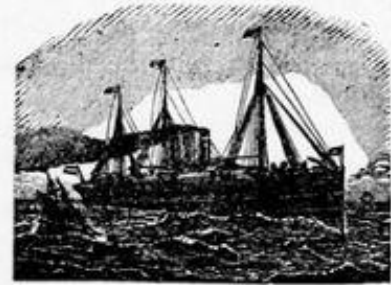


Re-visiting the emigration discourse in the Finnish newspapers in 1870–1910

20.5.2016

Amerikkaan!



Vaasasta Newyorkkiin

10–12 päivässä.
„Hampurin-Amerikan Paketti-
liike-Osakeyhtiön“

Kaksipropellisilla höyrylaivoilla kestää
Valtamerimatka ainoas-
taan 7 vuorokautta.

**Huokeat hinnat. Suotuista
matka.** Suoranaisia pilettiä hyt-
tisiijoille ja välikannelle saadaan
Jaakko Sandäsilta Vaasassa,
Joel Forsbergilta Kristiinankau-
pungissa, 217

A. N. Brenneriltä Seinäjoella.
Lähtö Helsingistä lähes joka päivä.

Migration from Finland to North America

Mass emigration from the 1870's to 1920's.

About 300 000 Finns emigrated before WW I.

Male-dominated, all groups of society but especially farmers, cottagers and workers.

Ostrobothnia region's dominance.

Reasons for emigrating e.g. political changes and oppression, economic pressure, lack of job opportunities, and hope for a better life.



source: *The Finnish Migration Collection*
Department of European and World History,
University of Turku

Research questions

How much is the emigration to North America discussed in the Finnish newspapers, 1870-1910?

1) Reality vs. newspapers

- What is the correlation between the amount of emigration and amount of articles on emigration on a given year?

2) Variation between papers

- How political affiliations of the newspapers affect the amount of articles? Also, how does the amount of articles differ regionally?

3) Advertisement

- The amount and nature of the advertisements

Earlier research on emigration discourses

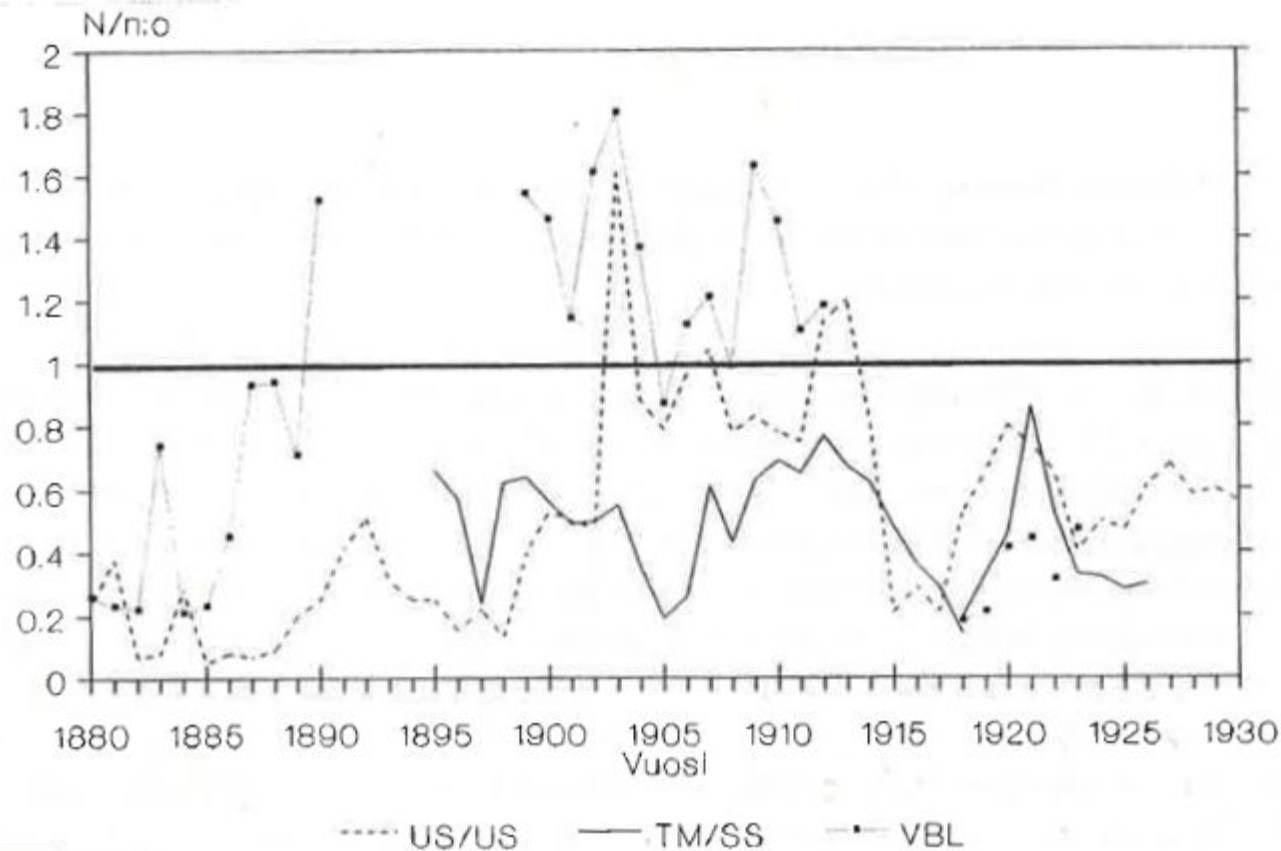
Siirtolaisuus suomalaisissa sanomalehdissä vuosina 1880-1939 ja 1945-1984.
Taisto Hujanen & Kimmo Koironen, published in 1990.

