

Intentional language change: written Estonian ca 1900

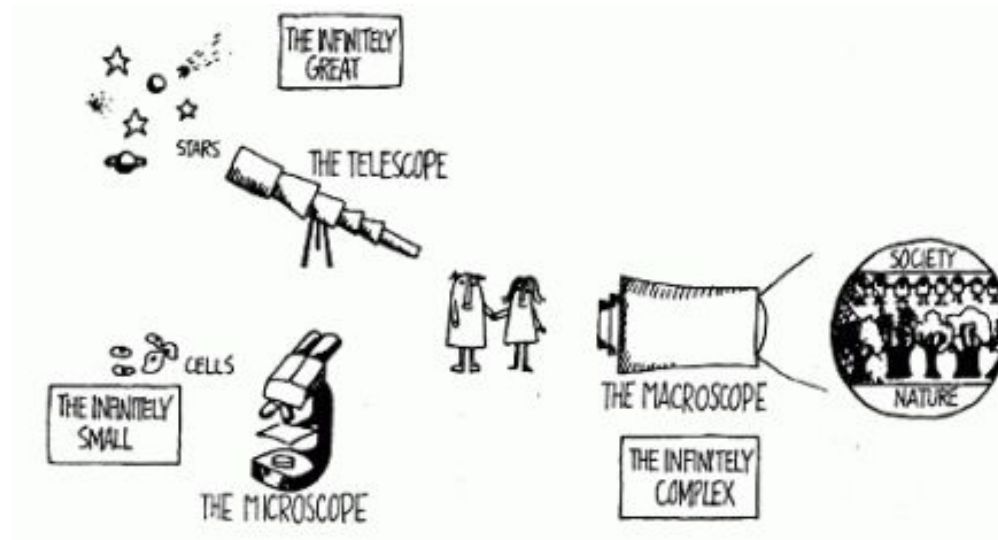
Peeter Tinitis

Congress “Juri Lotman’s Semiosphere” 2022

Tartu, 27.02.2022

Computational approaches in humanities

Looking at the connections, differently?



de Rosnay 1979: The macroscope is unlike other tools. It is a symbolic instrument made of a number of methods and techniques borrowed from very different disciplines. It would be useless to search for it in laboratories and research centers, yet countless people use it today in the most varied fields.

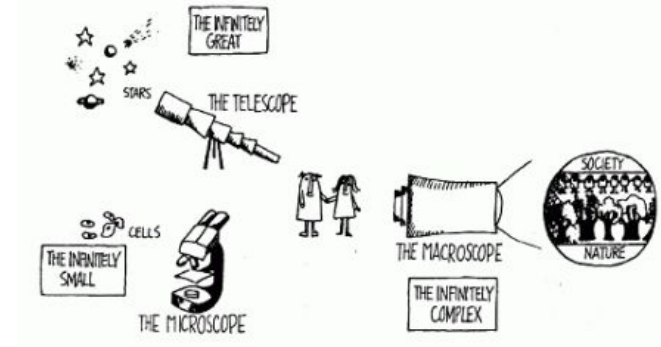
(de Rosnay 1979, Graham et al. 2015)

Computational approaches in humanities

Quantification in the humanities

Clarity, testing, big picture

Naturally, theory & thinking matters still



Here: historical sociolinguistics - find out what happened (*how past really was*).

- We don't really know, can't ask
- Need to put the puzzle together from incomplete data

Language

Language is ... (here, today :))

- Communicative habits in a community
- The speech acts they have been able to observe, and what they can generate from it.

In principle, it can all be quantified.

- What sentences have been observed,
- What biases and learning in the brain,
- How many sentences per day or per lifespan



Language and social value

Language has social value:

- Standard vs dialect
- High variety vs low variety
- L1 speaker vs L2 speaker
- Us vs them

Status has implications for use and existence. Which varieties do you learn and use?

Ideologies and discussions go into this. A lot of it is implicit and unnoticed.

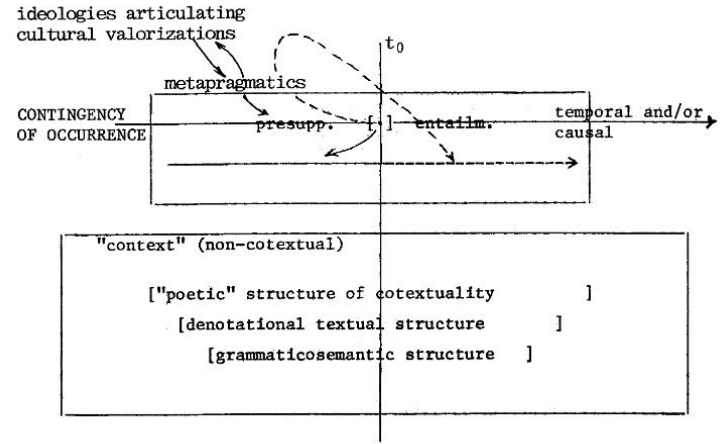


Fig. 1. Micro-contextual semiotic of indexicality.

Table 2
Accent speech levels in Britain

Highest ranked	
↑	Mainstream RP ('unmarked' RP)
	Near-RP: educated English accents of Scotland, Wales & Ireland
	Provincial accents (Yorkshire, West Country, Tyneside Geordie)
↓	Urban accents: London (Cockney), Liverpool (Scouse), Glaswegian
Lowest ranked	

Source: Giles, 1970, 1971.

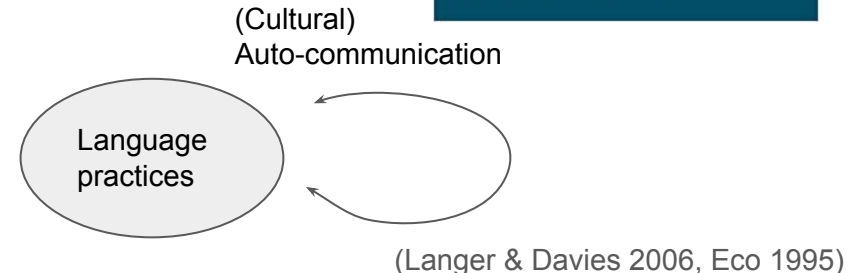
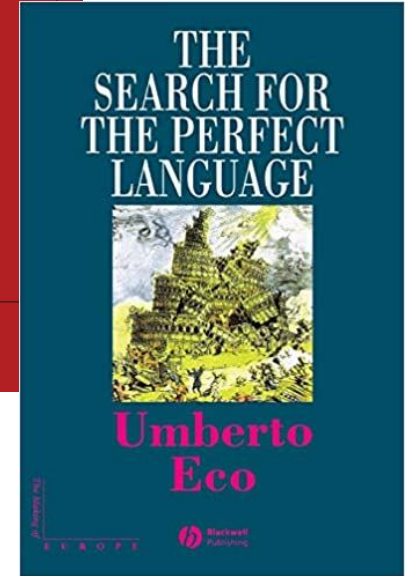
(Silverstein 2003, Auer 2005)

