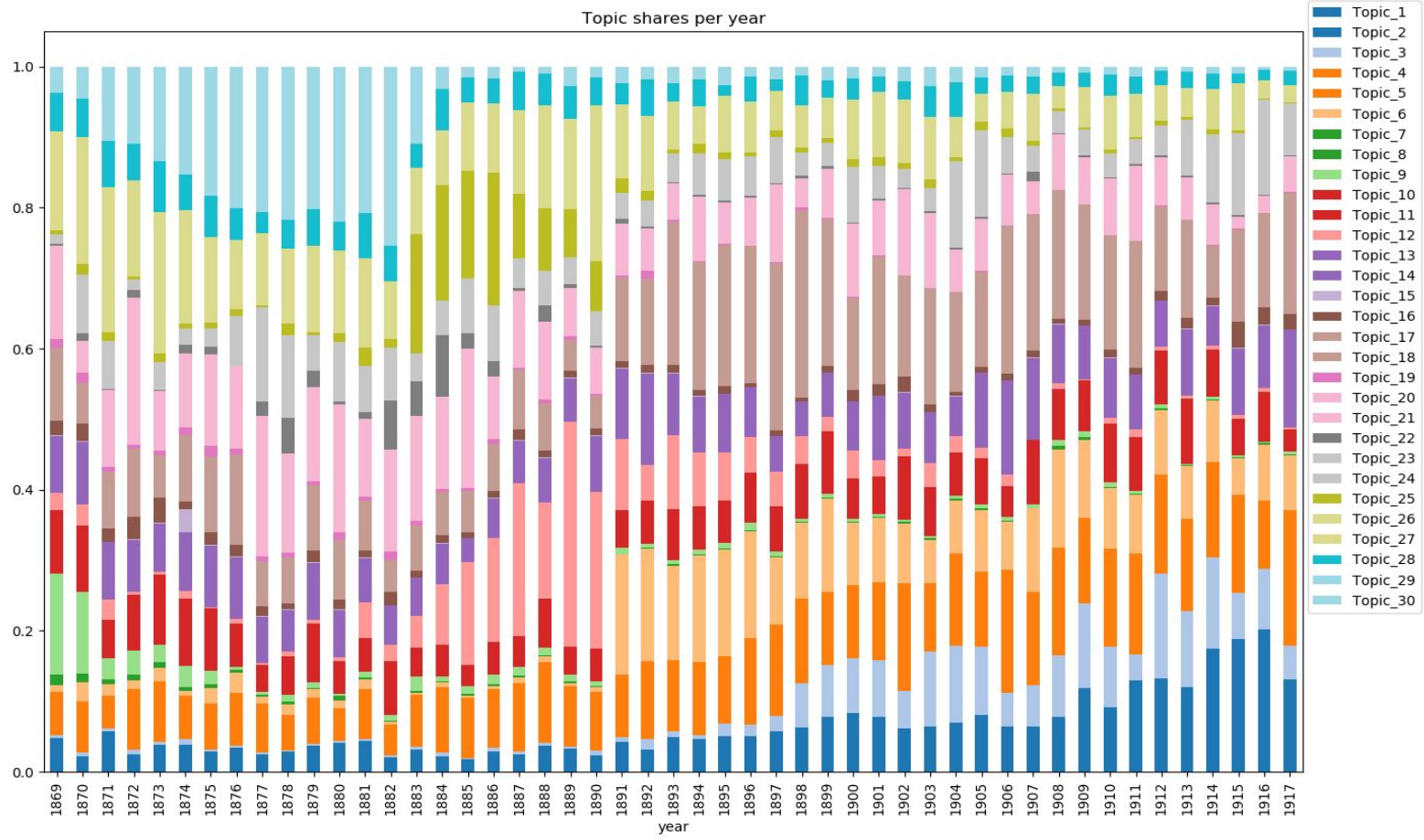


Dataset and training

- Dataset: *Uusi Suometar* 1869-1917 from the National Library of Finland
- Pre-processing: stopword removal, lemmatization
- Trained a Dynamic Topic Model with 30 topics
- Dynamic Topic Model: models topic evolution