Quantitative content analysis of three different newspapers in our timeframe: *Työmies*, *Uusi Suometar* and *Vasabladet*.

Q: How much was emigration discussed in the Finnish newspapers? Did the amount of discourses correlate with the actual emigration?

A: Emigration was discussed most actively in 1903. The amount of emigration articles was almost the same in both the bourgeois and socialist newspapers in 1895-1910. The discourses generally correlated with the actual emigration.

Kuvio 7. Varsinaisten siirtolaisuusjuttujen lukumäärä suhteutettuna ilmestyneiden numeroiden määrään tutkimuksen peruslehdissä vuosina 1880–1930.



Source: Hujanen & Koiranen 1990, 44.

Research Plan

- I) Develop a method for extracting emigration related texts from newspapers.
- II) Study how the articles are distributed in the corpus according to:
 - a) time (looking for correlations with actual emigration)
 - b) political affiliations of the publishing newspaper
 - c) geography (again looking correlations with actual emigration)
- III) Study what kind of a topic emigration in newspaper media is
 - a) what else is discussed in context with emigration
 - b) what is the distribution between for example of articles and advertisements.

Data



The National Library of Finland's corpus of Finnish newspapers, 1870–1910.

Accessed through ALTO XML format raw data.

The corpus contains around 3 billion words.