Language development projects

A long history of trying to control the language, make it fit an ideal

- Language standards: central, valued language forms, imposing uniformity on diversity
- Some forms or expressions come to be seen as “bad language”, “bad Estonian”, “bad German”, “I know X is bad language, but I still use it.”
- Can go quite deep: “How could Estonians be virtuous when they don’t even have a word for virtue” (Arvelius 1792), Cf. Estonian future tense
- Or neologisms of Johannes Aavik et al. in trying to beautify the language while they can.

Often significant effort is made trying to alter language practices to correspond to some ideal. This could be seen as an instance of cultural auto-communication.



Discourse entering into practice?

What happens when correct forms are taught in school?

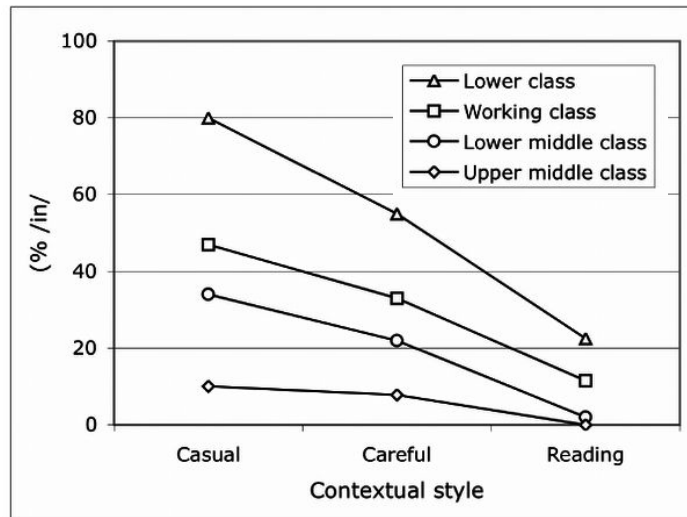
- Yes, use will follow this, but not always and not fully.

Sociolinguistic studies show how usage may vary depending on expected context.

- E.g. Casual, Careful, Reading

Ideals can be taught, but people use them as they want.

We want to understand how these attempts can reach a wider population. Generally, the role of intentional changes is considered rare for language history, but less rare in writing.



She is walking. Walking is fun. (Labov 1989)

(Labov 1966, Milroy 2001)

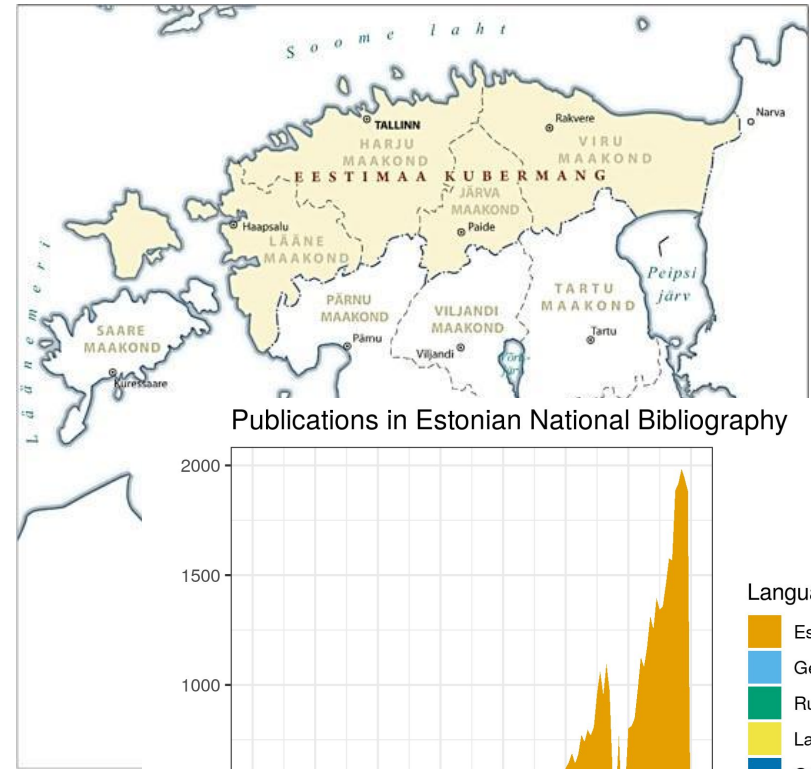
Background on Estonian

Around 1800 in the Russian Empire:

- ~400,000 speakers
- Rural, low prestige, oral language
- Written done by outsiders

