Standard language culture and the case of Estonian (around 1880-1920)

Peeter Tinits

Salos, 2015

Main topic

- The processes of standardization of written Estonian as they unfolded around 1880-1920.

- Looking at it from two angles:
    - Qualitative look at contemporary descriptions
    - Corpus work on orthographic variation

- For sociolinguistic history of Estonian (but also, how attitudes may influence language diachrony)

What's the problem?

- Some rather intuitive things:

- Remarks on early literate cultures, saying that then „people did not worry too much about how they wrote / felt not much stigma for it" (e.g. Raag 1999: 135).

- And there is generally more variation in orthography in early literate cultures than later ones (e.g. Scragg 1974, Rutkowska & Rössler 2012).

- Are there 'early' and 'mature' literate cultures?

- How can we observe it in the data?

A main argument

- Language culture as a typological characteristic (e.g. Milroy 2001; Silverstein 2003; Woolard 1998)

    - Has notable variation between communities

    - Can possibly explain language dynamics (how parts of languages change or remain the same)

    - Requires comparisons between different communities to understand

- Focus on standard language culture or standard language communities in this case

Outline

▶ Standard language culture (concept)

▶ Ethnometapragmatics (relax, I will explain ☺)

▶ Background info on Estonian

▶ Standard language culture & Estonian

▶ Variation in Estonian

Standard languages and standardization studies

▶ A research programme on standardization

▶ E.g. Otto Jespersen, 1925:

„The greatest and most important phenomenon of the evolution of language in historic times has been the springing up of the great national common languages — Greek, French, English, German, etc. — the "standard" languages which have driven out, or are on the way to drive out, the local dialects." (*Mankind, Nation and Individual from a Linguistic Point of View*, p. 45)

▶ Thus it would be worthwhile to find out:

„the most important factors which — though in rather different ways and especially with different degrees of strength in different countries — have operated everywhere where a standard language has arisen" (p.46).

Towards a typology of standard languages

- Different ways:
  - Every nation, each tribe can decide whether they have a „standard".
  - A standard language has a dictionary, and army, and a navy.
  - There is a standardness in all linguistic norms.
- For a linguist, there should be something about the language dynamics for it to be useful.
  - Standard languages are all that have gone through a process of standardization, and are similar on some characteristics in this sense.

Towards a typology of standard languages

▶ E.g. Auer's (2005) practical definition for Europe, now:

▶ A language variety is a standard language iff

  ▶ (a) it is orientated to by speakers of more than one vernacular variety (which does not necessarily imply that it is mastered by everybody),

  ▶ (b) is looked upon as an H-variety and used for writing, and

  ▶ (c) it is subject to at least some codification and elaboration (Haugen 1966) or *Ausbau* (Kloss 1967).

  ▶ Importantly, (c) is an attitudinal issue.

Towards a typology of standard languages

- Thus, in this case, a standard is:
  - More than a 'common language'
  - More than an H-variety
  - Tied to people's attitudes
- Some examples:
  - (Queen's) English, Standard High German, Standard Lithuanian
- Also (of course):
  - Standard Serbian, Standard Croation
- Near-hits:
  - Middle High German in the medieval literature (not codified)
  - *Zürichdeutsch* – there is a grammar, but no Swiss would consider the variety a standard (i.e. an H-variety)

Standard languages and variation

- Standard languages (Milroy 2001)
  - Standardization: imposition of variation on a class of objects (assumes that there is a 'natural' level of variance that would be higher.
  - Strive towards a one-to-one mapping between words and meanings in languages.
- This is often correlated with a standard language culture – particular attitudes towards language in a community
  - Low tolerance for variation
  - Strong distinction between formal/public and informal/private registers, often characterized by attention to speech/language
  - Low tolerance for mistakes / active discourse on mistakes
 - Heightened legitimacy for high language register, often built through a prestigious history

Standard language culture

- Beliefs
  - Languages exist in highly idealized stable states
  - Some phrases are wrong (even though I say them often)
  - Wrong speech (e.g. lower registers) is what I do when I don't pay enough attention to my speech. In formal settings I should use good language.
  - Good speech can be learned from experts and grammars

- In sum, a standard language culture supposes that a language exist 'out there', partly independent of the community & a high functional differentiation between registers.

Communities without a standard language culture

- Modern communities without standards
  - Mühlhausler (1996), Grace (1991), Heryanto (1990) comment on Austronesian communities where the people seemed to have no sense of belonging to a language community or a differentiable language. With no centralization these communities seem to make much fluid appearances

- Early literate cultures
  - E.g. Written Middle English (1100-1500) writings showed so much variance, and also much internal variability. Example: 'through' had at least 500 different spelling variants(Smith 1996).

Attention to speech (at New York ca. 1965)

Class stratification of (r) in *guard, car, beer, beard*, etc.
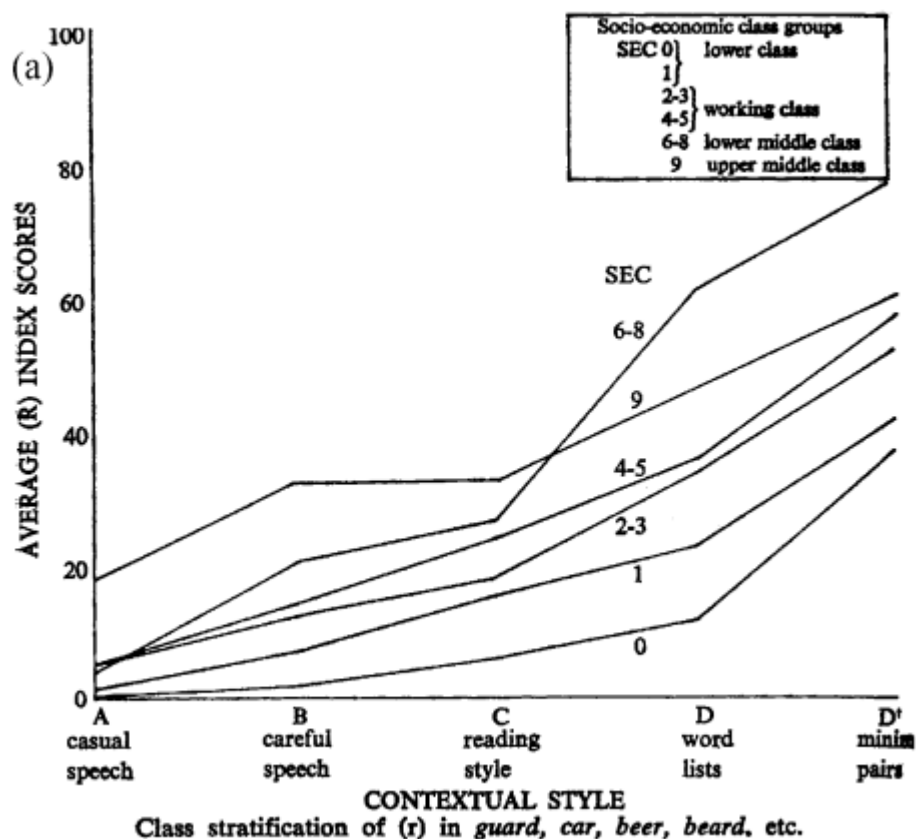
Rates of „Standard" pronunciation of NYC sociolinguistic markers (Labov).

Language culture

- General properties of language culture / language attitudes

  - Intervenes in linguistic dynamics

  - Can be altered by discourse, rituals, institutions

  - Has to do with beliefs in the community

  - Especially close to linguists

- Anthropological linguistics, sociolinguistics

Different drivers for standardization

- Standardization usually has two kinds of rather different influences intertwined:

  - Language attitudes, and the language culture (that seeks to build and maintain a common H-variety).

  - Language and dialect contacts (that can be relatively independent of people's awareness and is a natural result of accommodation in communication).

- Quite a difficult task to try and distinguish them, but arguably, the processes are quite different (e.g. see Vandenbussche & Deumert 2003).

- Possibly indistinguishable in each case, but with some trends when comparing communities.

How to approach this problem of language attitudes?

- ▶ Reflexivity in language
- ▶ Ethnometapragmatics
- ▶ Metapragmatic stereotypes

- ▶ Standard language cultures and metapragmatic stereotypes

Reflexivity in language

- In every language it is possible to talk about talk (Lucy, 1994).

  - „*Get* is an irregular verb"

  - „*Cilantro* is another name for *coriander*"

  - „One shouldn't say 'damn it' in front of children"

  - „Spanish consonants are pronounced slightly differently in Mexico than in Spain"

  - „A joke is a kind of humorous story"

- Much of language use tells listeners how to interpret what they are hearing

  - „A priest, a rabi and an imam walk into a bar…"

Reflexivity in language

▶ Speech is full of expressions that remark on language, report utterances, index and describe aspects of their speech events, invoke conventional names, guide listeners to the proper interpretation of their utterances, etc, to various degrees of explicitness.

▶ This is most obvious in dictionaries, and grammar uses, but to a degree, every utterance can be argued to be normative (e.g. Johnstone 2006) in providing a model for a way utterances can sound and how reoccurring forms can be linked with reoccurring functions.

▶ Thus this metalinguistic, or reflexive function can be found quite everywhere.

Folk-linguistics

- Some diverse uses
  - Attitudes and beliefs on language that are held as common sense in a community
  - How language is understood in the community (of non-linguists)
  - Can include conscious and unconscious attitudes

- Similar to studies like folk-physics or folk-mathematics (the way people understand and percieve physics and mathematics in their daily interactions).

- No need to privilege certain groups in the society, but can be a general guide-line to understand cultures.
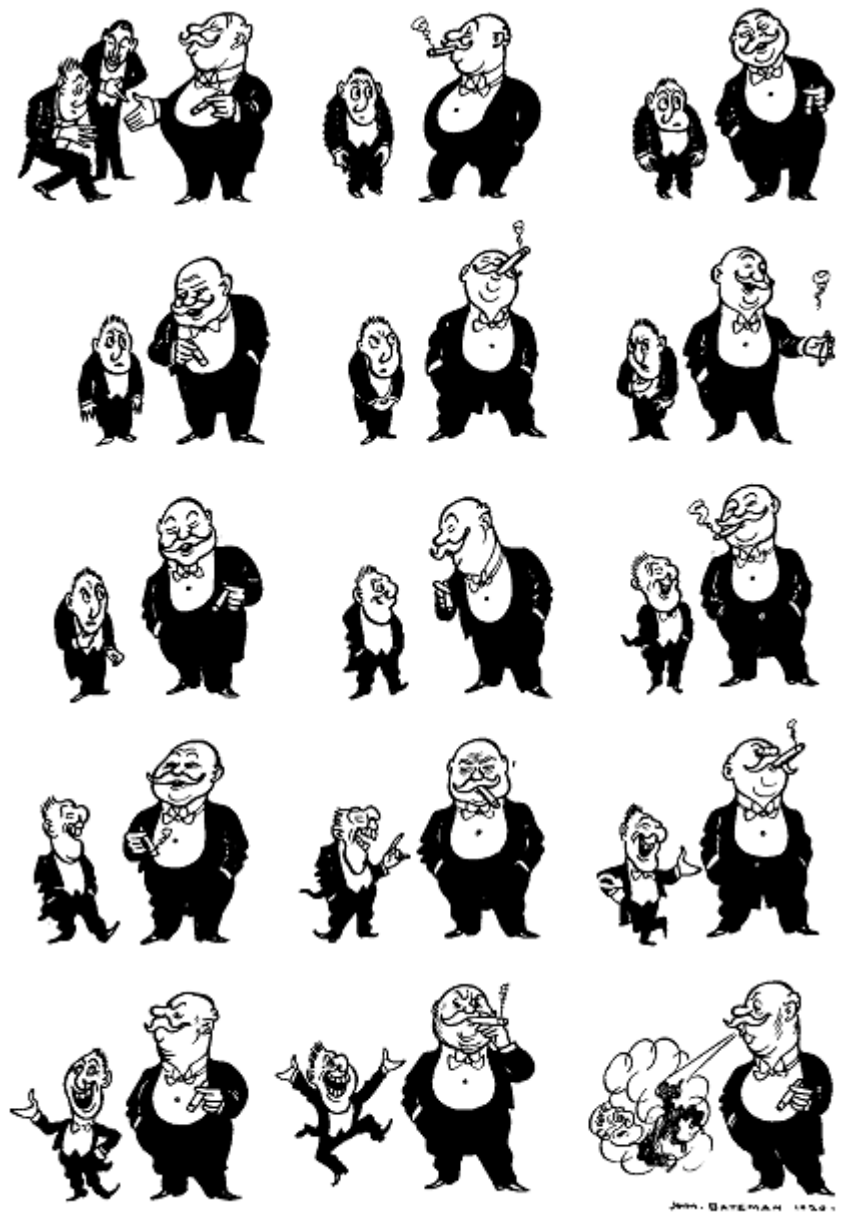
Ethno-metapragmatics

- Based on Ethno-methodology (from anthropology)
  - The study & practice of the methods by which people solve their common daily problems by.
  - Main issue:
    - In their daily lives people follow their habits and would not be described as being clearly rational.
    - But their actions do follow an internal logic, and this internal logic should be studied on its own.
- Same with ethno-metapragmatics
  - Metapragmatics – rules which people use to manage the pragmatics of language use
  - Ethno-metapragmatics – Emphasizes its habitual nature and cross-cultural variation

Ethno-metapragmatics

▶ The way people manage and maintain the metapragmatic functions utilized in language use.

 ▶ How they are interpreted

 ▶ How they are negotiated

 ▶ What kind of normative expectations are for them

 ▶ Etc

▶ The key to this is of course that there is variation within the community and between communities in how some linguistic actions can be interpreted and how they will be interpreted.

Metapragmatic stereotypes

▶ Surround us in all sorts of modalities and messages.

▶ Carry the metapragmatic habits in the community.

▶ Are necessarily legible by (new) members of the community.

▶ Thus, also linguists can read these cues.

(Bateman cartoon, 1920)

Metapragmatic steroetypes

▶ Culture-internal models of utterance indexicality associated with speech variants. These models set indexical text-defaults on the construal of utterances for persons acquainted with them. (Agha 2007: 148)

▶ Easiest: all members of a population sampled typify criterial speech forms in the same way, e.g., assign the same metalinguistic predicates to the forms at issue (viz., 'X is slang,''X is polite,''X is used by older persons,' etc.) ; or all treat the use of these forms as appropriate to a given scenario of use.