Methodology

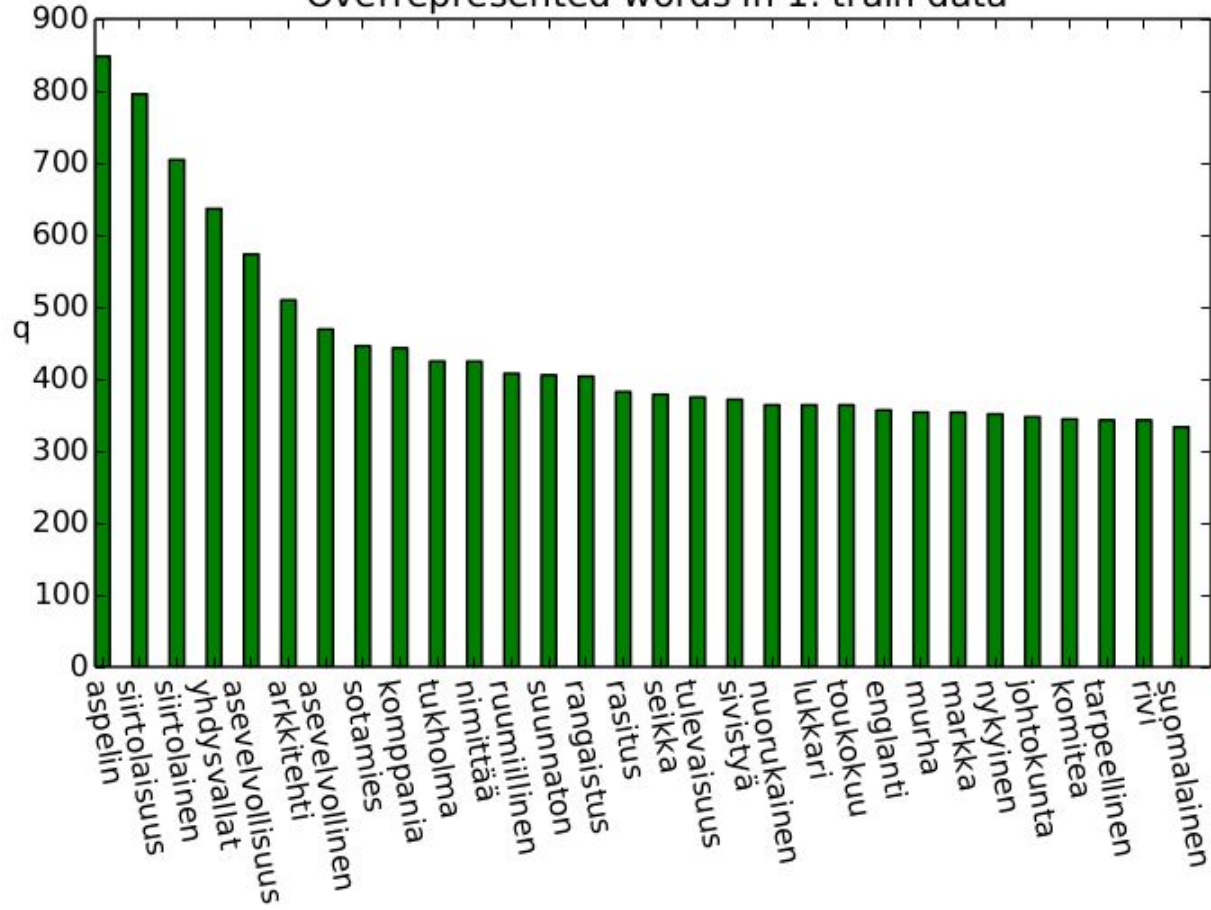
The first step of the methodology was to extract newspaper articles that talk about emigration to North America.

For this purpose, a training data of emigration related articles was manually collected from peak years of emigration (1887 and 1902).