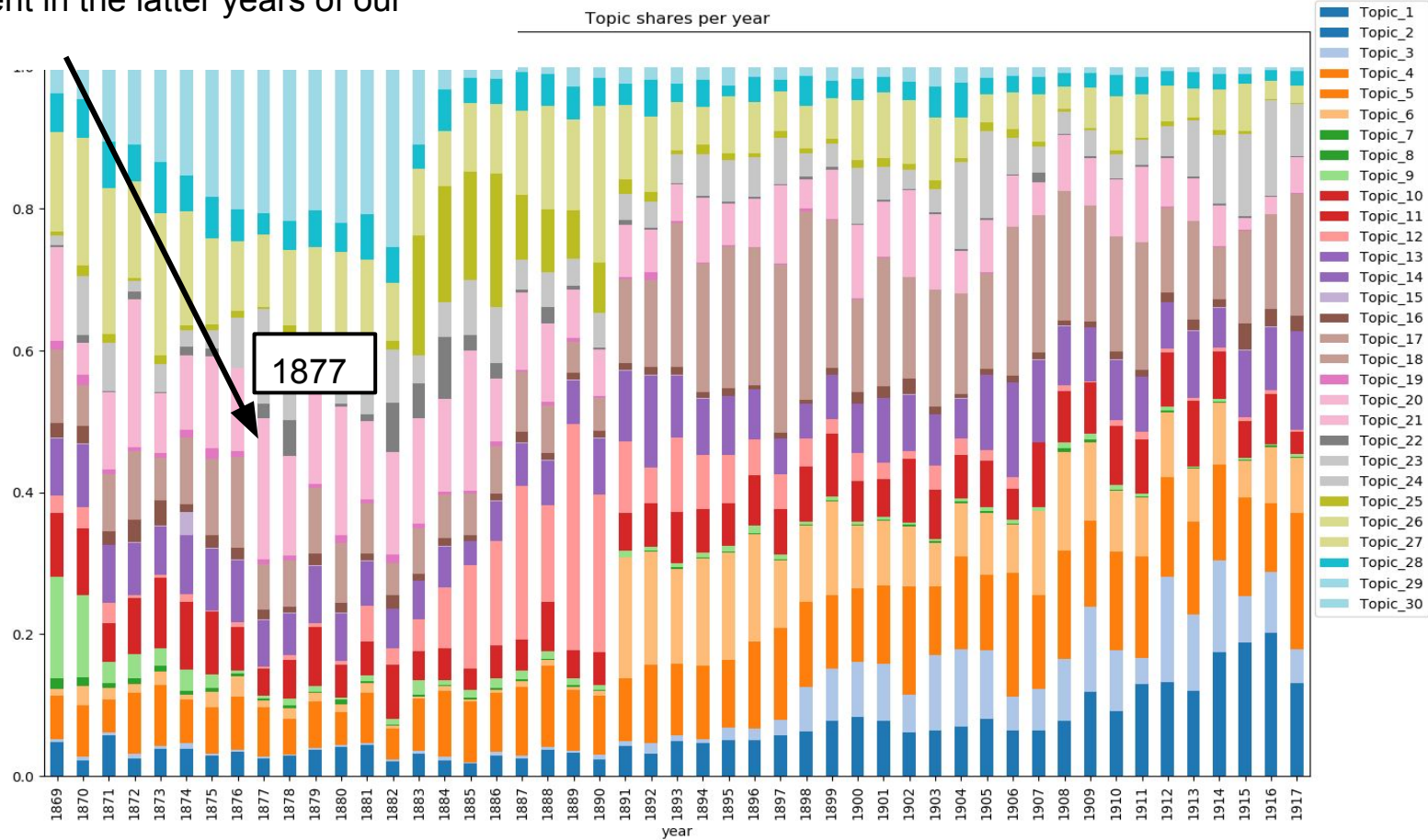


Topic shares



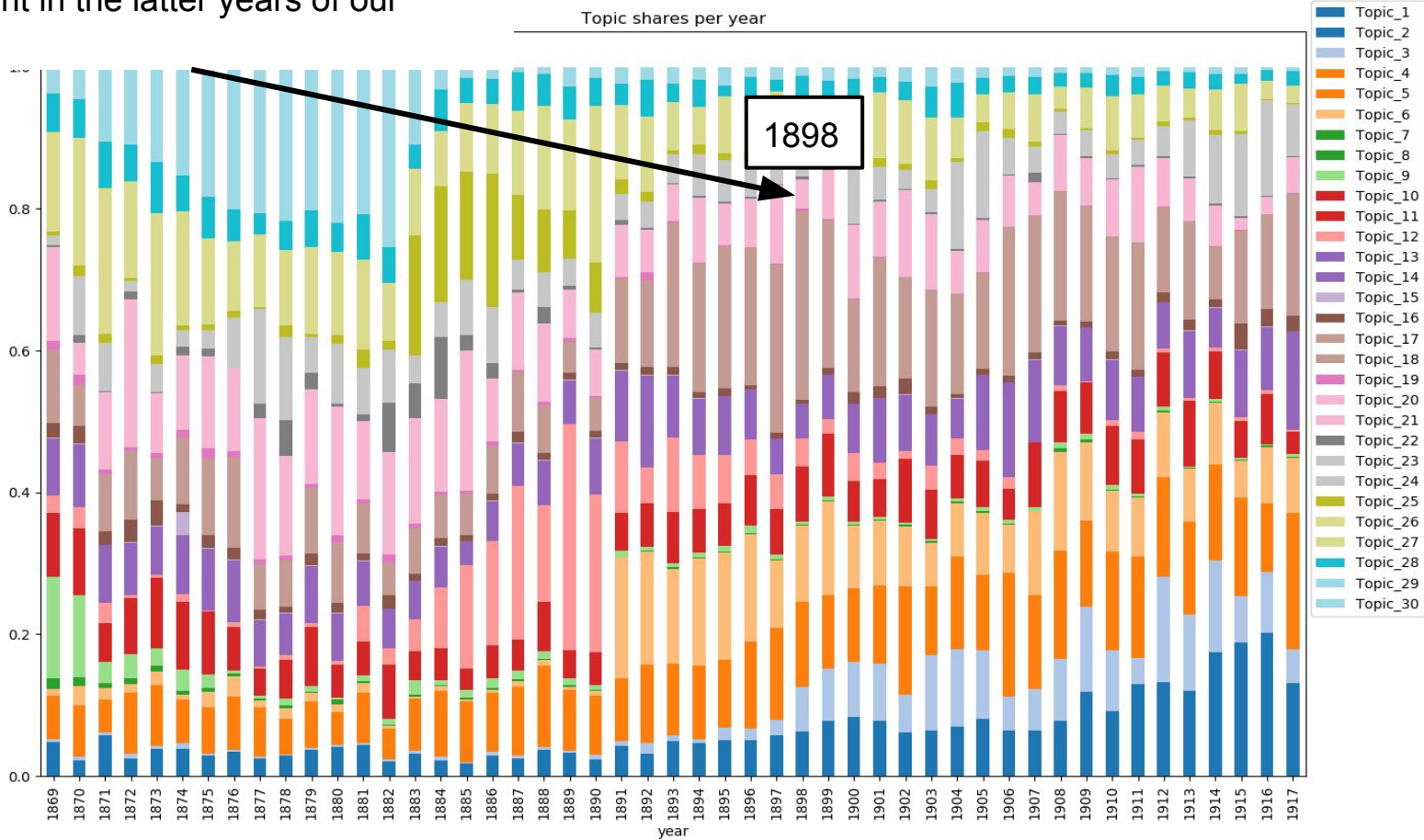
Topic 20 (in light pink) becomes less prominent in the latter years of our corpus