▶ But usually not everyone has the same stereotypes, and the ones you have, depend on exposure and interests of the people involved.

An epistemological note

▶ Metapragmatic stereotypes are not ideas in the head.

▶ The main evidence for their existence lies in overt (publicly perceivable) behaviors that evaluate the pragmatic properties of linguistic expressions.

▶ You can ask for introspection about beliefs, but that is a special case.

▶ Orderliness of the social life is a mutual coordination problem, a question of how partially overlapping perspectives on a register's forms and values can yield orderly forms of interaction among its users.

▶ The existence of social norms of language use depends on the possibility to recognize and acquire them.

Case of Queen's English

- Agha's analysis on the spread of Standard British English (or Received Pronunciation, or Queen's English): ample evidence of how accent values moved in space and time in the 18th – 19 century.

- Start from a prestige variety of the elite around London, Oxford, Cambridge elite in 16th century.

- Small number of speakers, regional prestige variety only.

- Recommended as a literary variety, but not recognized as a standard at wide.

Emergence of a standard

► Regional sociolect -> supra-local standard

► Via characterological constructs tied to sound disseminated through discourse communicating them.

► Gradual process by which the register becomes known and recognizable in a wider population.

► 18th century, the previous prestige court sociolect comes to be championed as a model for a national standard, but initially only for a small population.

$$[S \longrightarrow R] \quad [S \longrightarrow R] \quad [S \longrightarrow R] \quad [S \longrightarrow R] \ldots\ldots$$

time
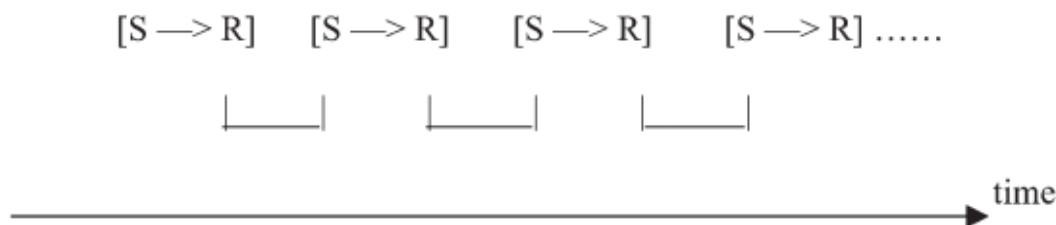
Social personae linked to sound

- 1760-1900 proliferation of a range of genres of metadiscourse on language use.

- Different in functions and mechanisms, but all of them promote this RP, by formulating social personae linked to sound.

- Recievers of the message can respond to them by aligning themselves in relation to this.

- For each text, it is possible to sketch their diffusion and audience within the community, and the interests of these communities.

Messages that travel

▶ Writings on pronunciation (17th cent descriptive, 18th century prescriptive, very narrow circulation)

▶ Even: Orthography for pronunciation
1761 – Thomas Sheridan „ To fix a standard of pronunciation by means of visible marks" (1761: 29), and make it available for everyone via grammars and dictionaries.

▶ Result: some unification among the educated, but heightened class consciousness at wide.

Table 4
Rough estimates of genre circulation

Genre	Immediate circulation	Demographics: circulation extends to
(a) Early prescriptivist works	Small	(a) aristocracy & intelligentsia
(b) Popular handbooks	Larger	+ (b) upper middle classes
(c) Novels & literary works	Large	+ (c) middle classes
(d) Penny Weeklies	Very large	+ (d) lower middle & upper working classes

Messages that travel

- Popular handbooks (early to mid 19th century): etiquette manuals, handbooks on pronunciation, elocution and grammar

- Non-technical style for a wider audience

- Result: increasing anxiety on accents, arising professional of orthoepists and elocution masters

- Describe differences within the community and everyday interactions – „how to be educated and well-mannered"

- „A person who uses vulgarisms will make but <u>little way in good circles</u>... A proper accent gives importance to what you say, engages the respectful attention of your hearer, and <u>is your passport</u> to new circles of acquaintance." (*Talking and Debating*, London, 1856)

Differentiation of registers

- Specifically emphasizing differences:

- „There are certain arbitrary peculiarities of manner, speech, language, taste, &c. which mark the high-born and high-bred. These should be observed and had. They are the signs-manual of good-breeding by which gentlemen recognize each other wherever they meet." (*Advice to a Young Gentleman on Entering Society*, 1839, p. 138)

- Hence knowing this language is useful, and has a direct market value, when one wants to advance in the society

- Proper manners all around.

Dramatizing the characters

- Literary works (to include also the expanding middle classes)
  - Many of the famous novelists were avid readers of the first two genres
- Novelistic depictions of accent amplify and transform these characters into more memorable forms.
  - 'Me, Master Copperfield?' said Uriah. 'Oh, no! I'm a very umble person.
- They dramatize the effects of using some accents (if they are marked in writing)
- Repeating characters expand the awareness of social norms.

To the masses

- Penny weeklies (cheap literature for the masses in 19th century Britain), mid 1850s circulation ~750,000 – much more than the best novels, including also lower middle and upper working classes.

- Read often within families, passed on to children.

- Readers are aware of prospects of social mobility, and look for role models on purpose. Letters show aims to eradicate their heritage to become cultured.

- Popular stories of slightly higher social classes.

Transformation of habits

▶ The metadiscourse presupposes and amplifies the language habits as a means for social advancement.

▶ These genres so far mostly teach to recognize those speech accent values, but much less to produce them.

▶ 1870s onwards, boarding schools become a habit for the higher classes, where general education leads to 'public schools' differentiating from mass schools

▶ Class-room teaching of pronunciation

Table 4
Rough estimates of genre circulation

Genre	Immediate circulation	Demographics: circulation extends to
(a) Early prescriptivist works	Small	(a) aristocracy & intelligentsia
(b) Popular handbooks	Larger	+ (b) upper middle classes
(c) Novels & literary works	Large	+ (c) middle classes
(d) Penny Weeklies	Very large	+ (d) lower middle & upper working classes

# Asymmetries in competence

▶ RP becomes an object of scarcity, and due to its patterns of acquisition of value and social advancement.

▶ This scarcity itself maintains the value in time.

▶ Future exemplary speakers – 1900 public schools, 1930 British army officers, 1970s BBC radio & TV – (descriptive role models!)

▶ 2000 – Even the Queen speaks the Queen's English no more.

▶ Nowadays, most people recognize RP, but still only 3% of the British native speakers have full command of it.

Attention to speech via ethno-metapragmatics

- ▶ Attention to speech is a type of ethno-metapragmatics.

- ▶ Attention to speech works via metapragmatic stereotypes and through the reflection of the people involved.

- ▶ Attention to speech includes mechanisms by which newcomers to the community can learn and acquire this way of organizing their patterns of language use.

Attention to speech (at New York ca. 1965)

Class stratification of (r) in *guard, car, beer, beard*, etc.
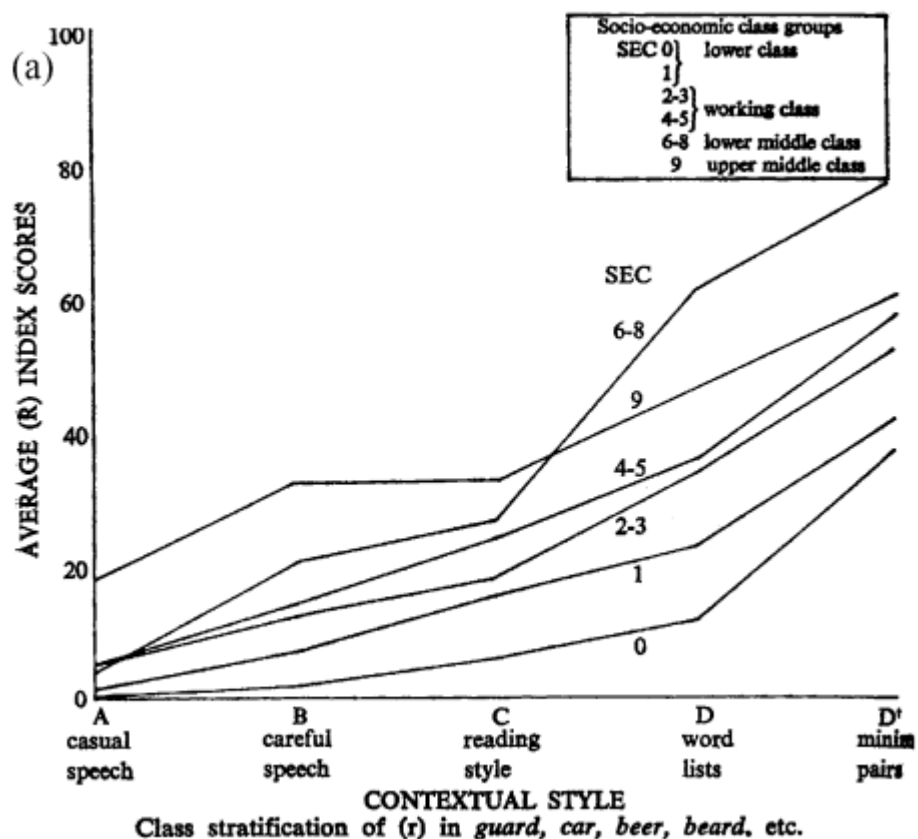
Rates of „Standard" pronunciation of NYC sociolinguistic markers (Labov).

Tools for thinking about language culture

▶ Ethno-metapragmatics in a community – habits of interpretation, and solutions to novel situations.

▶ There is variation in ethno-metapragmatics of a community.

▶ Standard language culture is a certain canonical community type, not present in all communities around the world.

▶ Compared to the ideal type, both in diachrony and synchrony these communities may differ from each other.

The case of Estonian

- Estonian language community 1880-1920:

- Special part of Russian empire run by the Germans

- Purely peasant community until the 1850s (reforms of 1802, 1804, 1809, 1816/1819, 1849/1856 changed this)

- Quickly urbanizing and modernizing

- Spread of romanticism, 1860s – 1885 known as national awakening

- 1885 – 1914 Russian policies of assimilation significantly changed the governates for both Germans and Estonians.

c. 1890

Estland (Reval) Gubernia
Livland Gubernia
Kurland Gubernia

WHKMLA
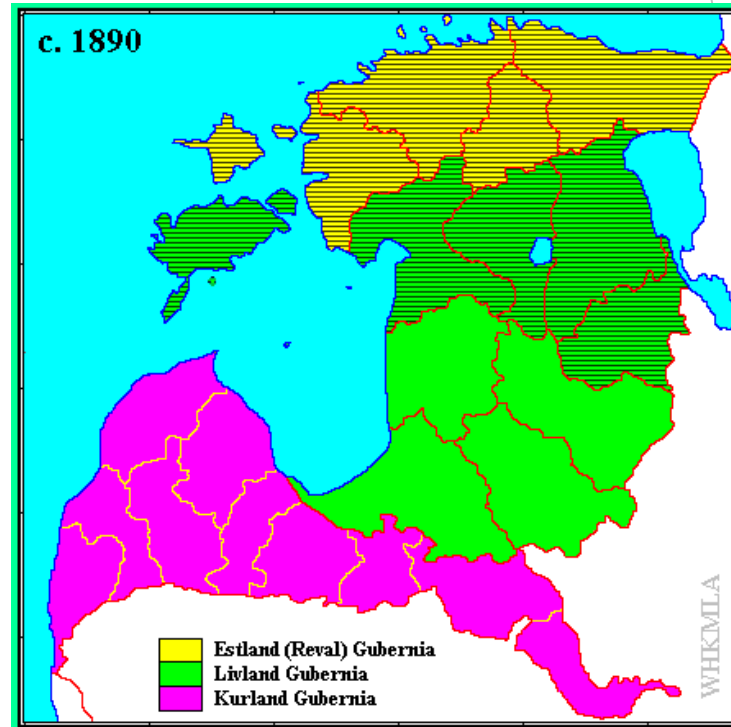
Language situation for Estonian

- Pre 1850s:
  - German language community, Estonian language community
  - Writing traditionally for German clerics and romanticists
- Since 1857:
  - An increase in Estonian to Estonian written communication (journalism and then other fields)
  - Emerging Estonian to Estonian communication, urbanization, social mobility increase variation in writing
  - Some implications of a 'lack of stigma' on language use initially
- 'Language improvement' discussions emerge in 1870s

Variation

- Described to be rather high by later linguistic assessments (of course their baseline is whatever happened next) (Kask 1984; Laanekask 2004; Raag 2008)

  - Used to be Tallinn and Tartu enregistered written varieties, this changed into one with one central variety sought after

  - Influx of dialect speakers from the country, language contacts in the cities

- Schooling for most basic levels in Estonian (interfered 1885-1914), but teachers are complaining they don't know what to teach (1876, 1886, 1912)

  - Notably quick changes possible: (e.g. transition from 'old' to 'new' writing: 1869 20% of books use 'new system', 1873 50%, 1879, 90%)

  - High literacy (1897, 96% of over 9-year olds can read)

Conscious intervention

▶ Note, linguist was a very high prestige job in late 19th century

▶ First emerges for vocabulary extension and graphization reforms

▶ The 'problem' of variance and parallel forms emerges in the awareness of language enthusiasts by 1900.