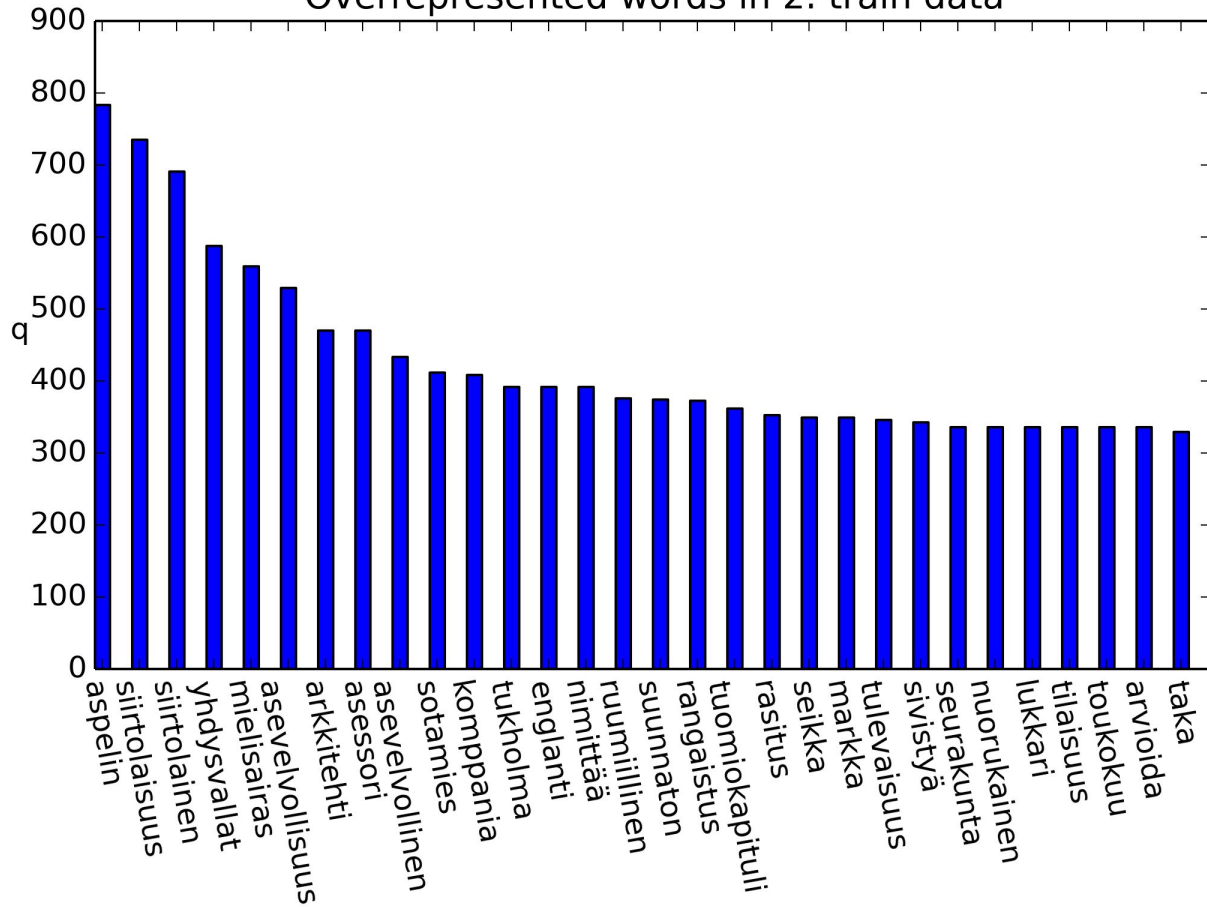
Word frequencies from this training data was compared to reference frequencies obtained from a random sample of all articles from the same period.

Those words that showed overrepresentation in the data were interpreted to be relevant for emigration discourse.