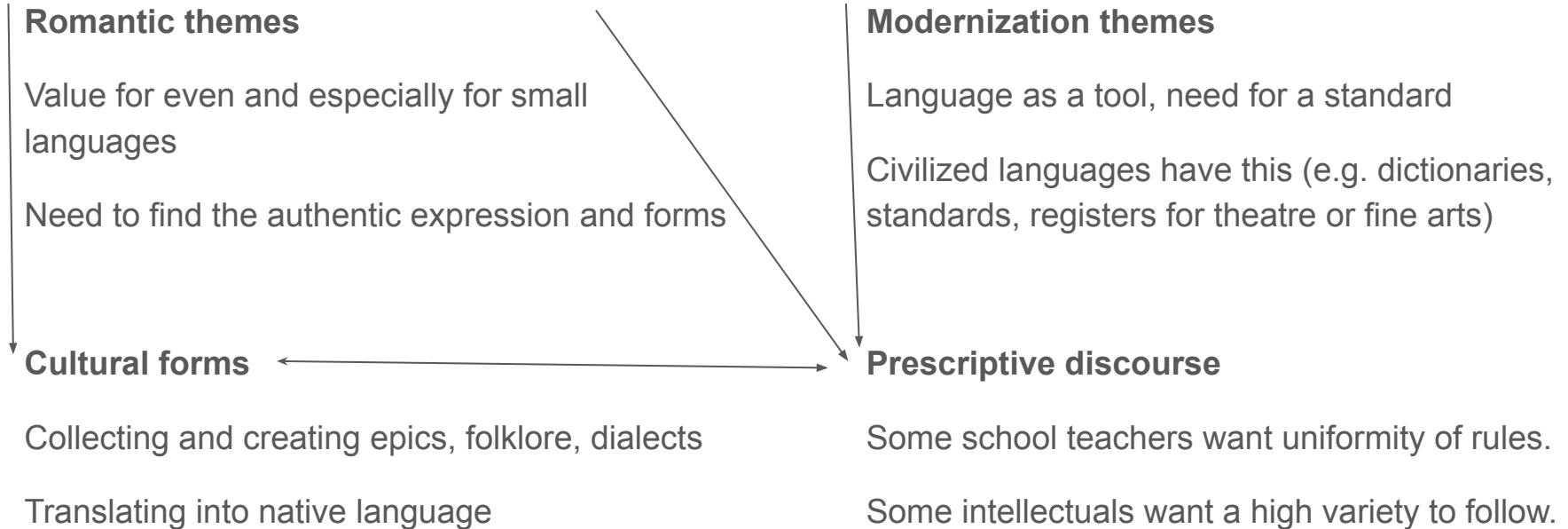
Late 19th century awakening

- Growth in literacy (20% -> 77%)
- Newspapers, printed works
- Moving to cities
- Social status of Estonian
- Cultural status of Estonian



(Raag 2008)

Language interest during the 19th century



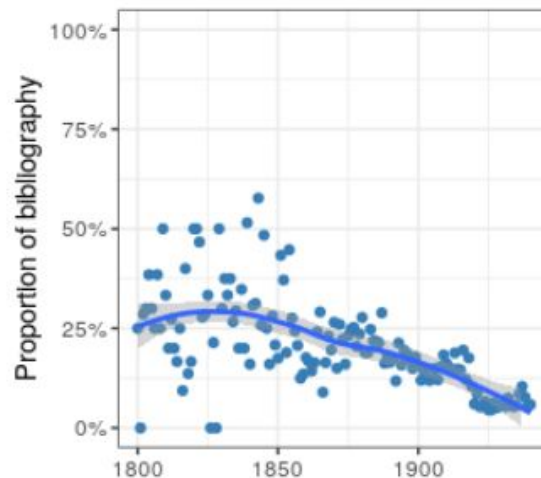
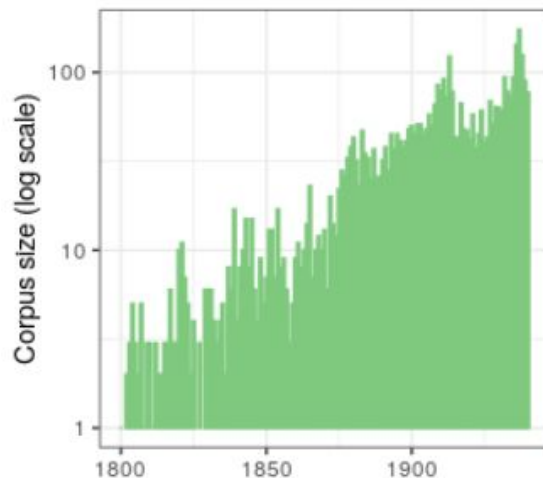
Did the prescriptive
influence carry over to
language use? How?

Data sources

Large collection of texts 1800-1940 (4608)

Metadata

- Biographic data (1k)
- Bibliographic data (39k)
- Demographic data
- Dialect data

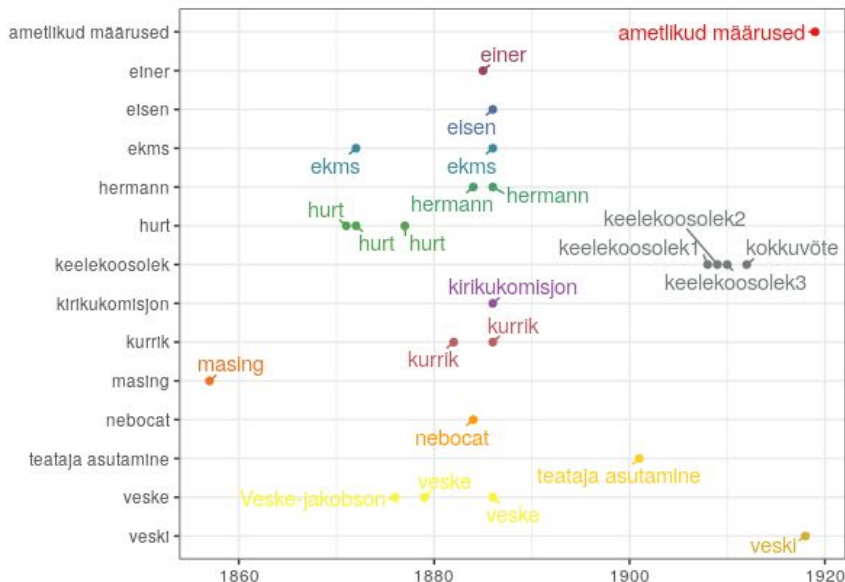


Data sources

Prescriptive works (25 here)