▶ A number of public and publicized discussions on particular linguistic variants (notably 1972, 1886, 1908-1911)

▶ Language arguments sometimes turned political

▶ 1920s was heavily prescriptive, 1930s seems to have reached a common standard considered as fixed

Discourse on variation

▶ Language was quite focal in discussions.
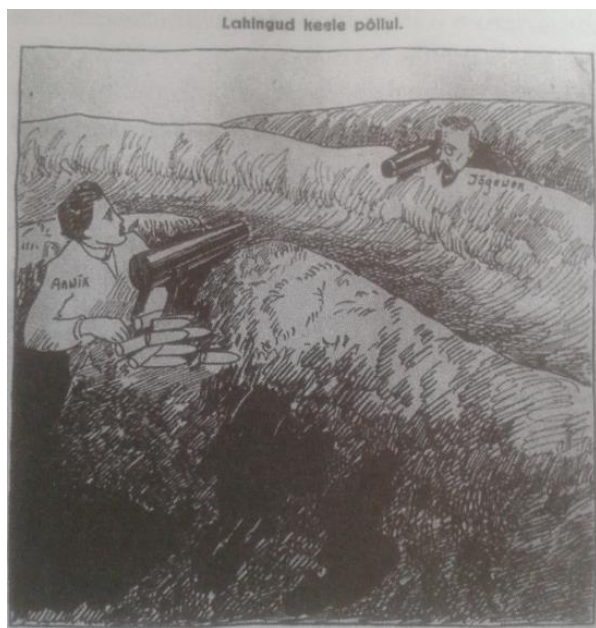
Aavik aims with his Linguistic Monthly, Jõgever defends with the wisdom of the folk. (*Sädemed* 10, 1914)
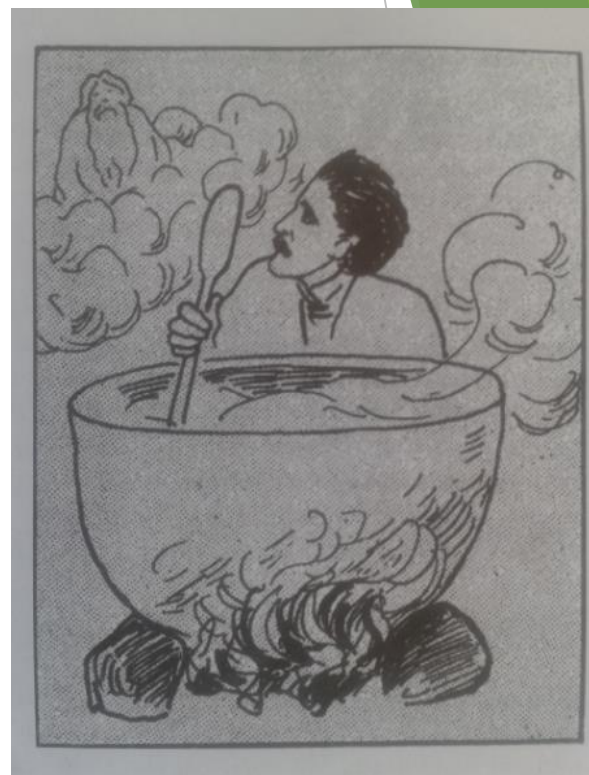
Aavik mixes the Old Man's recipe of language into language porridge according to Finnish cuisine with French spices, and blames the Old Man to not have been much of a cook. Sädemed nr 7, 1913.

Materials available

▶ There were frequent discussions on language in newspapers and separate publications

▶ Various aspects in focus, writing on grammar, functions and the history of language.

▶ Adds up to a few hundred articles, but uneven distribution with authors – not fully mapped, but a few authors wrote most of the texts, and some more authors wrote on just a few occasions.

▶ The audience was mostly conceived as the regular readers, although their lack of education was often commented on.

Discourse on standard language

- From the concept of standard language culture
  - Extent of variation
  - Elevation of a central register
  - Register differentiation
  - Attention to language use
  - Discourse on mistakes
  - Historical perspectives on language

Extent of variation

- ▶ Was made an explicit focus of language discussions (and the 'biggest worry with the Estonian language' around 1900.

- ▶ Under discussion since 1870s, initiators school-teachers, translators, linguists.

- ▶ Mostly about reduction of this variation: „We need to agree on standards"

- ▶ But with some counteractive reflections on this: „Maybe we could use some interregnum before we fix the rules."

- ▶ But general trend matches the 'reduction of variation' of Milroy, and in this case driven by conscious efforts (at least next to language contacts of increased mobility).

Influence of discourse on variation

- Societies and committees formed to improve language
- No legislative power only informal influence
- Lots of arguments with various argumentation
  - comparative-historical (old is good, related is good) & other linguistic theories
  - own dialectal background
  - if it sounds nice
- It seems to be an overly dominant paradigm to assume that there can be only one. Variance was assumed to be bad for the language.
- At the same time, choosing between parallel forms was one part of the story. In language elaboration many new word forms were borrowed, derived and created which increased the variation.

Extent of variation

▶ People were very reflexive on this issue, some wanted more:

▶ „Language must not only be organized, but improved and made greater. Now is the time for it. Ten-twenty years from now, when the language has become solid and fixed through common use, it will be much more difficult, almost impossible. Strike the iron while it's hot!" (Aavik 1914a: 2)

▶ „Some people, especially older ones, complain that too many novelties will bring confusion into the language in some type of an interregnum. [...] We dont see a danger in that, the bigger the confusion the better [...] language will come out cleaner, clearer, prettier and more complete from this purgatory." (Aavik 1913m: 99)

▶ Aavik sees a great opportunity here, arguing that „if our language was not complete" we would not have the amount of internal variation to choose from (Aavik 1912f: 461). He finds for example forms that were used frequently a few decades before that he considers more beautiful and argues that they could be returned to use (Ibid. 466)

Elevation of a standard register

▶ „Written language" is felt as naturally a different **kind** of language then common spoken language has been.

▶ Aavik: „written language is already by definition different from the common language: it has completely different tasks and demands to fulfil than a peasant dialect. It is naturally and necessarily richer and more complicated in its syntax and vocabulary: it has its own phraseology that has emerged mostly through books. (1913p: 156).

▶ Aavik : „written language, if it really wants to be written language – i.e. educated language – can't always be concerned with whether some phenomenon exists in common language and whether it can use it for this reason"(1912 i: 367).

▶ Thus part of a natural development of a mature civilized culture to have a 'written language register'.

Elevation of a standard register

- ▶ The written language should be a language for the educated.

- ▶ Aavik: „we must renounce the biased notion that any bloke without study and education would understand it. That the written or book language of educated people is difficult to understand for a 'common person' is unavoidable and even large culture nations have not been able to prevent this from happening"(1914b: 37).

- ▶ But also: 'common language should not be contrasted to the written language nor appealed to such common language's majesty'. (Aavik 1913p: 149)

- ▶ The dynamics of 'written language' are construed as partially independent of its users.

Register differentiation

- Written language is explicitly described as something that is not spoken natively by anyone.

- „It seems to us that this kind of attempted similarity with common language limits the free development of the written language style. Accordingly we think that now it is most appropriate and recommend to fall into the vice of complicated and artificial sentence structure and subtler style than sin with the rudely simple way of writing. Since Jannsen we have had enough of this type of common language fluency. The opposing extremity would not hurt us too much. There will always be time for reaction or backtracking in case that is needed." (Aavik 1913p: 150)

Attention to language use

- There were explicit complaints on people using language without enough attention, both in written and spoken language.

- Aavik: [people write] „however they please and how it happens, falling into all sorts of false analogies, rude and simple linguistic errors and grammatical foolishness" (Aavik 1914l: 2).

- Veski: „other nations go to the theatre to hear the best language, to learn, we however have to settle with a mish-mash of dictions tolerated due to our inferiority" (Veski 1914b: 176).

Discourse on mistakes

▶ Discussions on orthographic variation by the schoolmasters since the 1870s, e.g. Kurrik in 1876 'the legion of writing styles in the language impedes teaching the language and thus some organization is needed for it'.

▶ 1886, again these debates in the newspapers, trying to find some agreement on the presence of variation.

Discourse of mistakes.

▶ Awareness of dialects as not mistaken, but different correct items „what is correct for one, is wrong for another" (Aavik 1913p: 149); „there is no common language only singular dialects" (Aavik 1913p: 149).

▶ But wrong language is wrong language: „our books, even from the best authors, are still full of all sorts of mistakes, faults, lazyness […] which can't be considered correct according to the current written language standards" (Aavik 1913l: 315-316)

▶ An idea, that since we haven't agreed, there are no mistakes, or actually no correct language to speak of.

▶ Aavik: „there are no 'correct' speakers as there is no 'correct language' to speak of" (Aavik 1914a: 2)

Native instincts of language

- Jürgenstein (1914): I trust my (unspoiled) language instinct much more than a grammarian's.

- Aavik (1914): Nice joke! Educated Estonian with an unspoiled language instinct!?

- Common complaints on bilinguals having darkened language instincts.

Historical perspectives on language

▶ Age-old romanticist belief – legitimacy and prestige can come through a prestigious history.

▶ Some notable renegotiation of the historical interpretations of Estonian.

▶ An interesting conceptualization around 1914, Aavik:

  ▶ 1850s-1860s peasant language, 1870s German with Estonian words, 1880s rise towards semi-autonomy, 1900s cosmopolitan language in increasing translation, 1920 onwards – bright future of the beautiful cultural language

More on mistakes

- A genre of language reviews on publications

- A particular interesting example, a review on Kalevipoeg (written in the 1840s-1850s) at 1914:

- 'the epic has a „legion" of linguistic mistakes and is almost „exploding from them" (Ibid. 10). Aavik particularly describes this as a sign of the times when language „was in a confused and uncertain state, and there was little worry about correctness" (1914a:17),

- Comparing himself with one of the authors saying „learnt his Estonian from the people, like a bird learns her song"

- Both the language of the authors is wrong and lazy, as well as the dialects being wrong and lazy.

More details

- ▶ Frequent explicit comparisons with other languages

- ▶ Strive to become a „cultural" language

- ▶ Mentioned a reason for language improvement:

  - ▶ This is extremely important from the perspective of the nationalist struggle: an educated and beautiful native language will be considered as more dear and people will be less willing to let go of it in exchange of another. In giving the idea more value also the national vitality and resilience in the struggle for survival will be raised. It should be obvious to everybody. (Aavik 1918)

An intermediary conclusion

- The development towards a standard language culture is reflected in the contemporary commentaries.

- Standard language culture seems to have been an explicit target (with some variations) on several dimensions.

- These discussions on language, how to use it, and how it should be may have influence on the language change.

- These processes have been common around the world – but it is difficult to evaluate whether this would be in any way 'natural' for a community to develop a standard language culture (e.g. when writing becomes common) or to what extent it may have been due to cultural influence.

The Baltic similarities?

▶ Similar histories and possibilities in the history of the society?

▶ But different set of individuals leading the processes?

▶ Any similarities in descriptions?

▶ Any differences?

Discourse on standard language

- From the concept of standard language culture
  - Extent of variation
  - Elevation of a central register
  - Register differentiation
  - Attention to language use
  - Discourse on mistakes
  - Historical perspectives on language

Corpus analyses

- Orthographic variation
    - Easiest to change
    - Very often tied to social meaning
    - Very easy to notice by the speakers
    - Very often explicitly commented on
- Text corpus allows easy searching
    - Can be connected to metadata on authors and books
    - Distribution of variants can be easily compared

Looking for a standard in a corpus

- What is a standard?
  - An agreed-upon centrally negotiated means of using a language
- Evidence for a standard
  - In learning (you change your 'natural' expression on joining a community)
  - In renegotiation (community can be seen to correlate in the change of their language use)

Dynamics of language change

▶ Language change on standardization may differ from regular language change
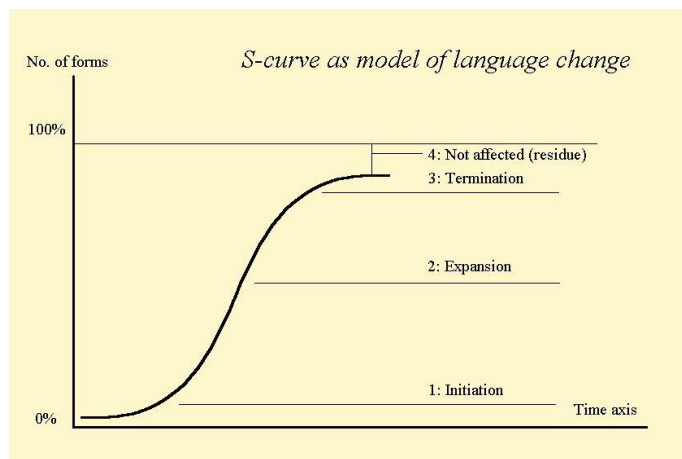
Regular language change
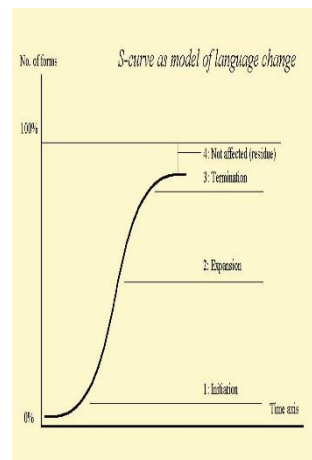
Standardization may have much sharper curves. (picture is only illustrative)

What patterns could you find?

- Ideally, you could find the allegiances in the group

  - What are the parameters that determined the choice in some variable in question

  - Connected to dialectal background?

  - Connected to education?

  - Connected to text genre (letters, print)?

  - Connected to residency?

  - Connected to explicit recommendations?

Variables

- 4 slightly different ones
  - 'ea' and 'ää' (*pea-pää* [head], *hea-hää* [good])
  - 'sti' and 'ste' (*kenasti-kenaste* (nicely), *armsasti-armsaste* (dearly))
  - 'ai' and 'ae' (*naine-naene* (woman), *laine-laene* (wave)
  - 'au' and 'auu' (*honour), 'õu' and 'õuu' (nõu* – advice)
- All mentioned in prescriptive works
- Most have a connection to dialectal background in pronunciation, but individual backgrounds are not always clear on their early life communties.