Overrepresented words in 1. train data



Overrepresented words in 2. train data



Methodology

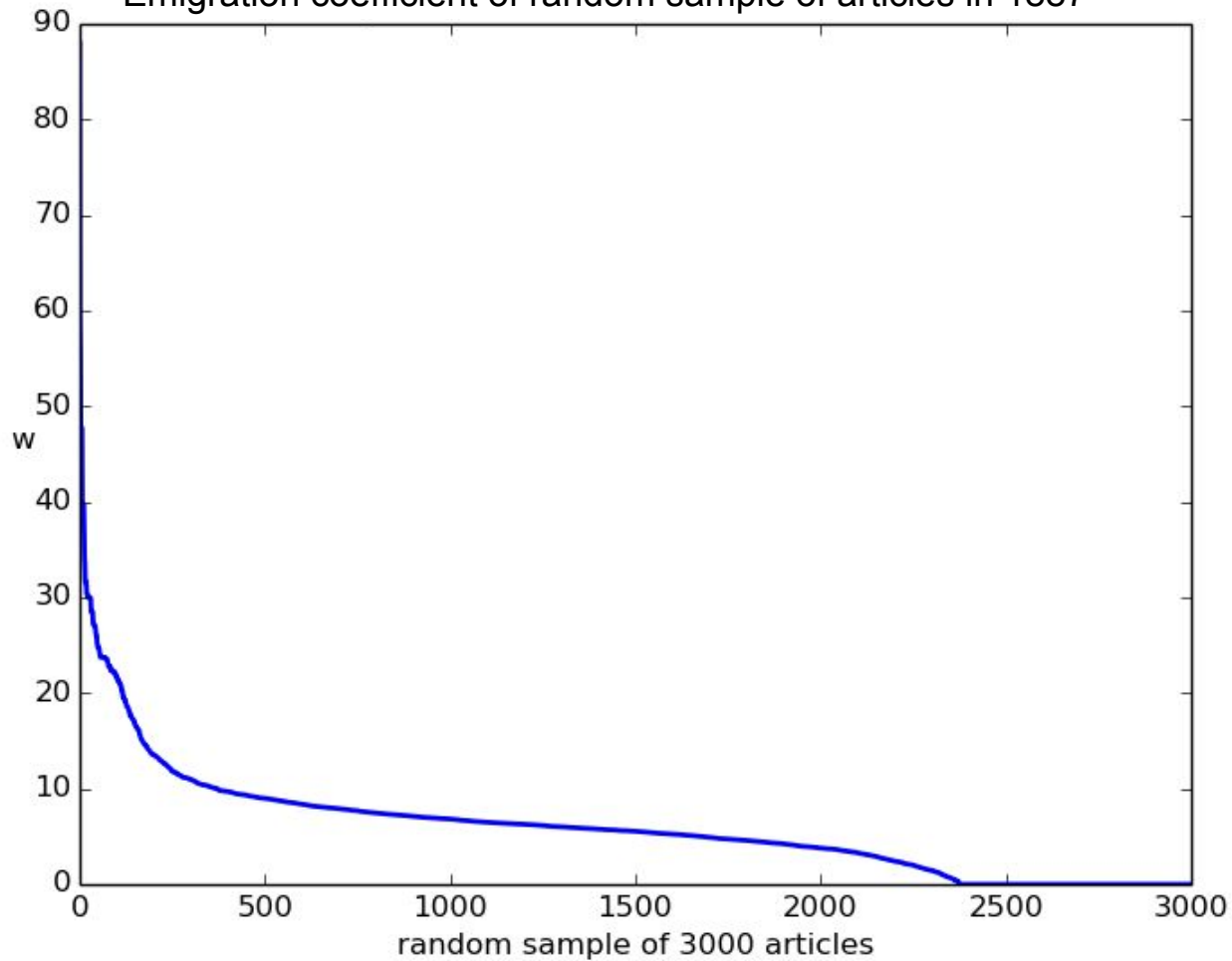
Next step:

Relevance of emigration to any article's content can now be estimated as the mean of (over) representativeness of its words in the training data.