- Dictionaries & committees
- Language conferences
- Newspaper articles

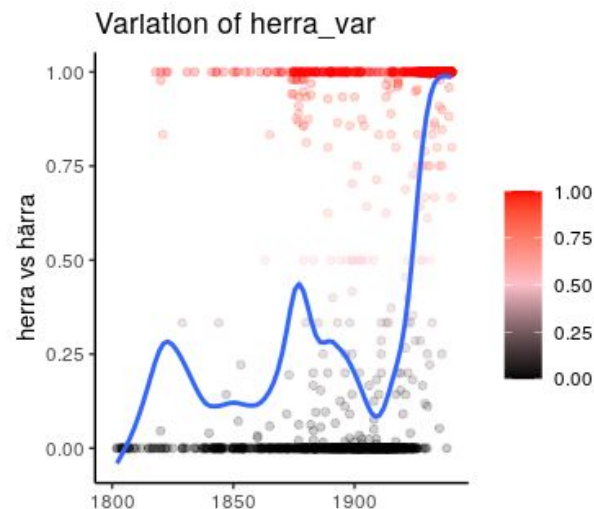
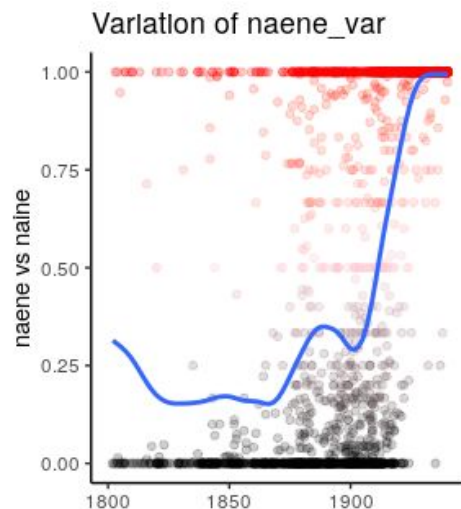
Example: *Please use “hää” instead of “hea” as this is the proper way to write.*



Dealing with texts: Linguistic variables (form vs form)

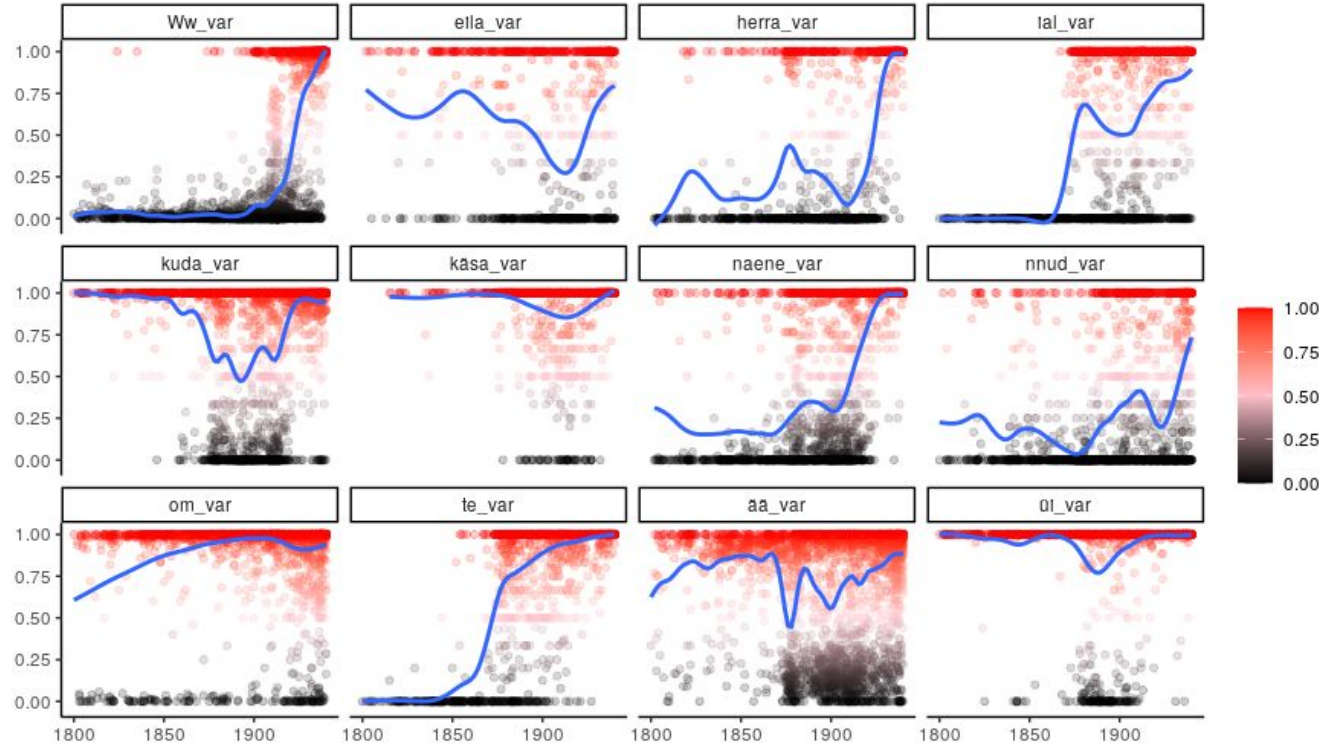
Linguistic variable:

- Competing forms that mean mostly the same thing (Labov)
- naene - naine
- herra - härra
- hää/pää - hea/pea



Many Linguistic variables

Some variables that showed interesting variation around ca 1900



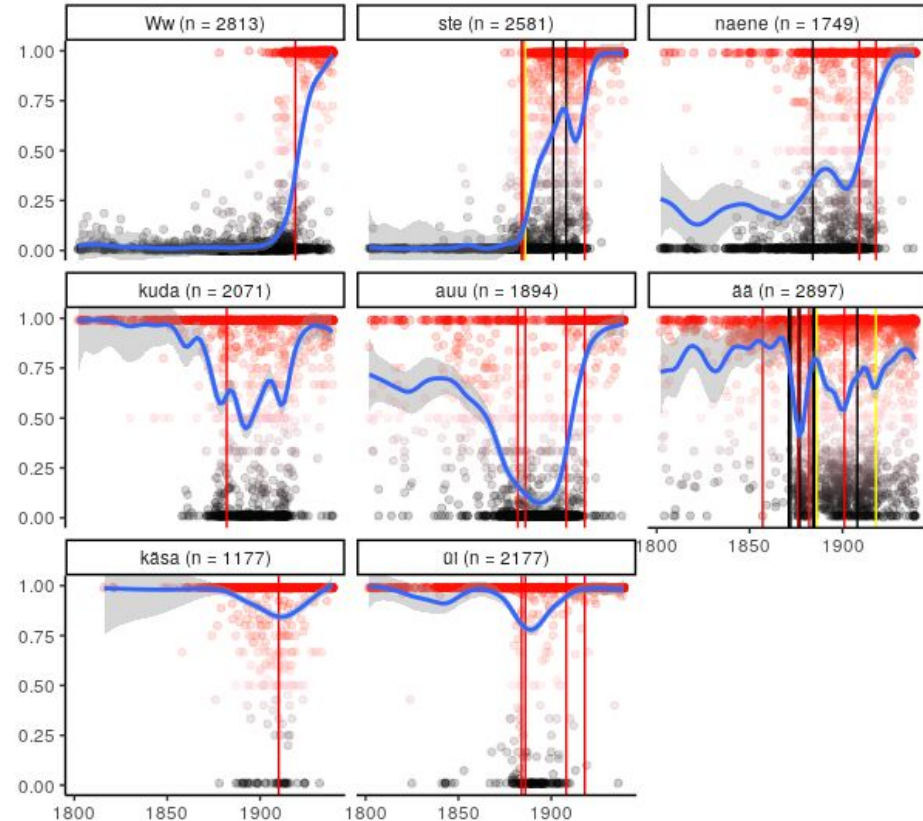
Impact of prescription

Combine data on variation and prescription.