Topic shares

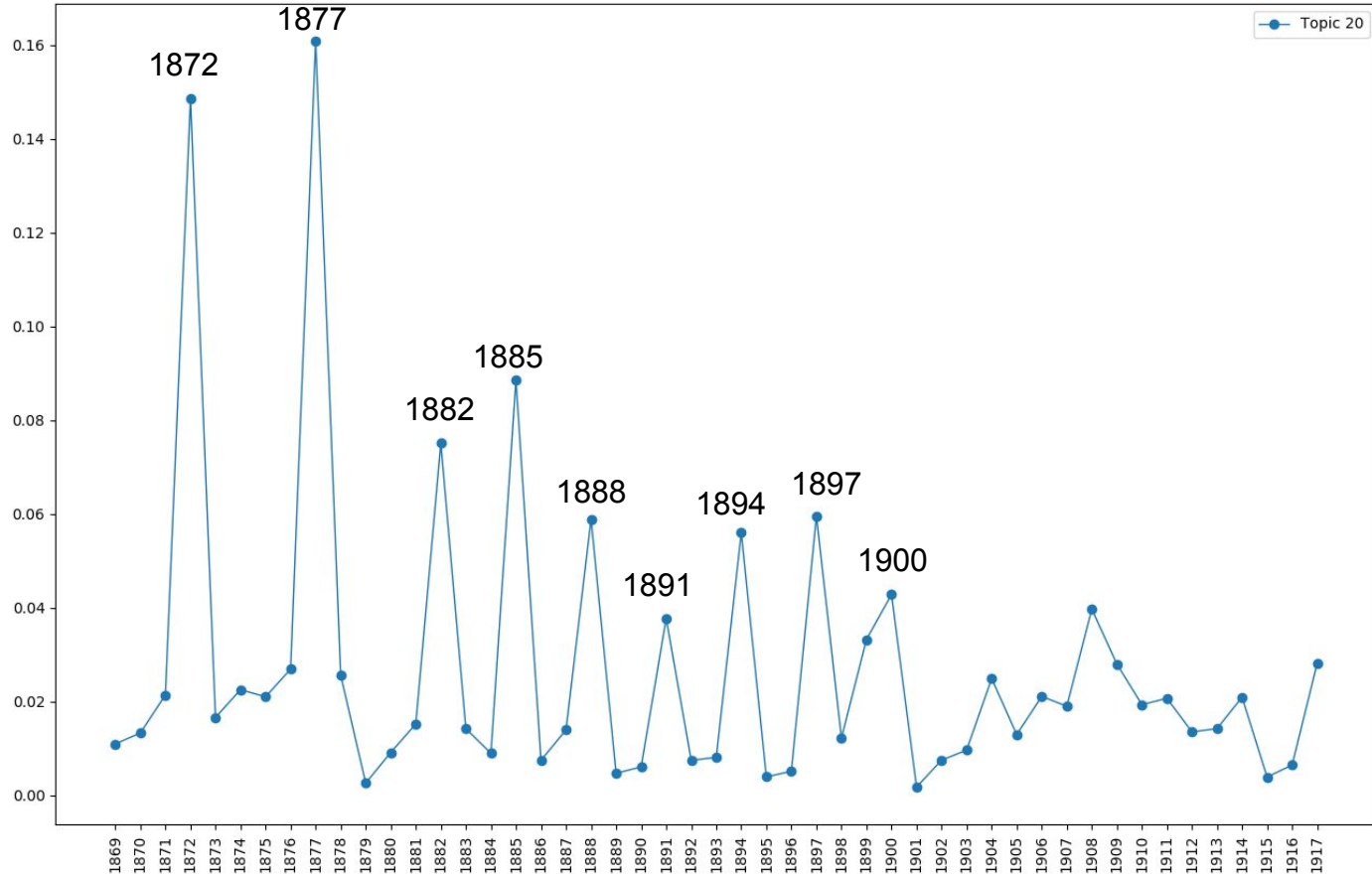


Topic 20 (in light pink) becomes less prominent in the latter years of our corpus