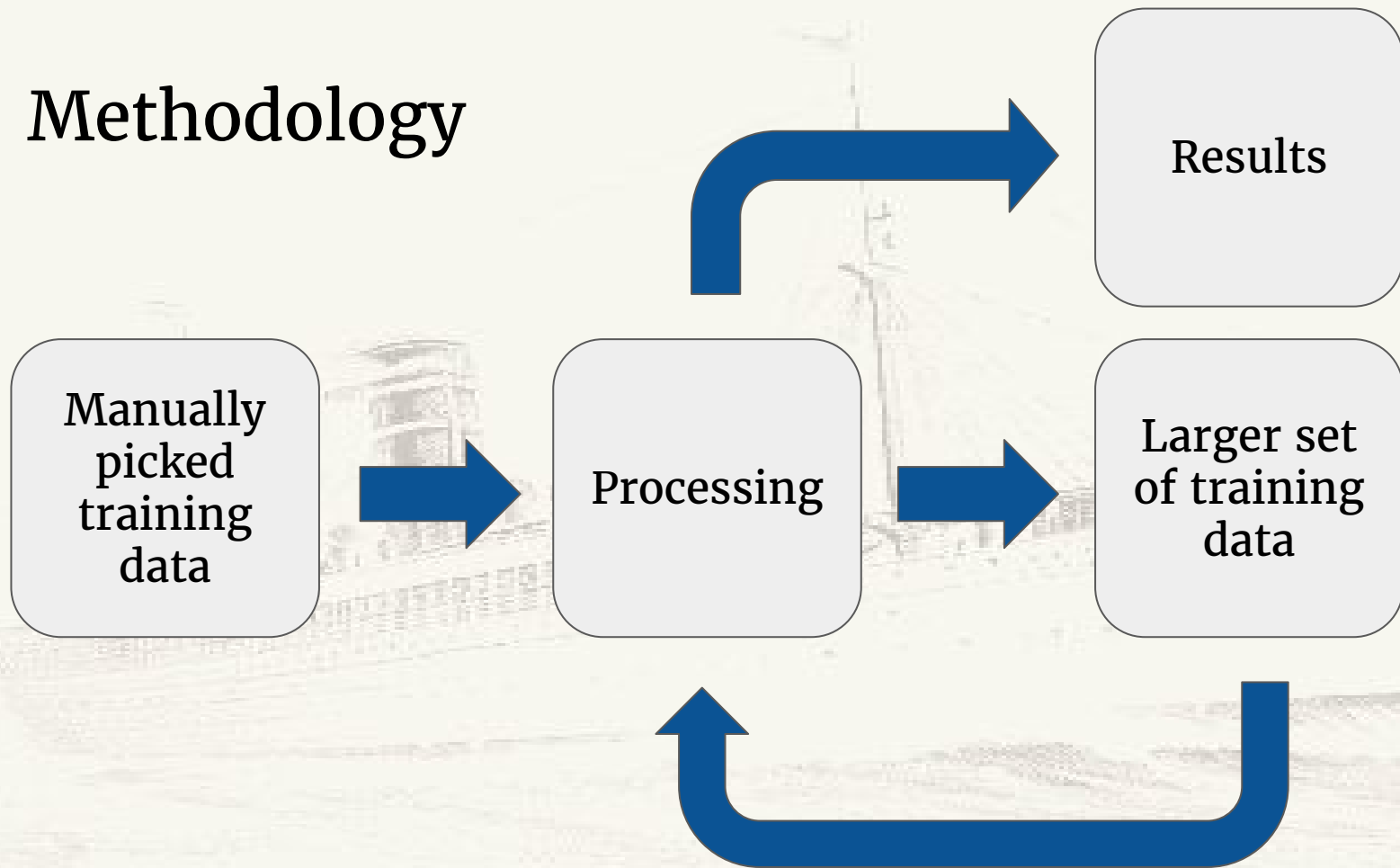
This measure of relevance can be used to extract candidate articles, which in turn can be manually evaluated to improve the training data.

As the end product we (hopefully) will get a decent measure of emigrations relevance to (any) article's content.

Emigration coefficient of random sample of articles in 1887



Methodology



Methodology

Experiences

The method seems plausible and preliminary results promising.