Did prescription influence language use?

Here, population mean based on model fit vs timing of prescriptive effort.

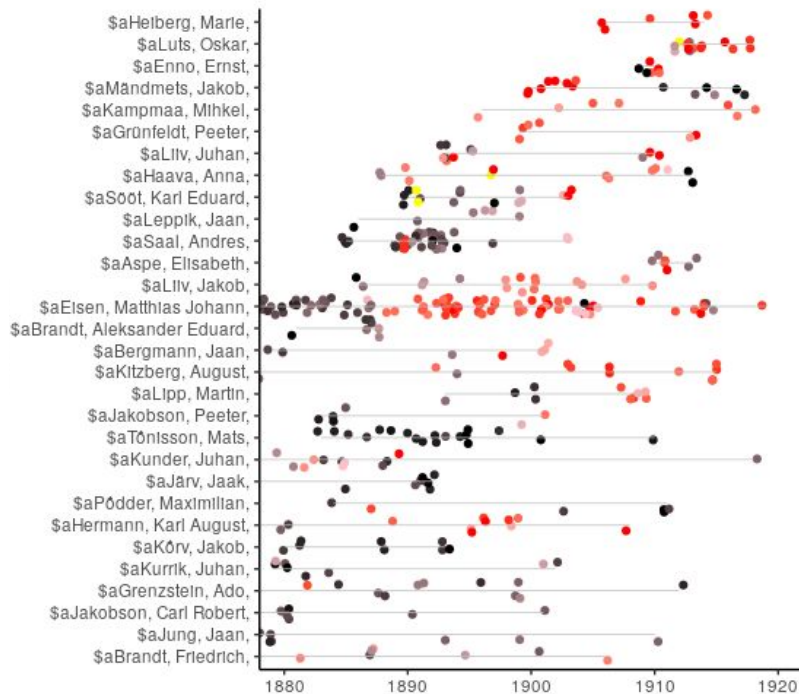
Timing co-occurs often with trend changes in population. Yes, it did play a role.