Corpora in use

▶ printed publications in digitization (left, yellow)

  ▶ Literary classics ~1879-1920 (~33 authors, ~90 publications, 1,233,170 strings), uneven authorship distribution, good info on life histories

▶ 2 collections of personal letters (1 – right red, 2 – right green)

  ▶ 1) Urbanizing writers ~1900-1920 (~10 writers, ~243 letters, 71,224 strings)

  ▶ 2) Letters from folklore collectors~1888-1896 (~269 wr., ~290 let., 51,908 strings), good geographic coverage, uncertain authorship

Time-span covered

▶ Not evenly distrubed, but covers the interval somewhat.
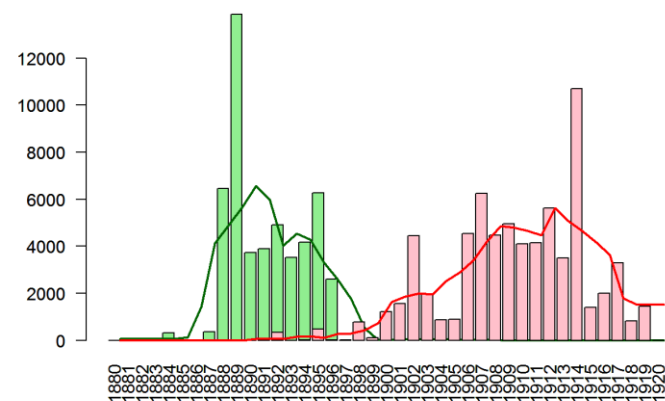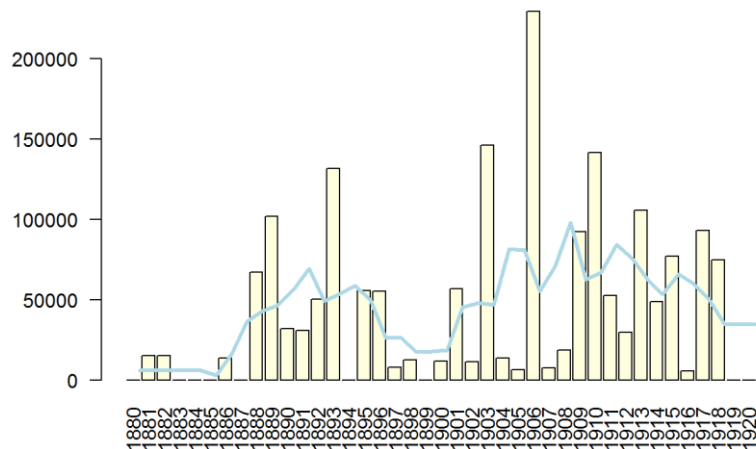
Language change

- How did particular linguistic variables travel within the community?

- Why did people choose the variants they chose?

- Change in time

- Age-gradient, education, chosen domicile

Calculations

- Basis is a set of synonym pairs (e.g. *naine-naene,naisele-naesele) where absolute frequencies are counted.*

- *How much (%) modern standard forms vs non-standard*

$$\hat{u} = \frac{\sum_{i=1}^{N} a_i}{\sum_{i=1}^{N} (a_i + b_i)}$$

- If many lemmas, take each as independent & equal source of evidence: e.g. (ratio of *naine* + ratio of *laine*)/N

$$\hat{u} = \frac{1}{N} \left(\sum_{i=1}^{N} \frac{a_i}{a_i + b_i} \right)$$

- For authors in group, the same, each author has a ratio (does not depend on corpus size in its contribution)

Social groupings among the authors

- We can look at a priori categories:
    - Birthyear
    - Birth location
    - Education
    - Chosen residence
- In principle we could track social networks and movement in time
    - But: lack of data
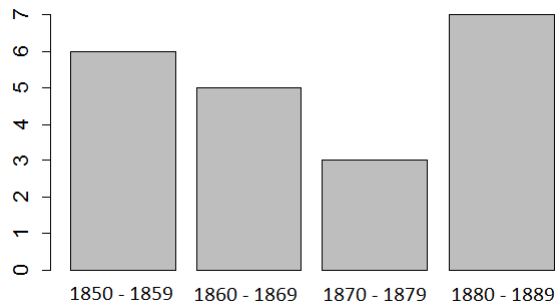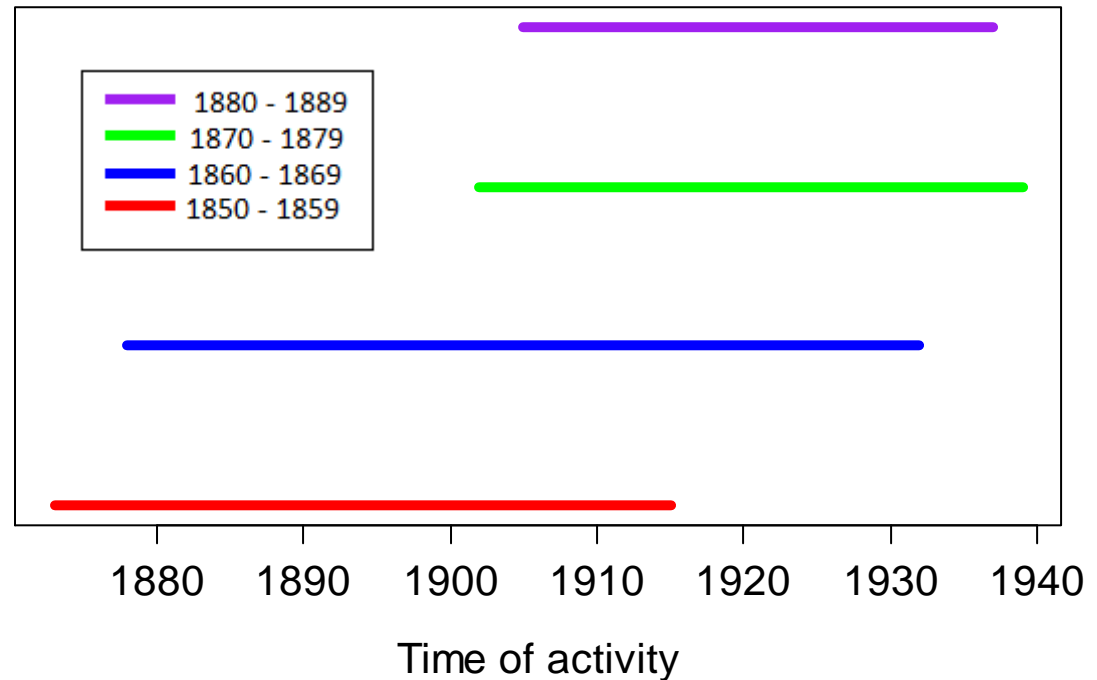   - &: difficult to model

Publication corpus

▶ Cohort

Cohort 4: 1880-1889, 7 ind-s
Cohort 3: 1870-1879, 3 ind-s
Cohort 2: 1860-1869, 6 ind-s
Cohort 1: 1850-1859, 5 ind-s

Author (ordered by birthyear)

Time of activity

1860   1880   1900   1920   1940
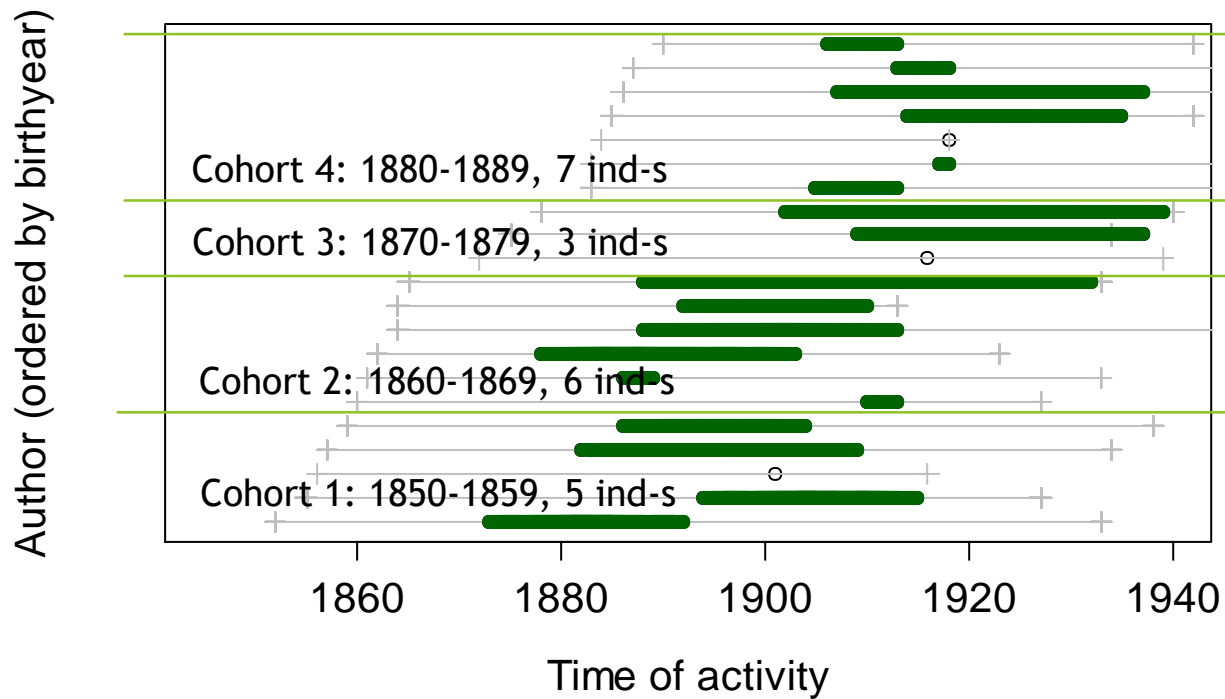
Grouping by birth

▶ Time of activity by

Time of activity per cohort

Number of authors per cohort

Legend:
- 1880 - 1889
- 1870 - 1879
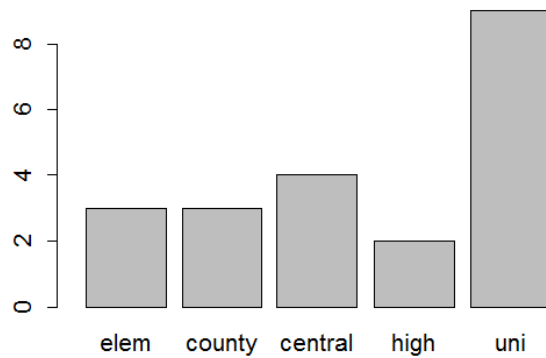- 1860 - 1869
- 1850 - 1859

Time of activity

Grouping by education

▶ Education cohort, time of activity

Time of activity per cohort

Number of authors per cohort
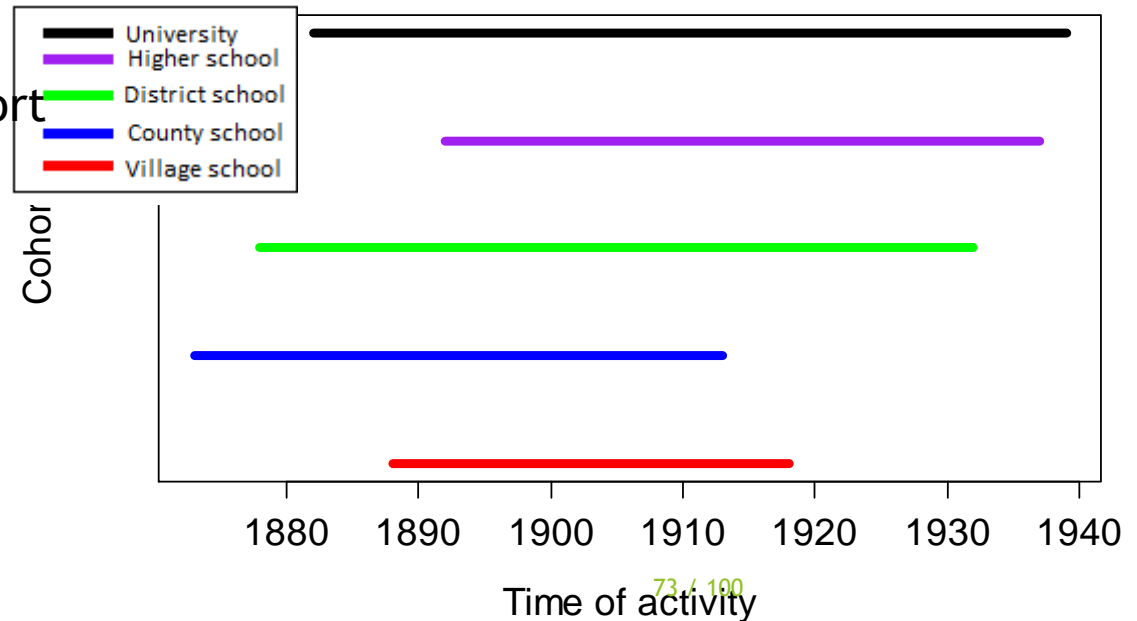
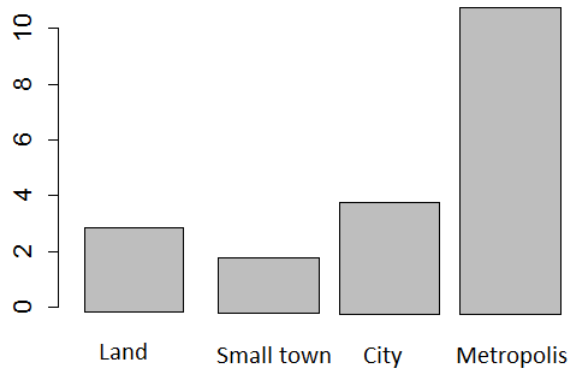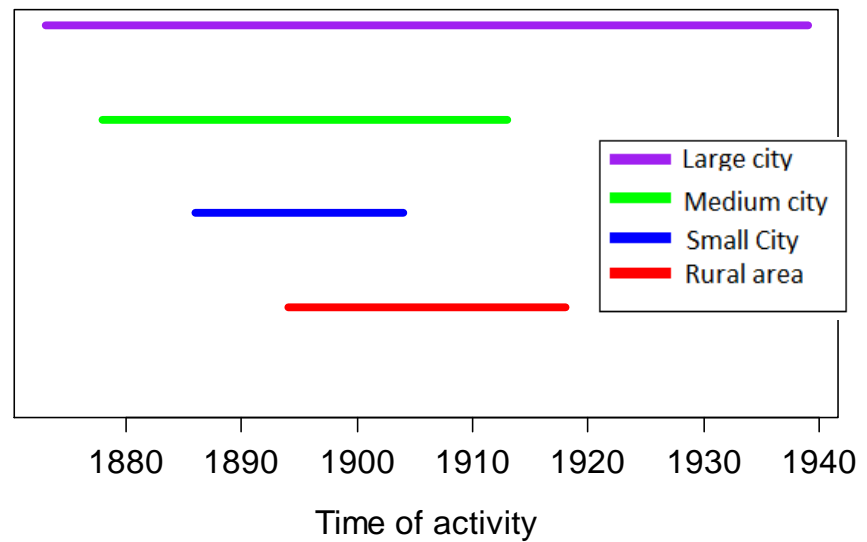# Grouping by domicile

▶ Activity in time by domicile

Time of activity per cohort

Number of authors per cohort

Some issues with the data

- Sparse and uneven data
- Representativity depends on availability (but what should it be representative of?)
  - Overrepresentation (?) of higher education levels is typical in historical sociolingiustics
- How to calculate across time – to give each piece of evidence a fair value?