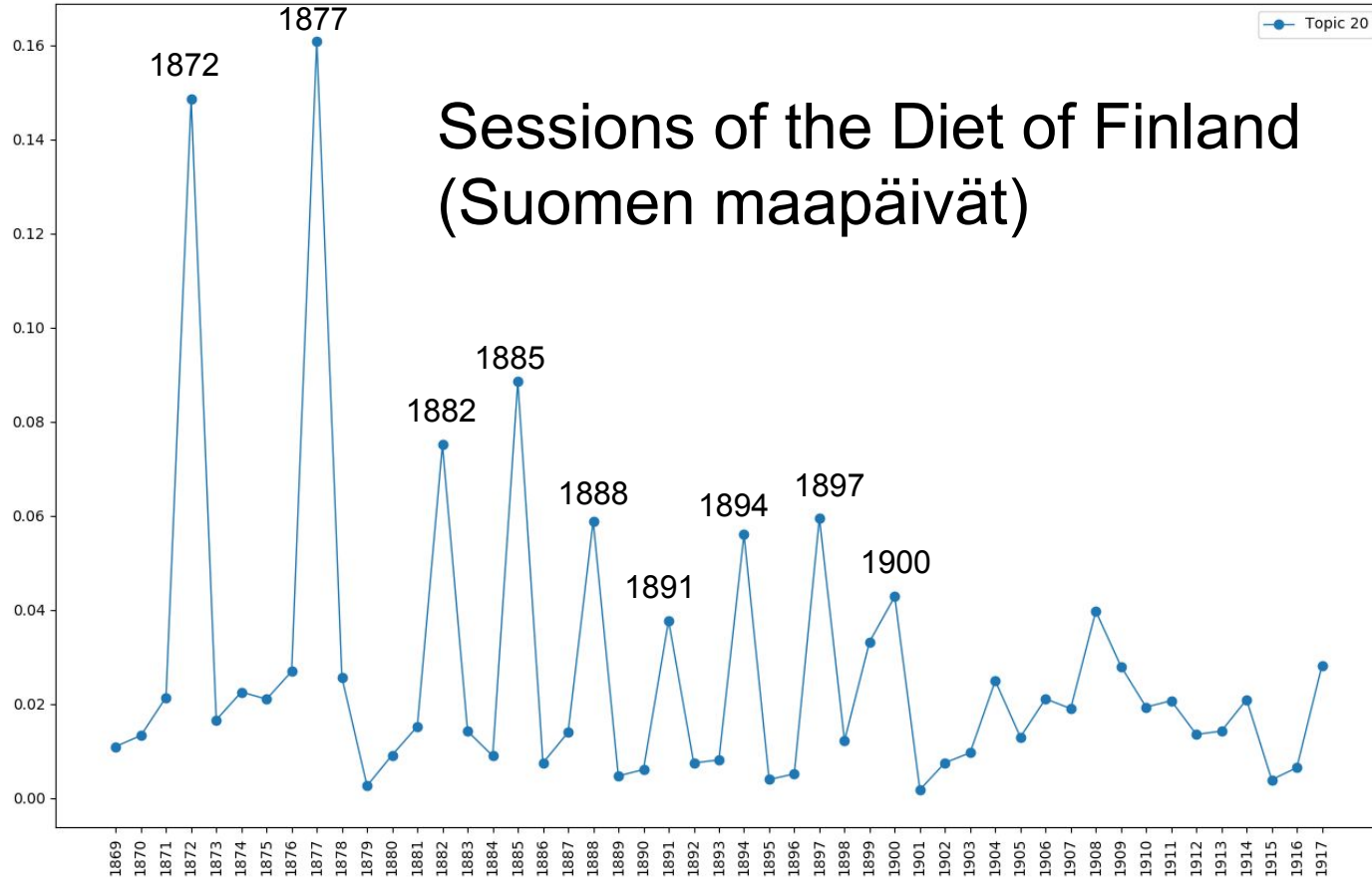
Topic shares



Topic 20



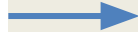
Topic 20



Topic - Diet of Finland

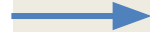
1869

asetus
suomi
oikeus
armollinen
tahtoa
julistus
pitää
sääty
puhuja
ilmoittaa



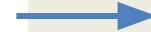
1870

oikeus
puhuja
tahtoa
asetus
suomi
pitää
armollinen
sääty
ilmoittaa
esitys



1871

sääty
oikeus
pitää
tahtoa
asetus
puhuja
esitys
suomi
armollinen
valtiopäivä



1872

sääty
mietintö
esitys
valiokunta
hyväksyä
puhuja
pitää
tahtoa
valtiopäivä
panna



Topic - Diet of Finland

1874

sääty
esitys
hyväksyä
protokolla
seurakunta
tuomari
valiokunta
päätös
ehdoitella
suomi