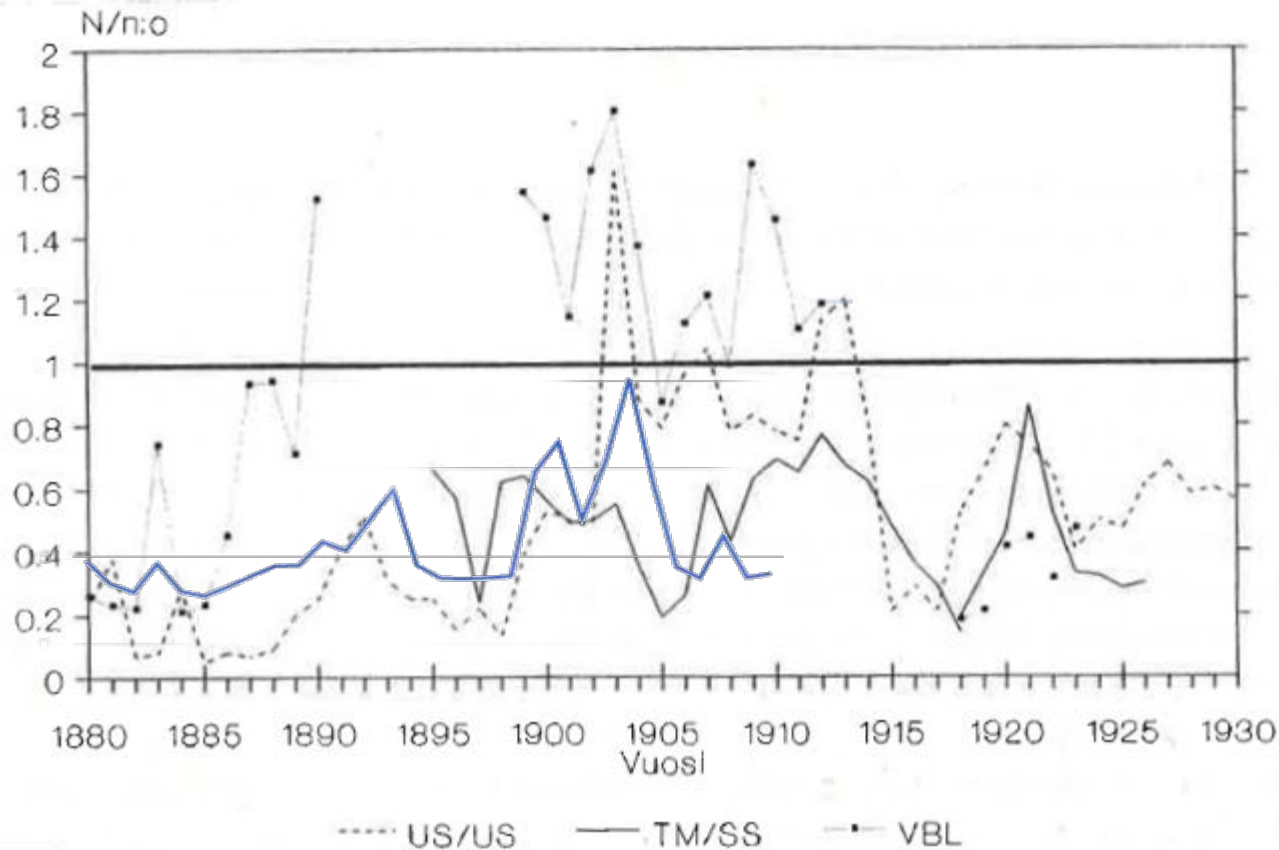
However, programming the pipeline turned out to be slower than expected, while human resources were abundant.

Problem:

Distribution of work did not take into account the whole process from the start, but proceeded from beginning to end.

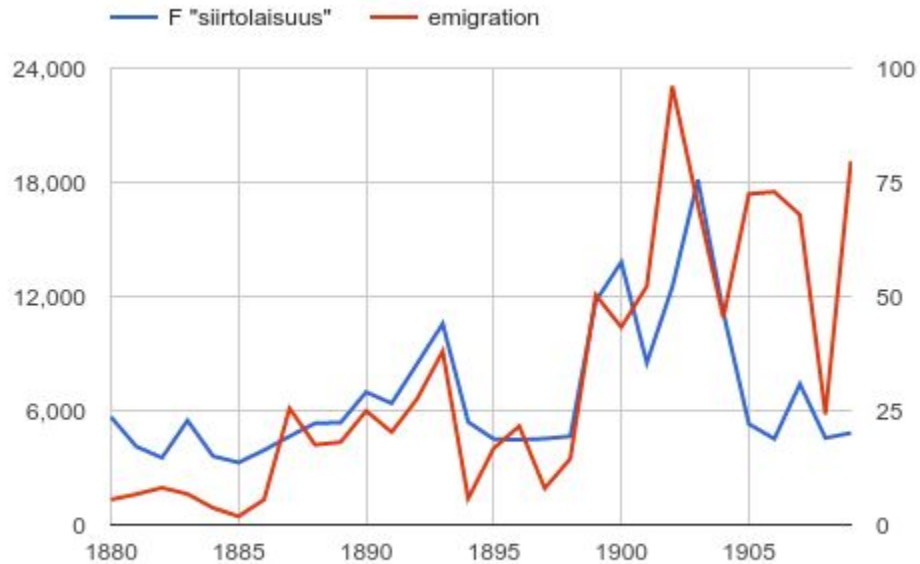
In order to avoid bottlenecks, the workflow should have been planned and explicated in a more detailed fashion.