'ea' – 'ää'

- Includes lemmas *hea, pea, seal*

- Quite frequent debates, recommendations in opposite directions; has ntable dialectal background (*ea* is in biggest dialect northern central one; *ää* is common elsewhere)

- **1872 – use *ää!*** [public and published speeches (J. Hurt): incl. „use new writing system", use *ää* instead of *ea* always, *ää* was argued to be more archaic and having wider presence in dialects]

- **1884 – use *ää!*** [K. A. Hermann published an influential grammar]

- **1886 – use *ea!*** [Committee for making an improved new testament; use *hea* and *pea]*

- **1908 – use *ea* in the words in focus, *ää* almost everywhere else!** [One of the main points in first Language Conference 1909: Solution was compromise: use long vowels generally: e.g. *hääl, jääma, päästma; hüüdma, nüüd, tüütama; võõras, mõõk, rõõm*, but <u>*ea*</u> in <u>*hea, pea, teadma, seadma, seal, teal]*</u>

- **1908 –** a literary movement Young Estonia, „We will use ää every time!"

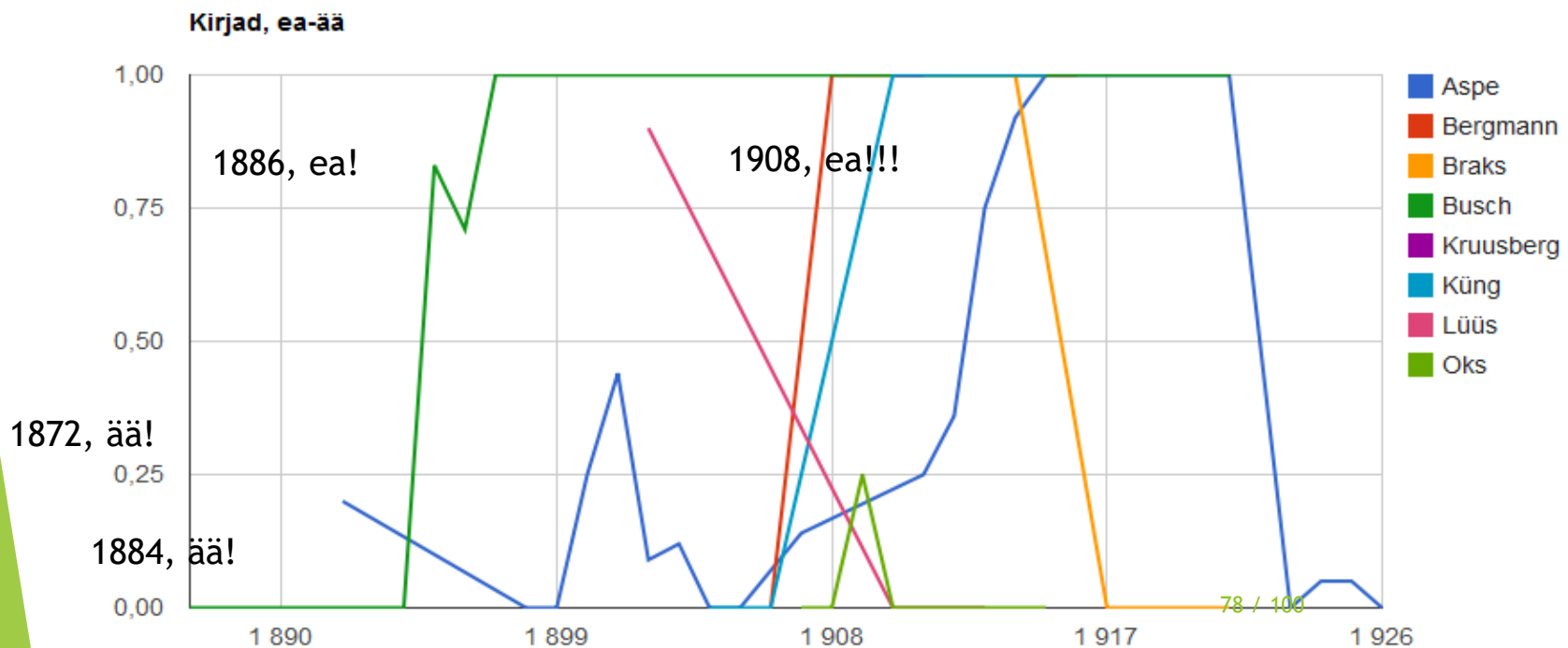- 1953 *ää* is removed from dictionaries, still in use nowadays as part archaism

Dialectal background and writing

▶ ea-ää writing in the letter corpus of rural writers
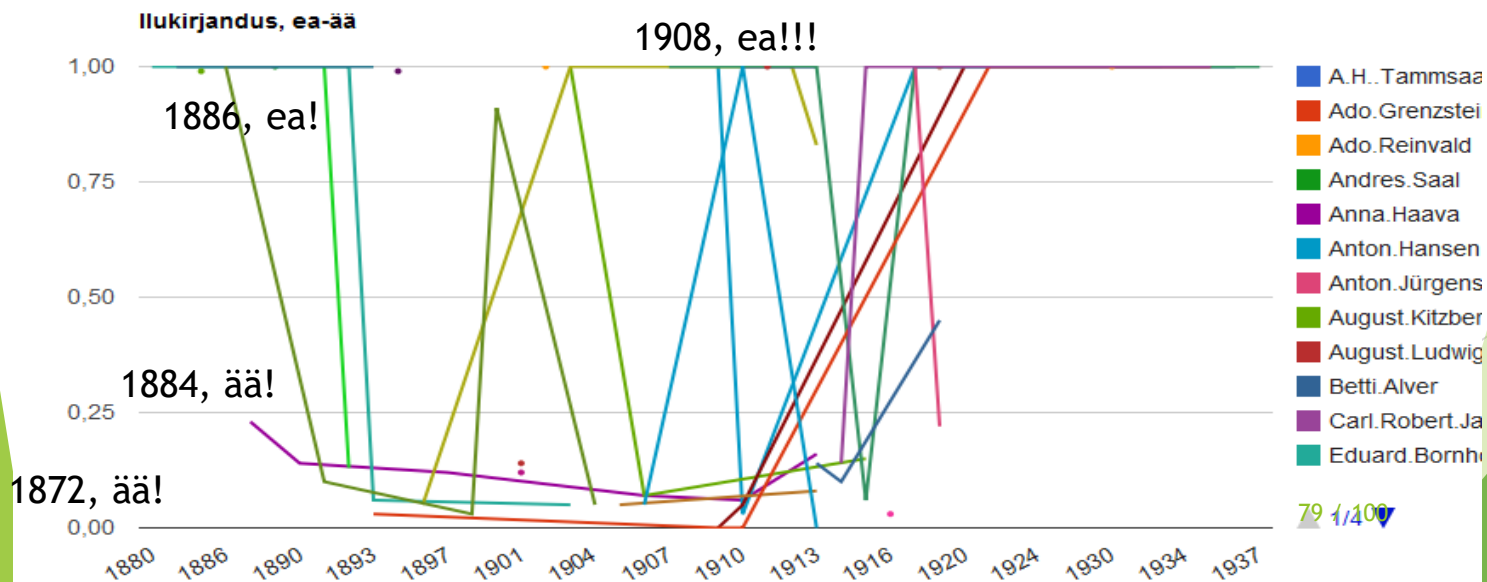 – light green „ää", dark green „ea".

ea–ää Hurda kirjades
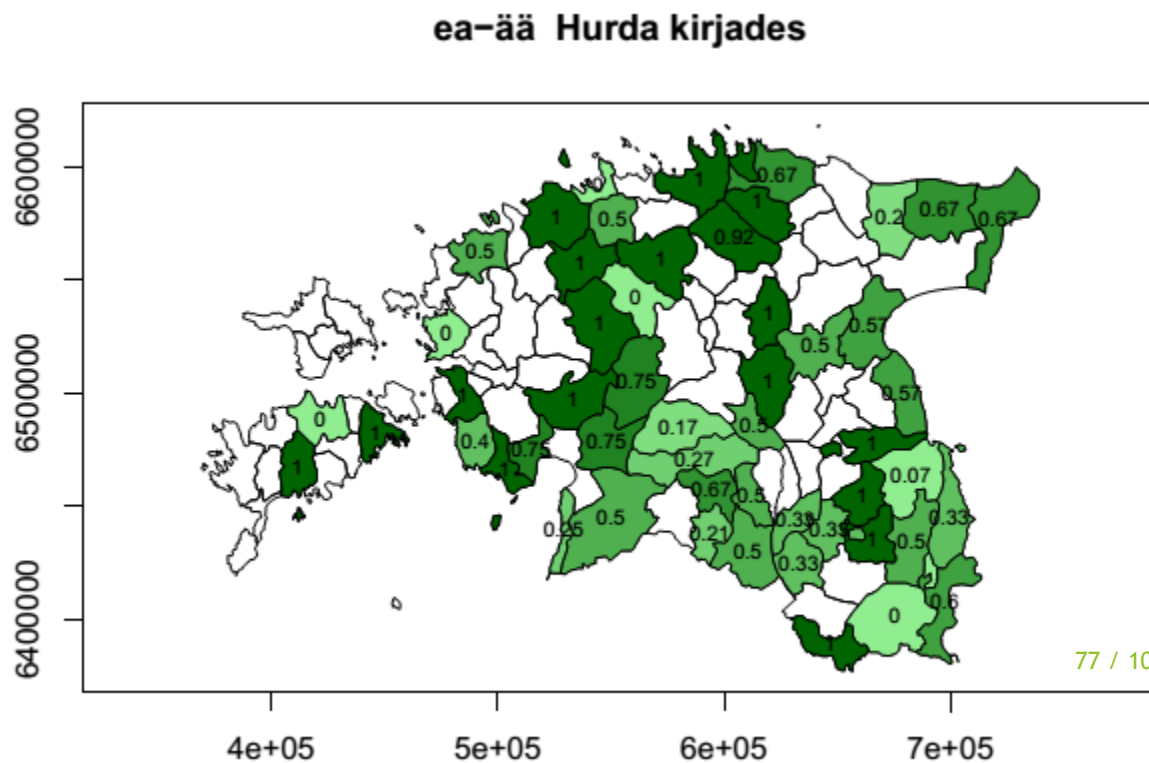
ea-ää, individual variation in letters

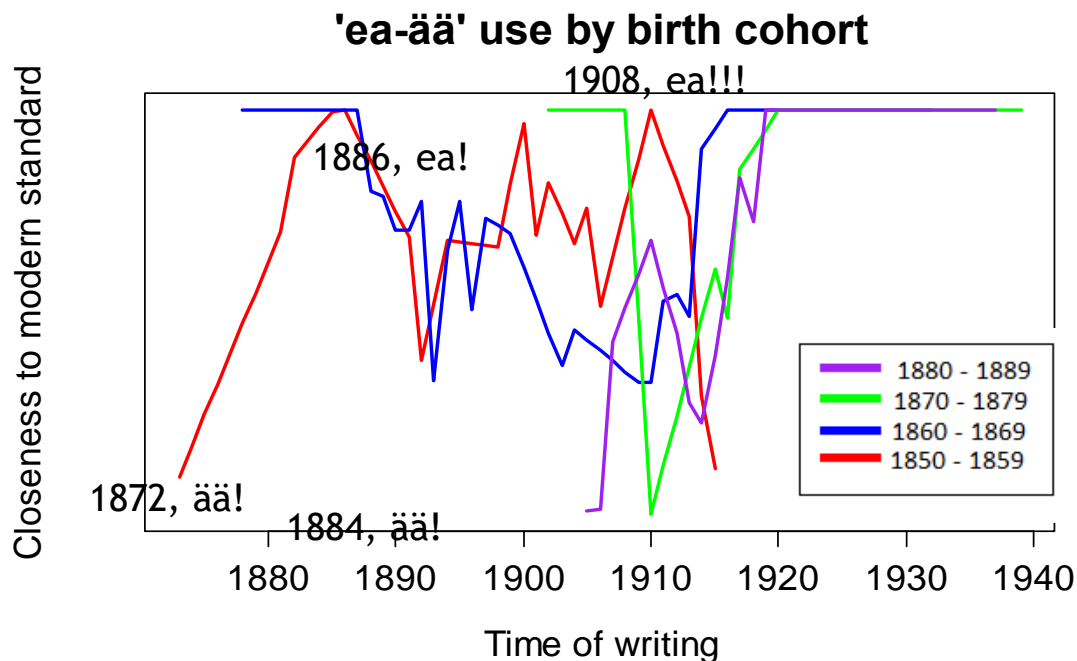▶ Letter corpus shows some change in time but is fairly consistent

Kirjad, ea-ää

ea-ää, individual variation in pub-s

▶ Printed publications shows a lot of variation particulary in time, inconsistent at first with a sharp conclusion

Ilukirjandus, ea-ää

1908, ea!!!

1886, ea!

1884, ää!

1872, ää!

Legend:
- A.H..Tammsaa
- Ado.Grenzstei
- Ado.Reinvald
- Andres.Saal
- Anna.Haava
- Anton.Hansen
- Anton.Jürgens
- August.Kitzber
- August.Ludwig
- Betti.Alver
- Carl.Robert.Ja
- Eduard.Bornho

ea-ää, by time of birth

▶ Birth cohorts show the conclusion, but the differences seem small.

ea-ää, by education

▶ Village schools slightly lower,
~~uni~~
~~inc~~
~~aut~~

'ea-ää' use by education cohort

1886, ea!   1908, ea!!!

1872, ää!   1884, ää!