1875

seurakunta
esitys
edusmies
valiokunta
pappi
sääty
hyväksyä
ehoitus
kysymys
pitää

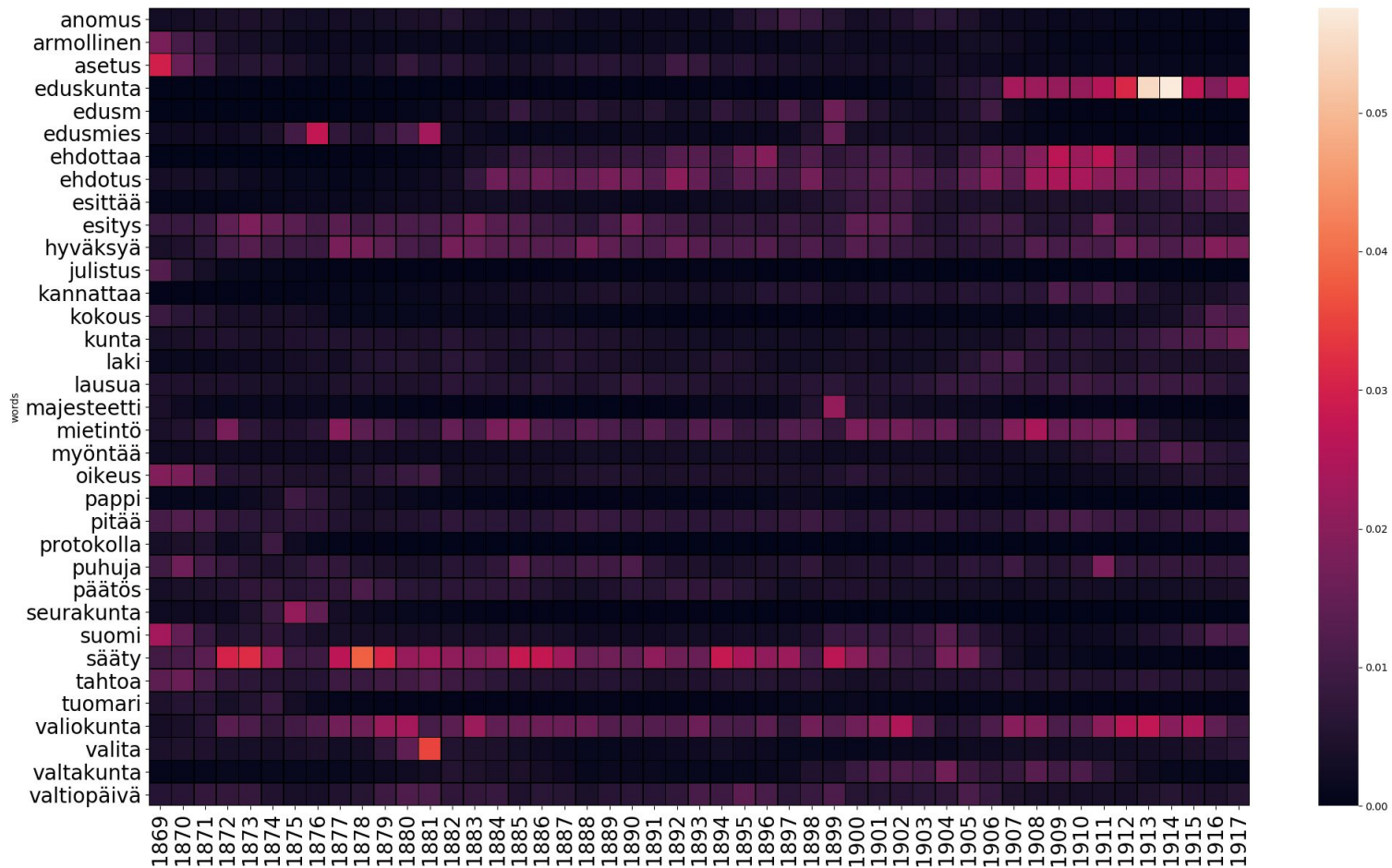
1876

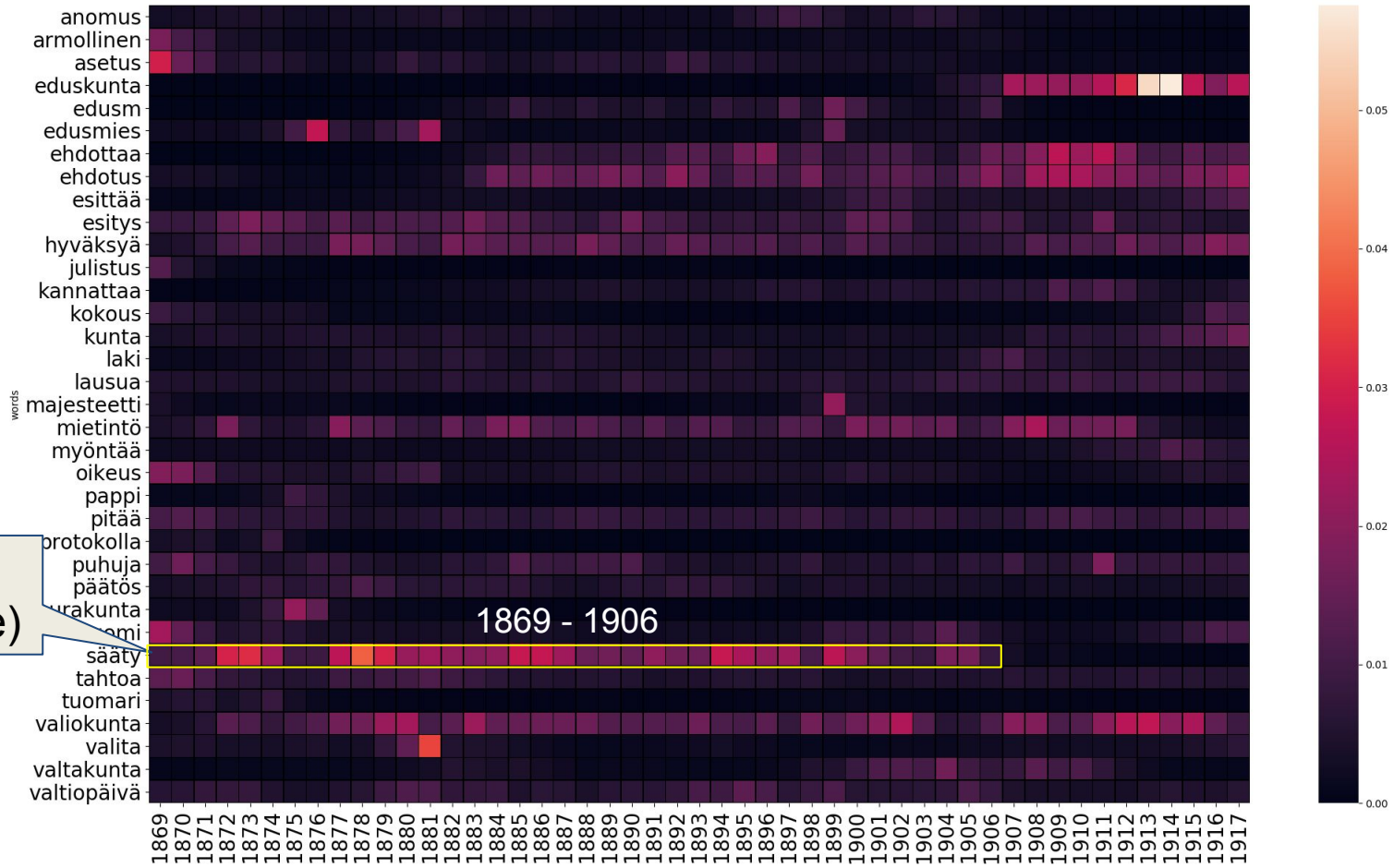
edusmies
seurakunta
valiokunta
esitys
sääty
hyväksyä
puhuja
ehoitus
muuttaa
ehdoitus

1877

sääty
mietintö
hyväksyä
valiokunta
esitys
tahtoa
päätös
puhuja
edusmies
muuttaa







säätö
(estate)

1869 - 1906



Next steps

- Train on longer time periods and include other newspapers
- Use a topic model that explicitly models time distribution of static topics (Topics over Time, Wang & McCallum, 2006)



More About NewsEye & Comhis

<https://www.newseye.eu/>

Twitter: @NewsEyeEU

<https://www.helsinki.fi/en/researchgroups/computational-history>



Comhis
Helsinki Computational History Group

The NewsEye project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 770299.