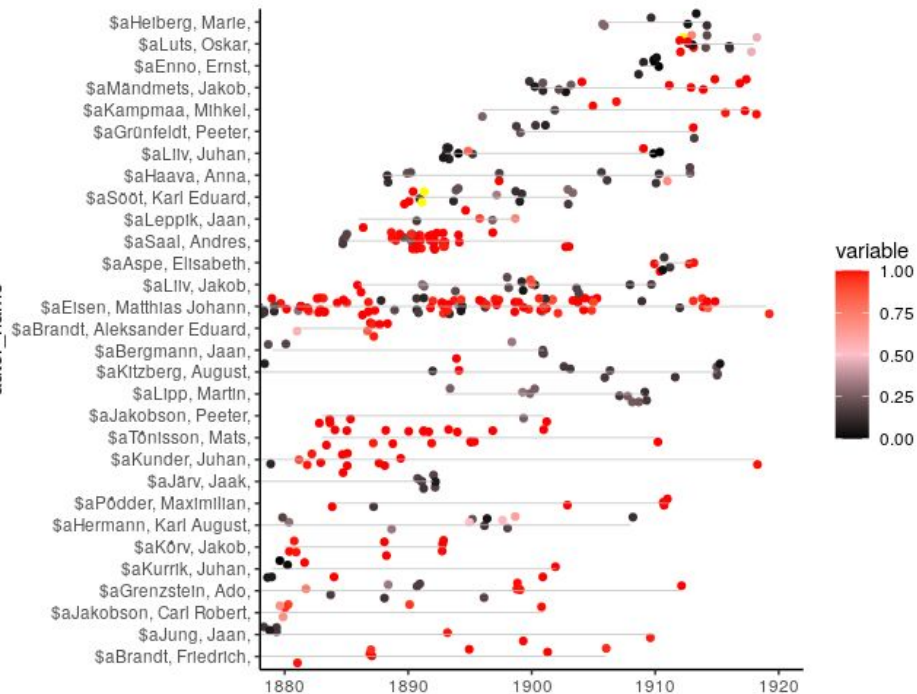
The mechanisms

Author level variation

Variation of ste_var

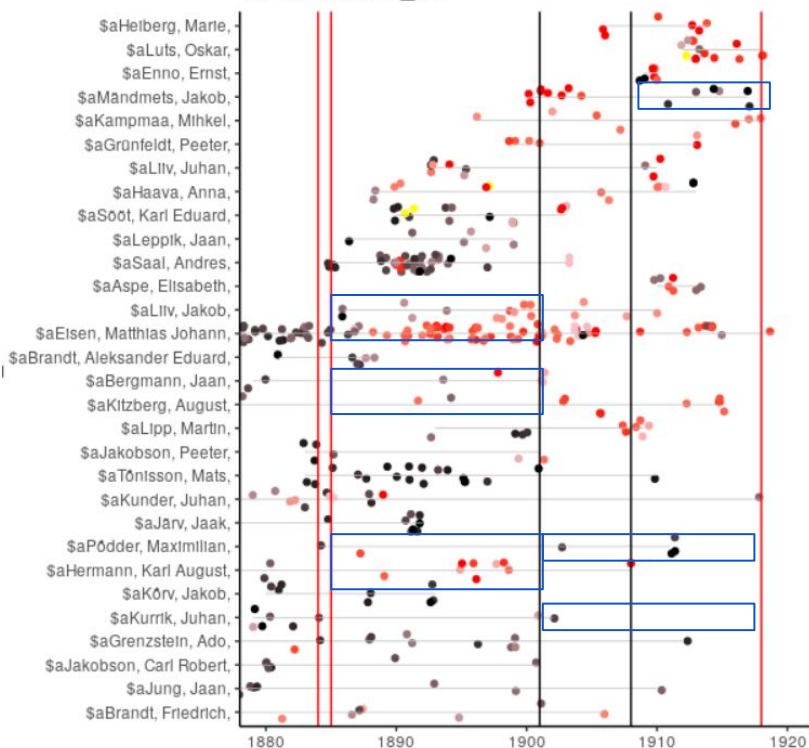


Variation of ää_var

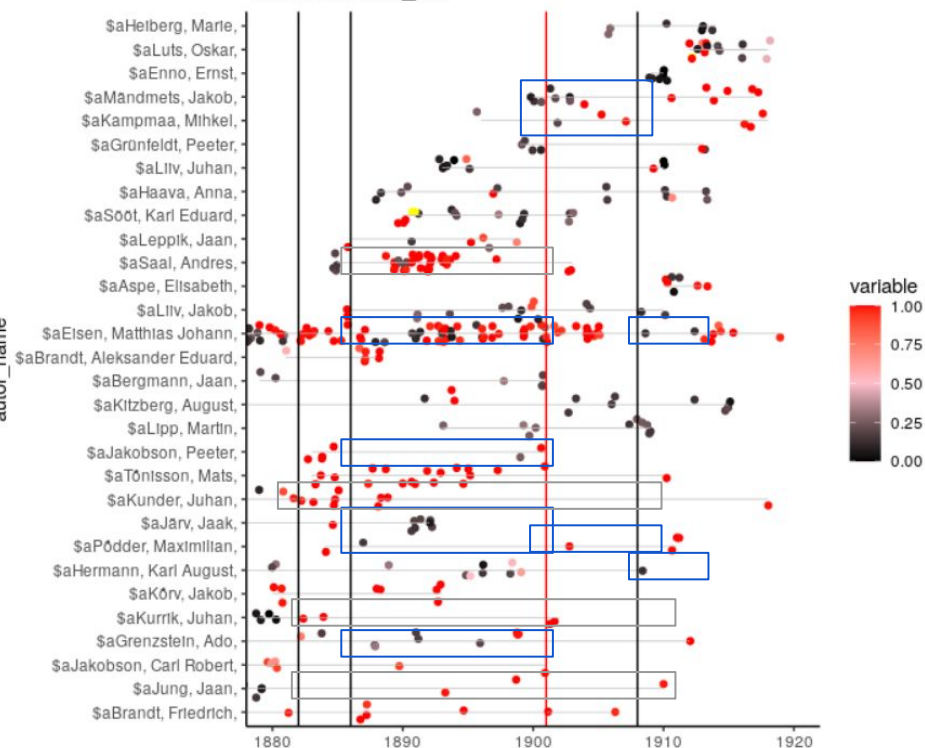


Author level variation

Variation of ste_var

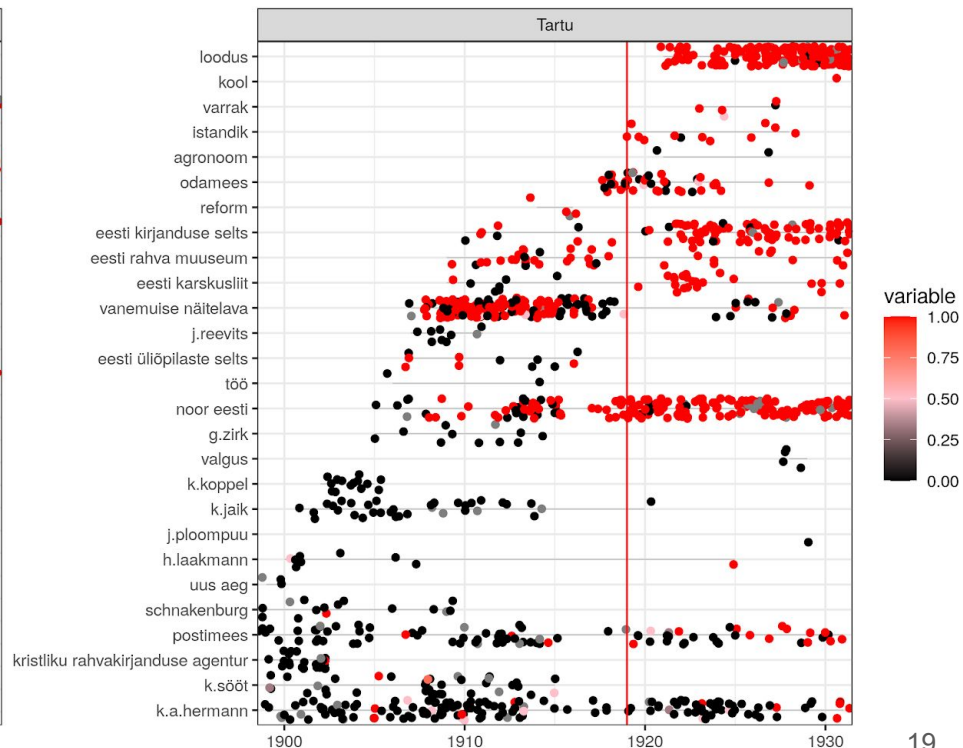
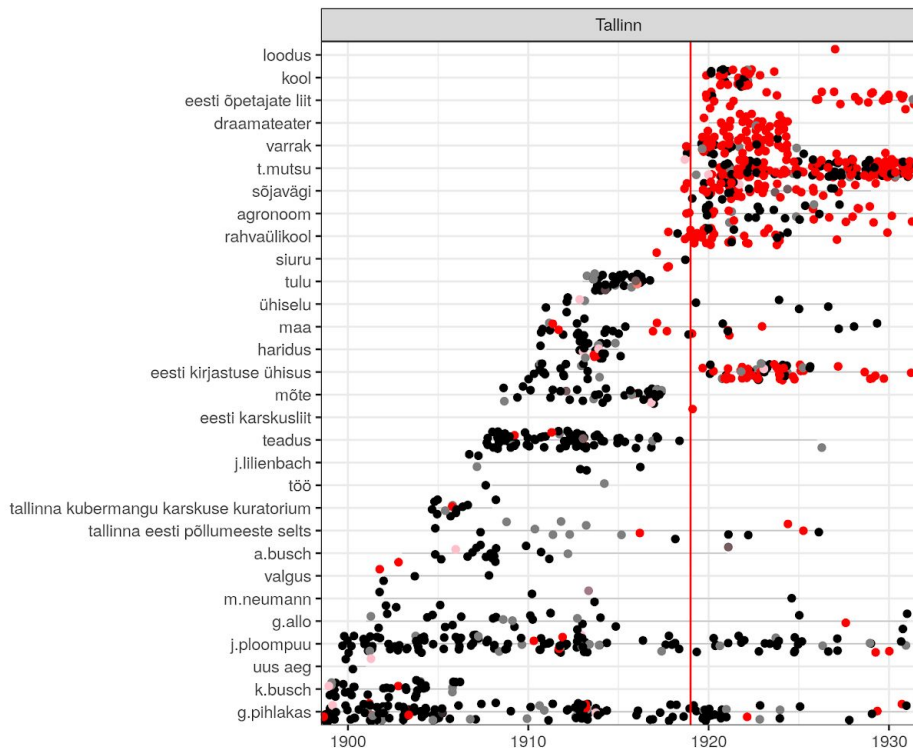