Closeness to modern standard

Time of writing

Legend:
- University (black)
- Higher school (purple)
- District school (green)
- County school (blue)
- Village school (red)

X-axis: 1880, 1890, 1900, 1910, 1920, 1930, 1940
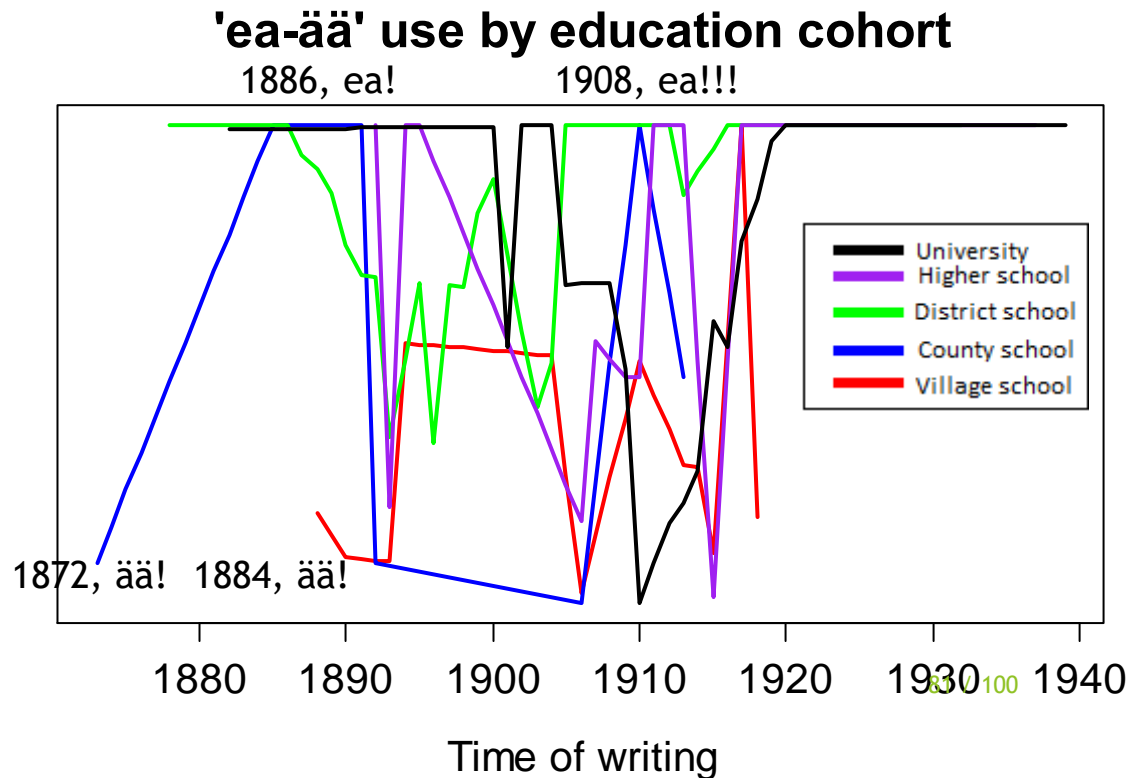
ea-ää, by chosen domicile

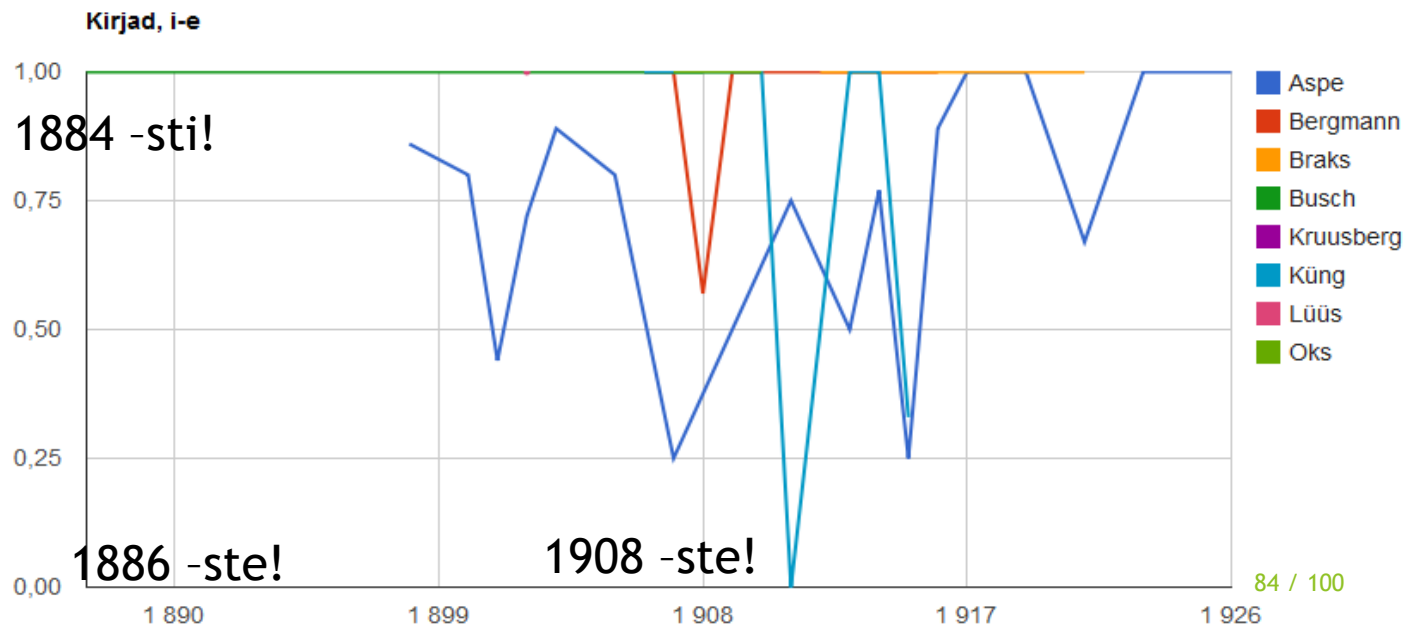▶ Large cities seem to be leading the *ea* variant

'ea-ää' use by domicile
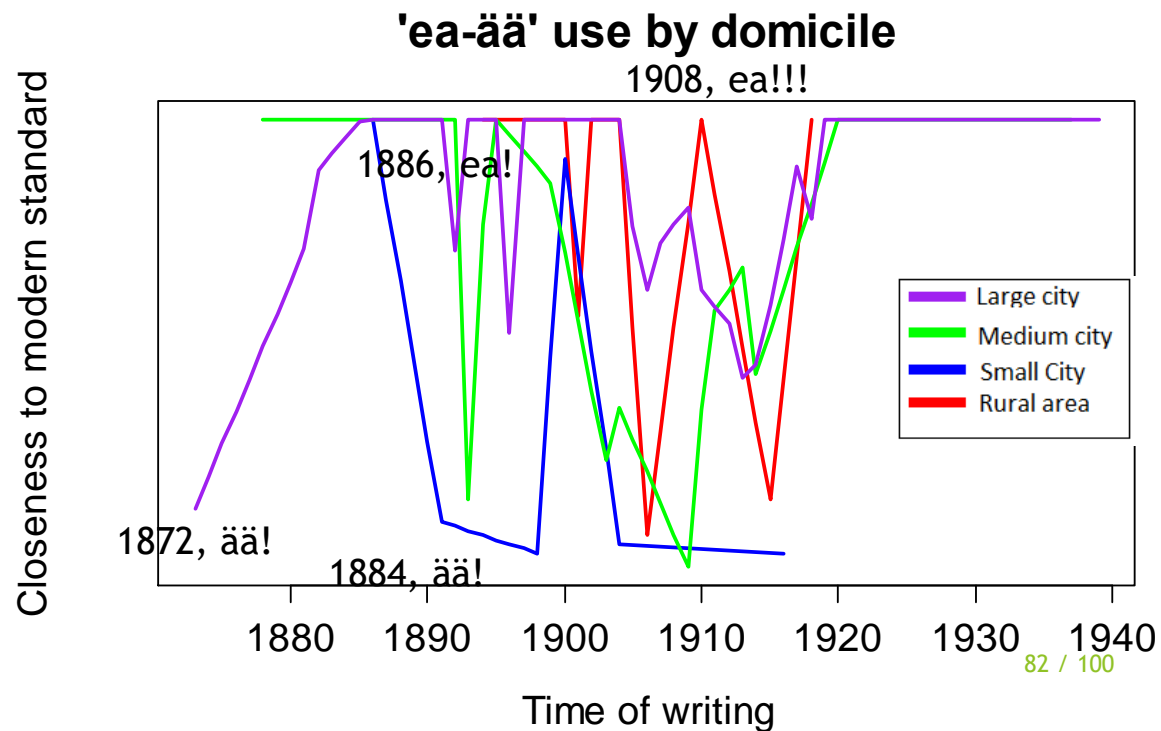
i-e variable

- -sti - -ste variable; 19 lemmas in total; for example täieste/i, õiete/i, vististe/i, lahkeste/i, uueste/i

- Known topic of polemic

- **1884 – Use –*sti!*** [K.A. Hermann recommends in his influential grammar – oma põltsamaa murde ja soome keele põhjal…; also Hans Einer recommends it so in 1885]

- **1886 – Use –*ste* in adverbs; use –*sti* in *õieti*, *kõigiti*!** [Committee on bible publication]

- **1908 – Use –*ste* in adverbs!** [First language conference]

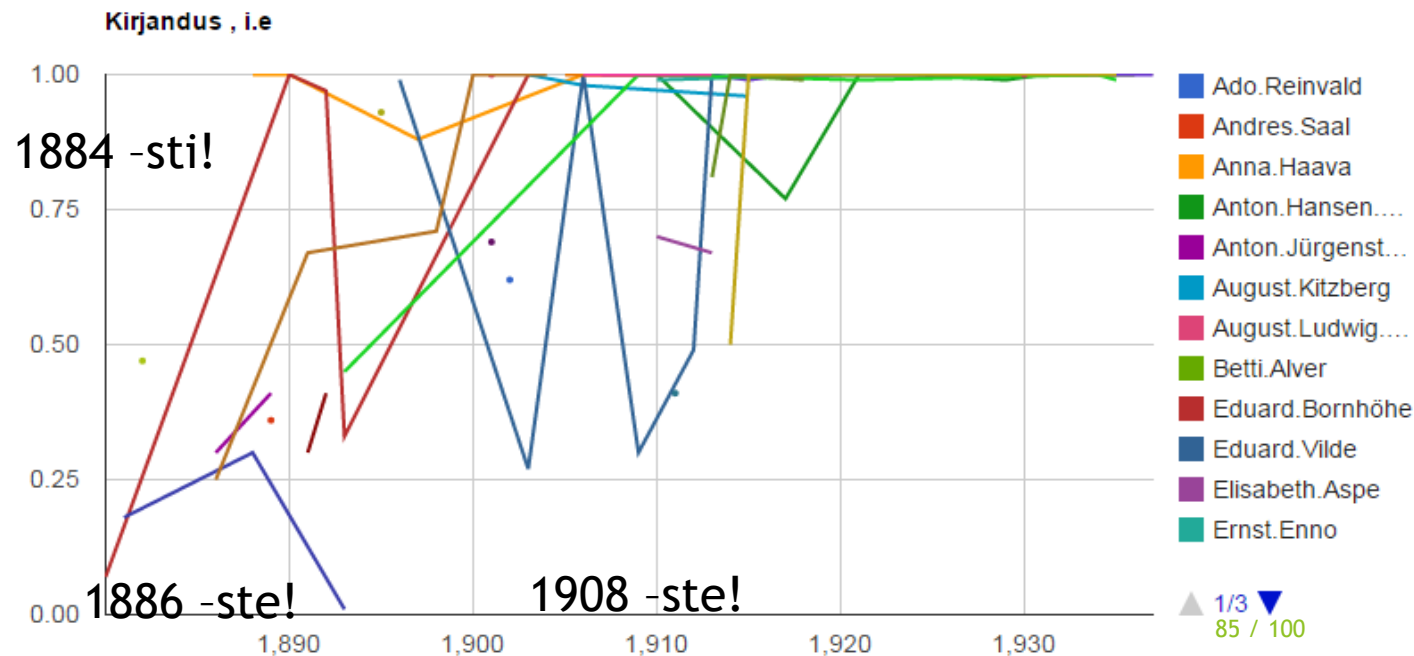# i-e, individual variation in letters

▶ Authors seem fairly consistent, *-sti* seems standar to begin with

Kirjad, i-e

1884 –sti!

1886 –ste!

1908 –ste!

Legend:
- Aspe
- Bergmann
- Braks
- Busch
- Kruusberg
- Küng
- Lüüs
- Oks

i-e, individual variation in pub-s

▶ Some variation and inconsistency, quick resolution

Kirjandus , i.e

1884 –sti!

1886 –ste!   1908 –ste!