Kuvio 7. Varsinaisten siirtolaisuusjuttujen lukumäärä suhteutettuna ilmestyneiden numeroiden määrään tutkimuksen peruslehdissä vuosina 1880–1930.



Source: Hujanen & Koiranen 1990, 44.

Reality vs. Newspapers



Advertising

Expectations:

- Steady and substantial amount of advertising
- Advertised trips are an established product

Increased amount of advertising, especially from the peak years of emigration

Siirtolaisille vaariin otettavaa.

Amerikkaan

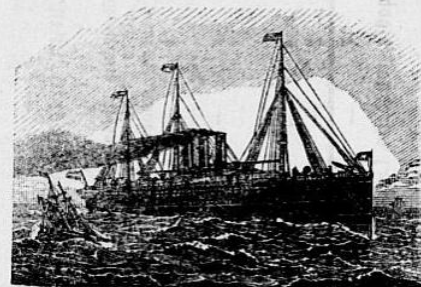
kuljetetaan siirtolaisia nopeimmin ja mukavimmin höyrylaivosakeyhtiö Nordin ensiluokkaisilla ja välikannen matkustajia (siirtolaisia) varten varta vasten sisustetuilla höyrylaivoilla

Advertising

Qualitative analysis of a small random sample
(3 newspapers)

- Not much advertising of cross-Atlantic trips
- Advertising is a complex phenomenon, including varying strategies, rhetorics and conventions

AMERIKAAN!



Turusta Newyorkiin

10—12 päivässä.

**Hampurilais-Amerikalaisen
Postikuljetus-Osakeyhtiön**
Kaksois-propellisella nopeasti kul-
kevalla höyrylaivalla.

Valtamerellä kulku vaan 7 päiv.

Huokeat hinnat. Mukava matkustus.

Suoranaisia kajutta-paikkoja ja
välikansi-pilettiä on saatavana

Holmudd & Hellbusch'illa

Turussa, sekä

P. von Knorring'illa

Maarianhaminassa

Lähtö Helsingistä melkein joka päivä.

Advertising

Advertising often dialogical:

For example, ticket agents commenting on other shipping line's quality and reliability

“Rumour control”, commenting on information from informal sources:

- third party travel agencies' policies
- speculation of coming changes in American immigration policies
- possibly fabricated eyewitness recounts implemented in advertisements

Amerikaan!



Helsingistä Newyorkiin
10-12 päivässä

Hamburgin—Amerikan
Höyrykukku-Osake-Yhtiön
kaksipropellisilla pikahöyryillä.

Valtamerellä ainoastaan 7 päiv.

Halvat hinnat. Mukava matka.

Koko matkalle piljettejä kajutta-
paikkoihin ja välikannelle myy Hel-
singissä

Otto Rohde.

Helsingin satama on auki;
tie Pohjanmaalta Helsinkiin on 90
virstaa lyhempi ja matkalla menee
1 päivä vähemmän kuin matkalla
Hankoniemeen.

Matkustavaisten tulee saapua Hel-
sinkiin Torstai-iltana.

Further research

Final product (in terms of original research questions):

- 1) Distribution of emigration related articles in terms of time, geography and political affiliations of the newspapers.
- 2) A new corpus of emigration related discourse for:
 - a) Qualitative research
 - b) Text mining of concurrent features & variation of content

**Wieläfin
siirtolaisuudesta.**

Further other work

Side product:

The pipeline in itself is reproducible and could function as a foundation for a simple text corpus search interface.



**Wieläfin
fiirtolaisundeSta.**

Thank you!

Satu Bennert
Antti Kanner
Johanna Komppa
Aaro Salosensaari
Ilari Sarén
Ville Vaara

University of Helsinki

Ilavarasi Radhakrishnan
Aalto University

Risto Turunen
University of Tampere

Taina Saarenpää
University of Turku, MAMK



UNIVERSITY OF HELSINKI



UNIVERSITY
OF TAMPERE



Turun yliopisto
University of Turku