Variation of aa_var



Publisher level variation

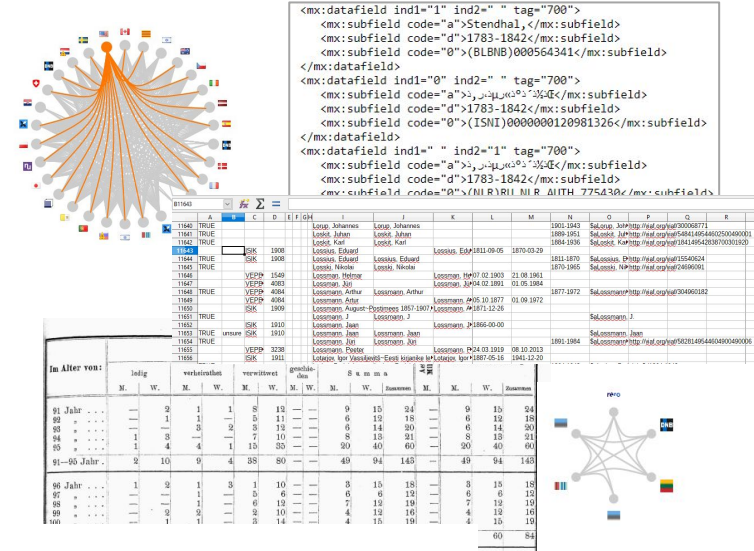
Variation of W vs V in print titles by publisher



Building author metadata

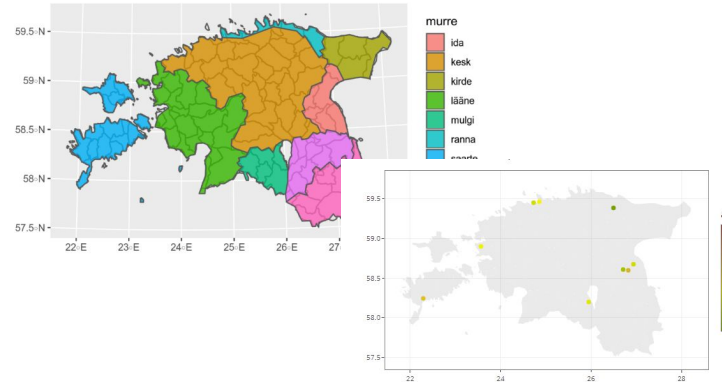
Author identities were linked to biographical information: birthdate, birthplace (dialect background), education, later home

Information from National Bibliographies,
Wikidata, ISIK+VEPER biographical collections



Questions:

- Leaders and followers
- Conservation vs innovation
- Dialect backgrounds

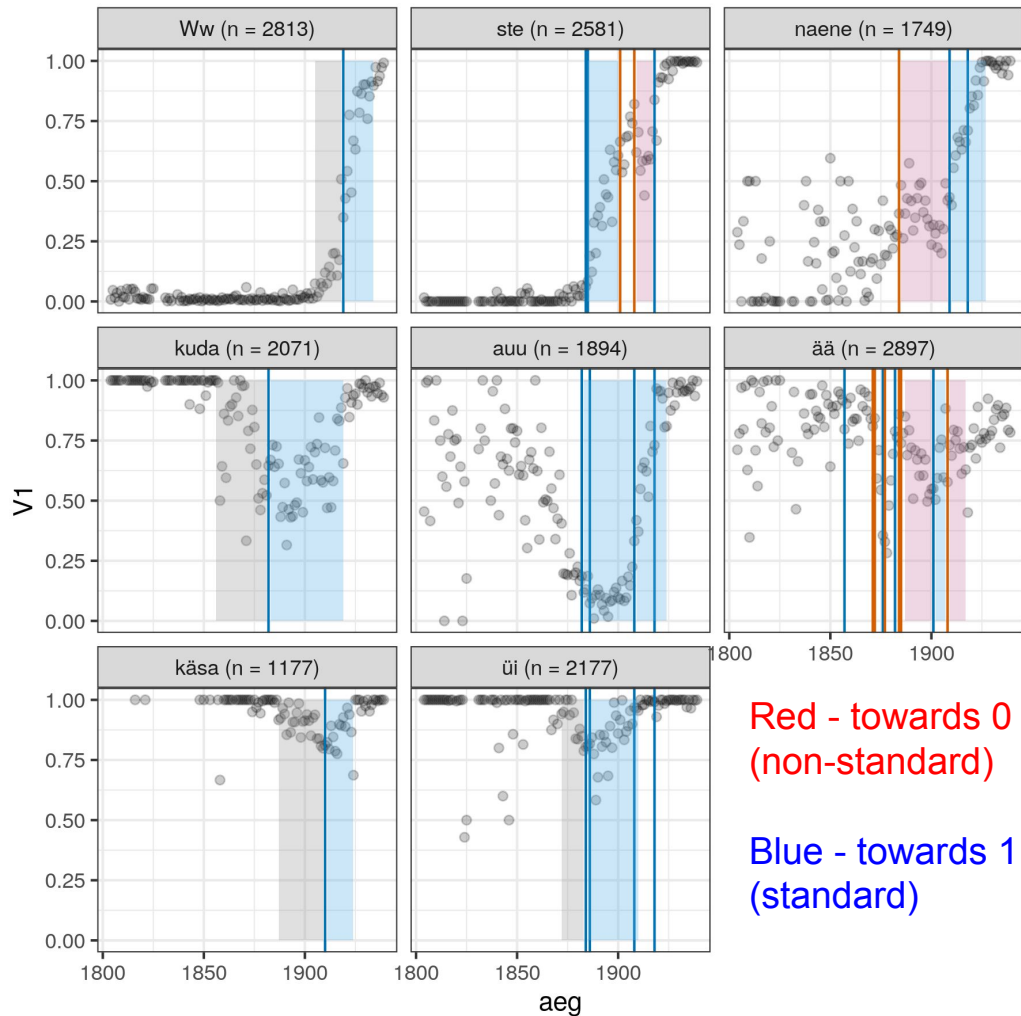


(Nevalainen et al. 2011, 2020)