Legend:
- Ado.Reinvald
- Andres.Saal
- Anna.Haava
- Anton.Hansen....
- Anton.Jürgenst...
- August.Kitzberg
- August.Ludwig....
- Betti.Alver
- Eduard.Bornhöhe
- Eduard.Vilde
- Elisabeth.Aspe
- Ernst.Enno

1/3 ▼
85 / 100

i-e, by time of birth

▶ Seems a rather smooth transition for all

'i-e' use by birth cohort
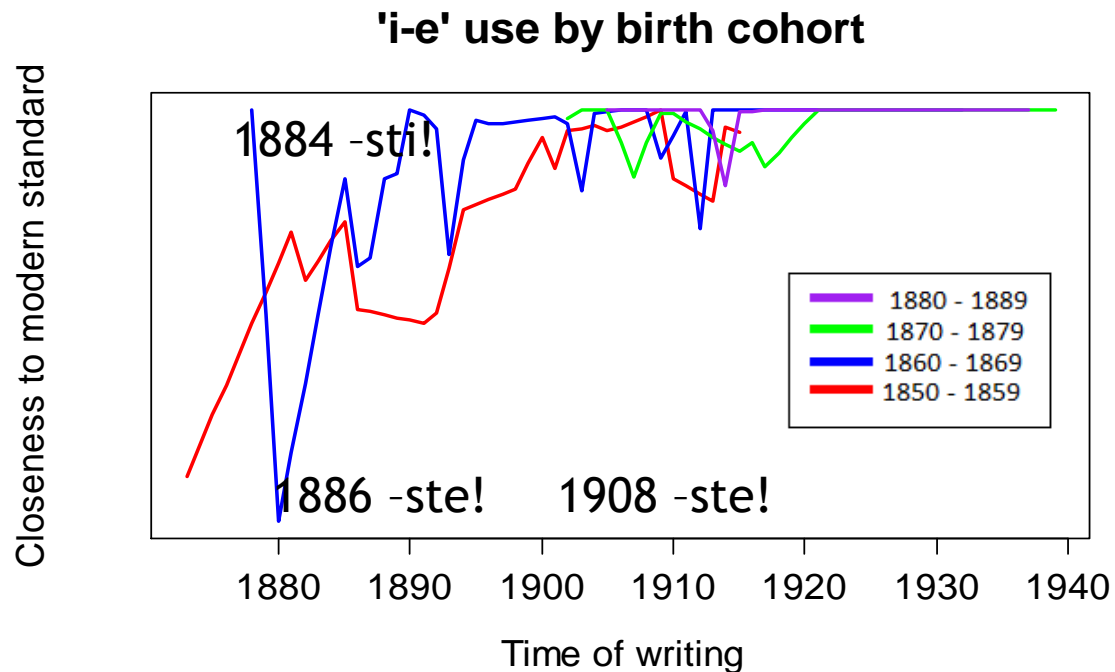
i-e, by education

▶ Seems smooth transition

'i-e' use by education cohort

Figure with y-axis "Closeness to modern standard" and x-axis "Time of writing" (1880–1940). Annotations: "1884 –sti!", "1886 –ste!", "1908 –ste!". Legend: University (black), Higher school (purple), District school (green), County school (blue), Village school (red).
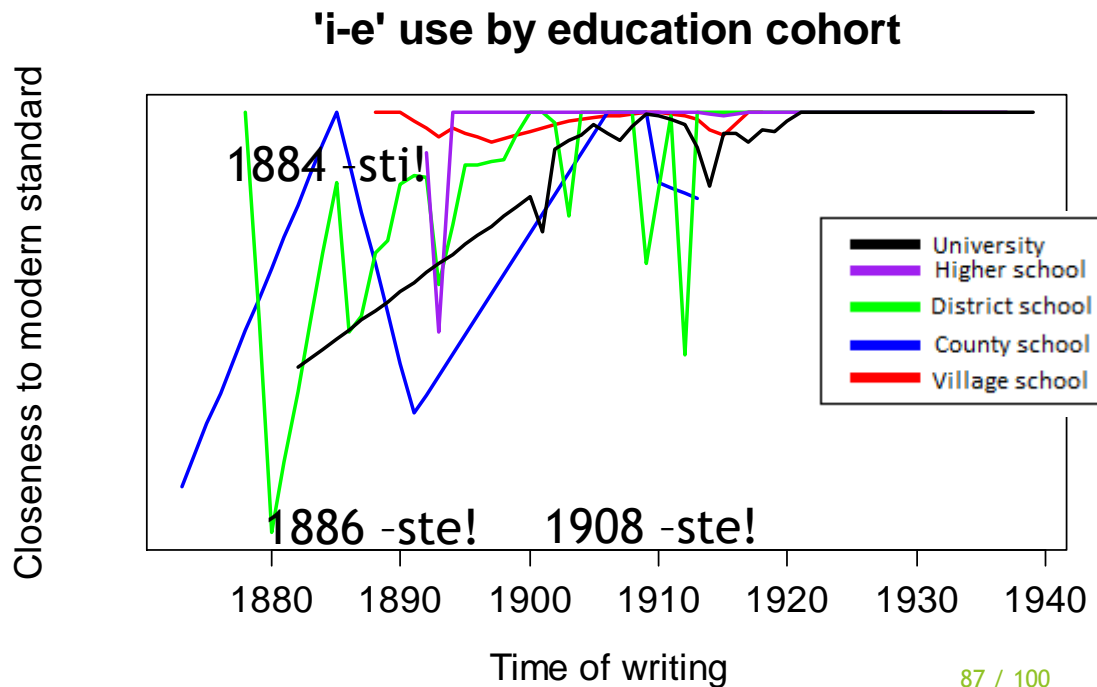
i-e, by chosen domicile
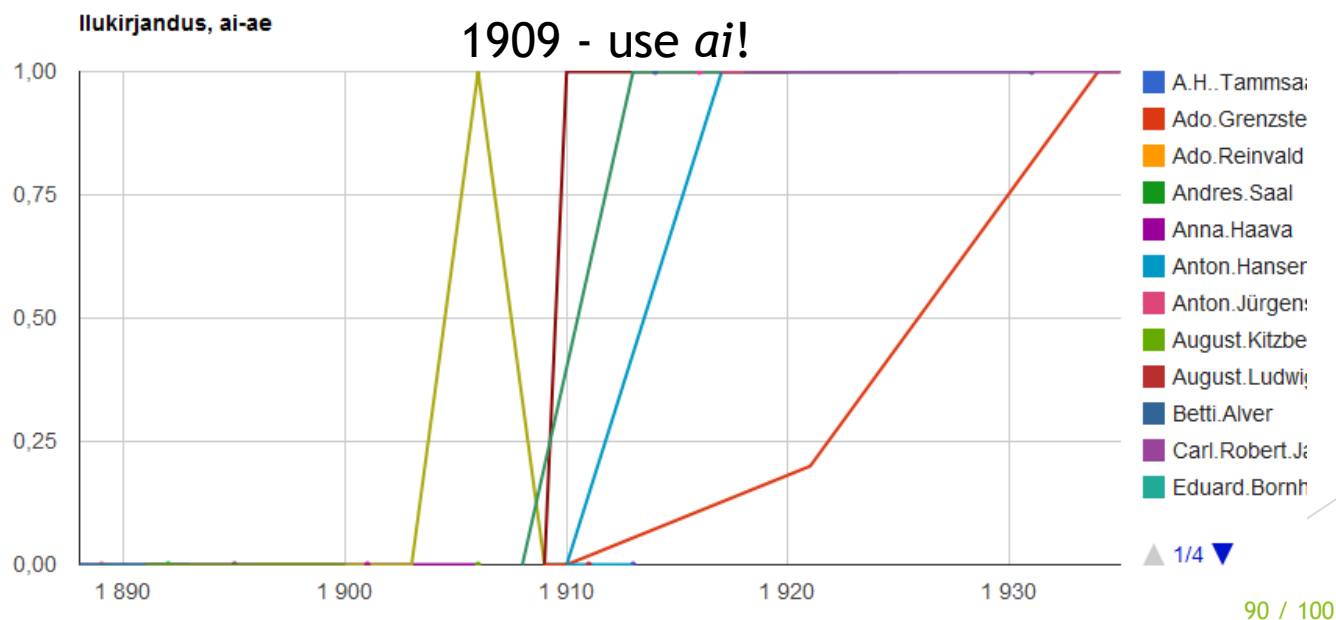
▶ Seems a smooth transition

'i-e' use by domicile
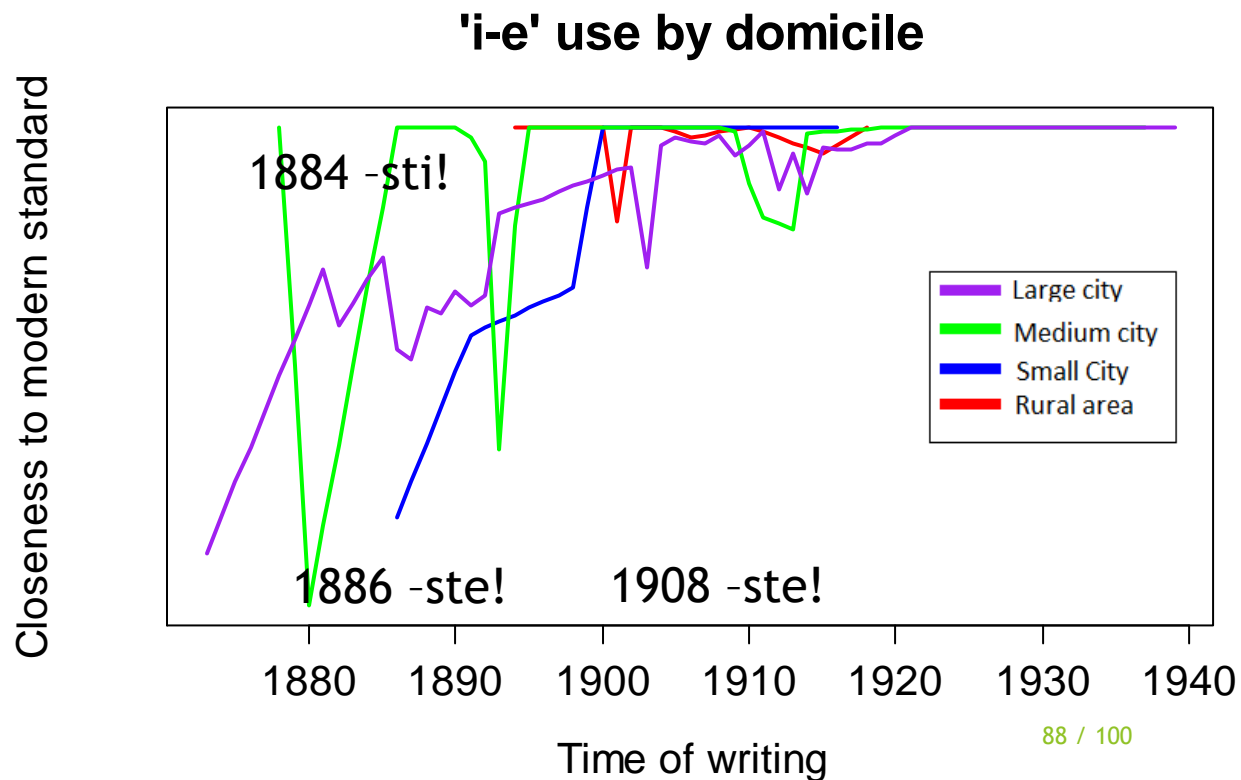
ai-ae variable

- lemmas: naine-naene, laine-laene

- Has some dialectal background, but was not really in focus

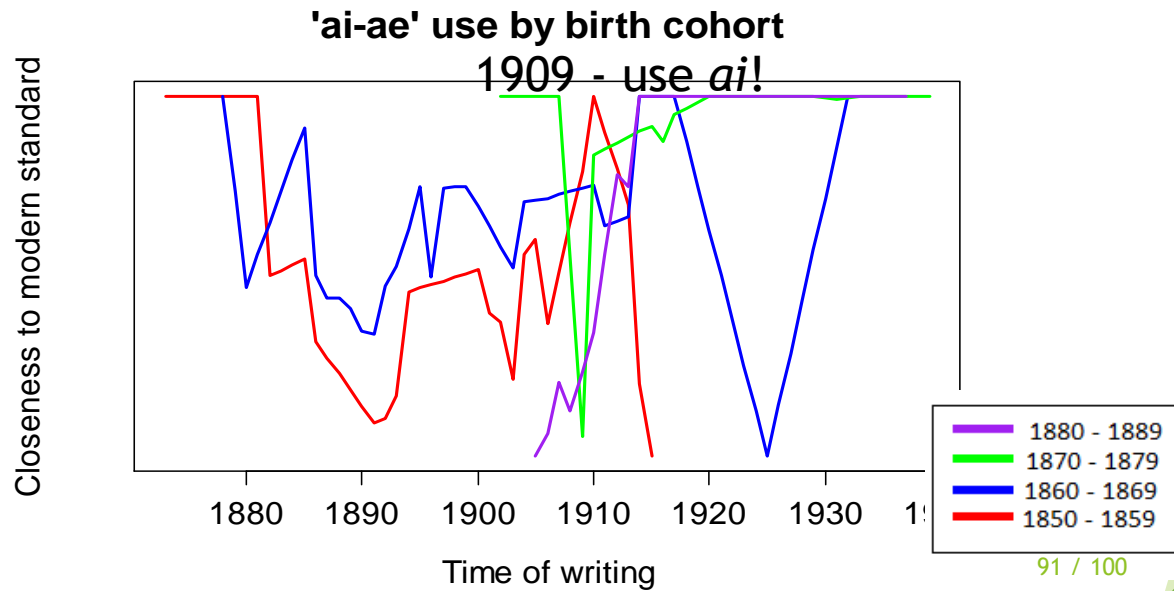- **1909 – Use *naine & laine*** [Second language conference 1909]

ai-ae, indidual variation in pub-s

▶ Almost everyone is consistent throughout

▶ change happens once

Ilukirjandus, ai-ae

1909 - use *ai*!

ai-ae, by time of birth

▶ Much interpersonal variation 1880-1914

▶ Crucial period seems 1910-1914

'ai-ae' use by birth cohort

1909 - use *ai*!

Closeness to modern standard

Time of writing

—	1880 - 1889
—	1870 - 1879
—	1860 - 1869
—	1850 - 1859

ai-ae, by education

▶ Some lag visible in lower education levels

'ai-ae' use by education cohort

Closeness to modern standard — Time of writing

1909 - use *ai*!

Legend:
- University (black)
- Higher school (purple)
- District school (green)
- County school (blue)
- Village school (red)
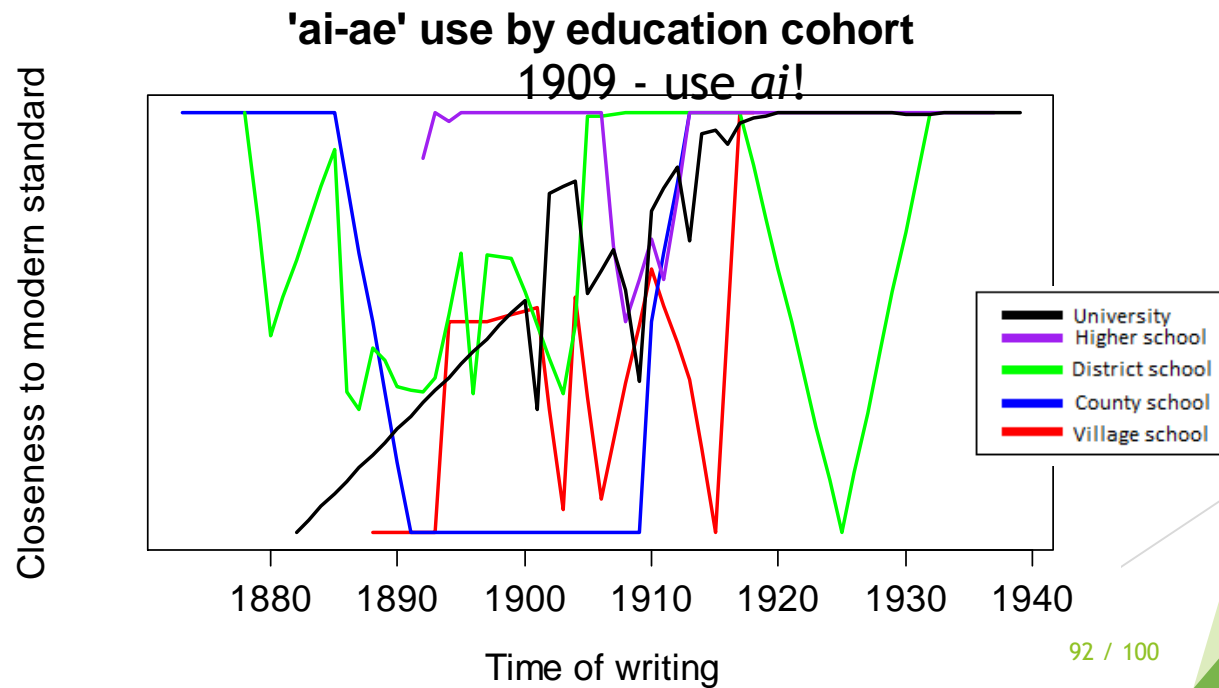
ai-ae, by chosen domicile

▶ Rural areas seem somewhat slower in use, inconsistency may be due to varying writers.