Test cases

Phases with clear directions:

- Who were the leaders or followers in this case?
- Use author and publication metadata to group them.
 - E.g. years in education, birth dialect areas, age
- Predict the variant used based on these groups



Building a model

One way to model - logistic regression:

Variant used (0,1) ~ predictor1 + predictor2 + predictor3

Publication information:

- Tradition
(during last 5 years)
- Big city, small town, rural
(log population size)

Author information:

- Age (birthyear)
- Dialect (birth geolocation)
- Accommodation (later home geolocation)
- Years of education

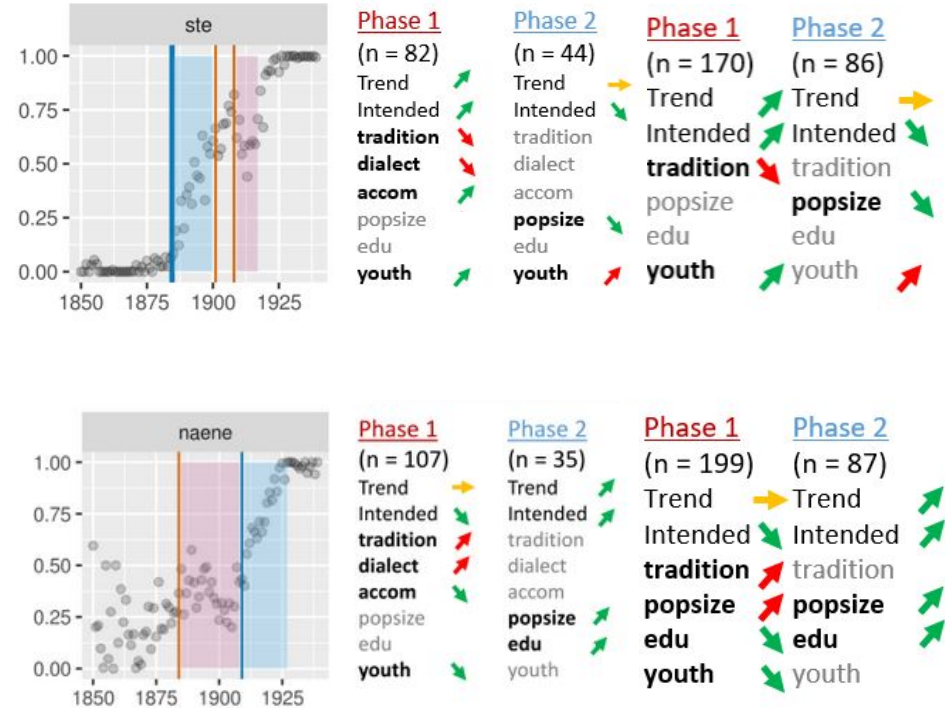
Assembling models

Model results (sig predictors in bold):

Model1: no geoinfo (more authors)

Model2: with geoinfo (+ dialect info)

- Effects vary
- A trend noticed: For successful prescription, either cities or youth had to be significant predictors.



Findings

Initial results:

Did prescriptive works matter?

Important prescriptive works often (but not always) turned the tide in use.

How?

For successful changes, big cities or young people were the leaders.

Native dialect does predict conservativeness in language use sometimes.

But these are models of convenience, more precision is possible.

General summary

- Large digital text collections + metadata open can be a very useful resource for cultural history.
- Quantified data allows for mechanisms behind historical change to be tested.
- This includes things potentially hidden from casual observation (e.g. did Estonian writers circa 1900 take language prescription seriously?)
- The uses of models can include semiotic mechanisms: e.g. the impact of cultural autocommunication in language prescription. But needs to find a good fit in data.

References

- Auer, P. (2005). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. *Perspectives on variation: Sociolinguistic, historical, comparative*, 7, 7-42.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., ... & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59, 1-26.
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson Education.
- Davies, W. V., & Langer, N. (2006). *The Making of Bad Language-Lay Linguistic Stigmatisations in German: Past And Present*. Peter Lang International Academic Publishers.
- De Rosnay, J. (1979). *The macroscope: a new world scientific system*. HarperCollins Publishers.
- Eco, U. (1995). *The search for the perfect language*. Search.
- Graham, S., Milligan, I., & Weingart, S. (2015). *Exploring big historical data: The historian's macroscope*. World Scientific Publishing Company.
- Janda, R. D., & Joseph, B. D. (2003). On language, change, and language change—or, of history, linguistics, and historical linguistics. *The handbook of historical linguistics*, 3-180.
- Labov, W. (1989). The child as linguistic historian. *Language variation and change*, 1(1), 85-97.
- Labov, W. (1966). The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.
- Leerssen, Joep, 2021. "Language interest : Europe - Introductory survey essay", Encyclopedia of Romantic Nationalism in Europe, ed. Joep Leerssen
- Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of sociolinguistics*, 5(4), 530-555.
- Nevalainen, T., Raumolin-Brunberg, H., & Mannila, H. (2011). The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change*, 23(1), 1-43.
- Nevalainen, T., Säily, T., & Vartiainen, T. (2020). Comparative sociolinguistic perspectives on the rate of linguistic change. *Journal of Historical Sociolinguistics*, 6(2).
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & communication*, 23(3-4), 193-229.
- Smaldino, P. E. (2017). Models are stupid, and we need more of them. *Computational social psychology*, 311-331.
- Raag, R. (2008). *Talurahva keelest riigikeeleks*. AS Atlex.
- Undusk, Jaan 2012. Luterlik, valgustuslik ja romantiline keeleideoloogia meie vanemas kirjakultuuris. – Vikerraar, nr 10–11, lk 73–90