'ai-ae' use by domicile

1909 - use *ai*!

Closeness to modern standard (y-axis)

Time of writing (x-axis): 1880, 1890, 1900, 1910, 1920, 1930, 1940

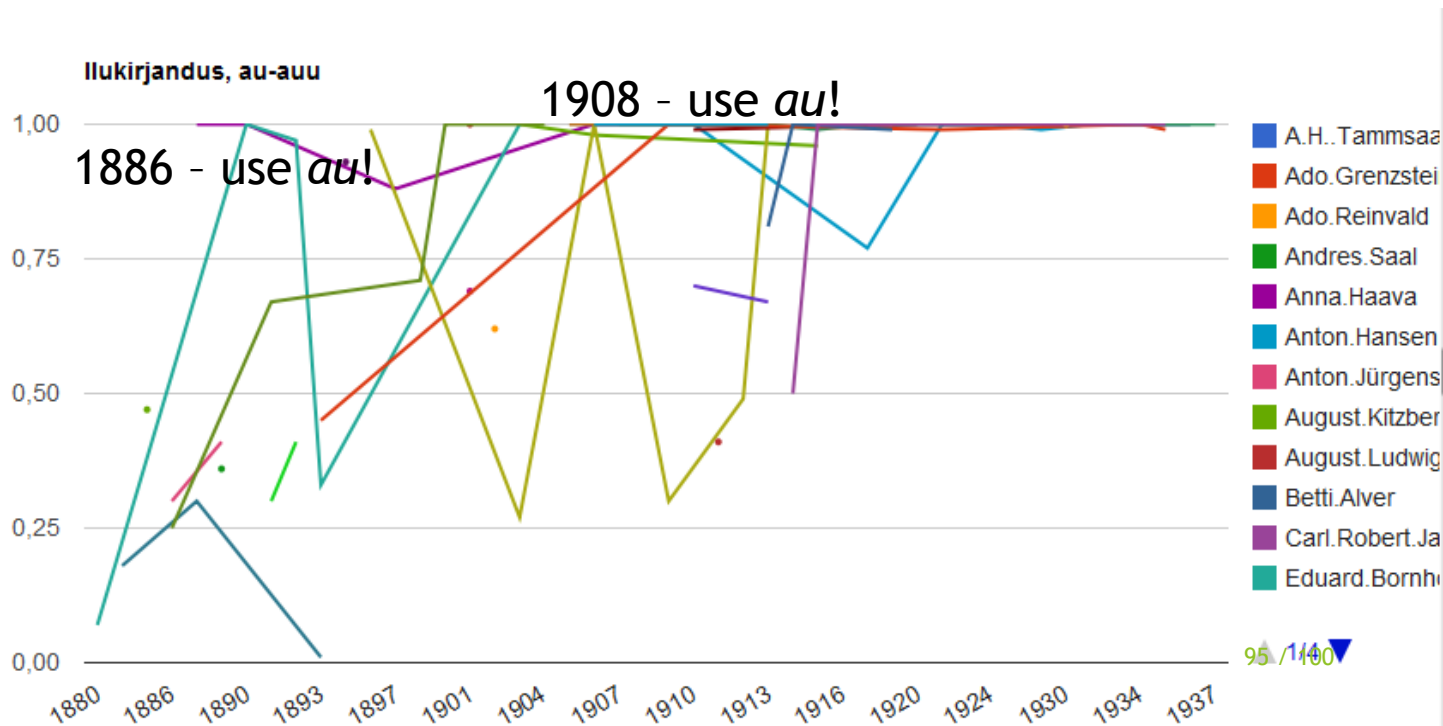Legend: ...ge city, ...dium city, ...all City, ...ral area
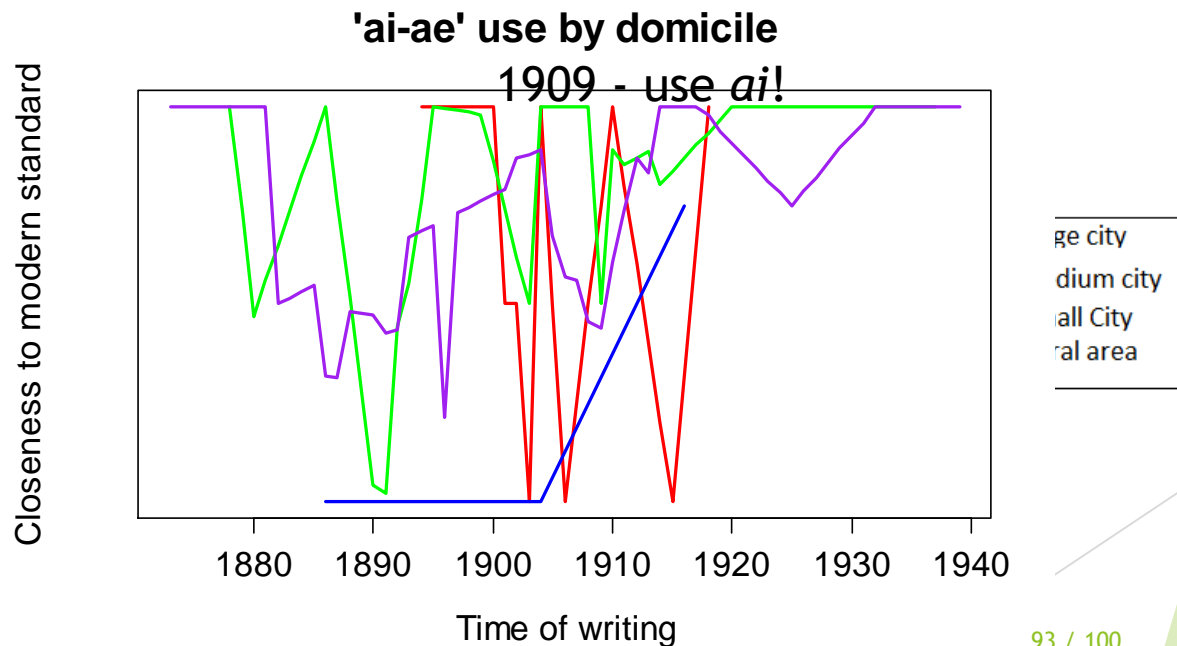
au-auu variable

- Here included as a single lemma: *auu* was a mostly archaic pronunciation *auvu* stored in the 'old written system of language; *auu* was standard during the 19th century

- Under discussion in some events of prescription, 'new writing system'

- **1886 – Use *au, nõu* in nominative; *auu, nõu* in other cases [New bible publication committee]**

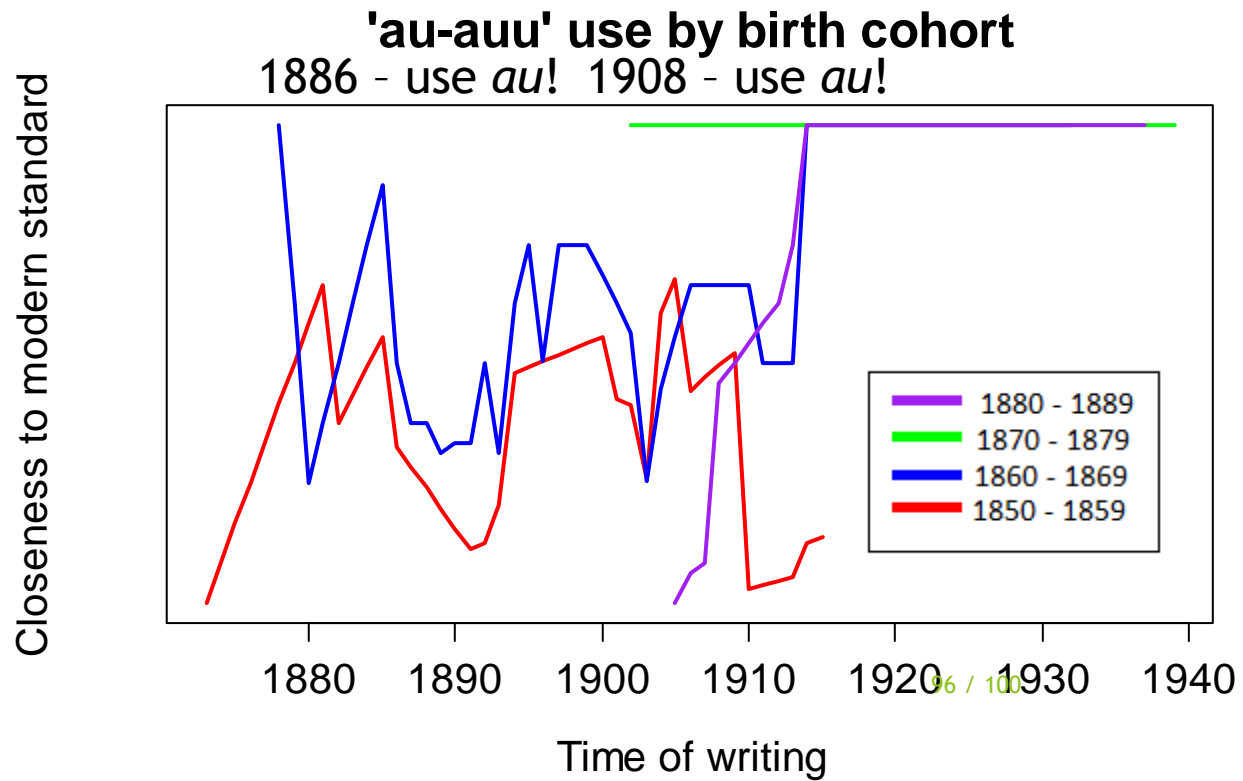- **1908 – use *au, nõu* in all case forms [First language conference]**

au-auu, individual variation in pub-s

▶ Some temporal inconsistency over individuals, not a lot, the variable is much rarer though.

Ilukirjandus, au-auu

1908 – use *au*!

1886 – use *au*!

Legend:
- A.H..Tammsaa
- Ado.Grenzstei
- Ado.Reinvald
- Andres.Saal
- Anna.Haava
- Anton.Hansen
- Anton.Jürgens
- August.Kitzber
- August.Ludwig
- Betti.Alver
- Carl.Robert.Ja
- Eduard.Bornh

95 / 100

au-auu, by time of birth

► Quick change around 1914, new generations did lead the change slightly?

'au-auu' use by birth cohort
1886 – use *au*! 1908 – use *au*!

Closeness to modern standard (y-axis)

Time of writing (x-axis)
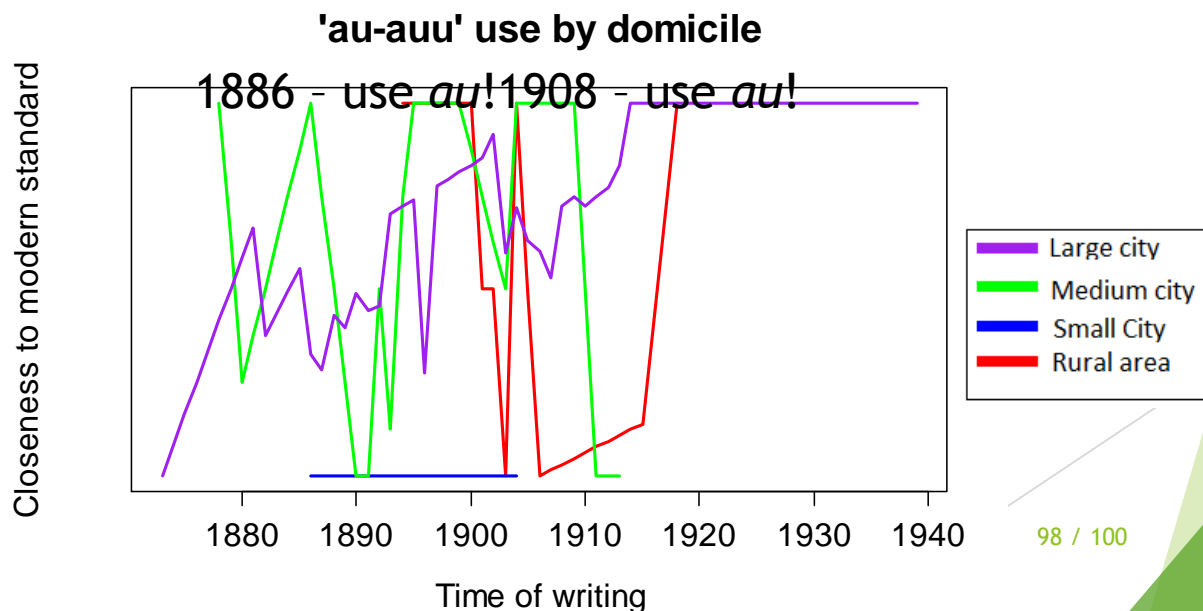
Legend:
- 1880 – 1889
- 1870 – 1879
- 1860 – 1869
- 1850 – 1859

au-auu, by education

▶ University/higher education is the quickest, village and county is quite slow at this.

'au-auu' use by education cohort
1886 – use *au*! 1908 – use *au*!
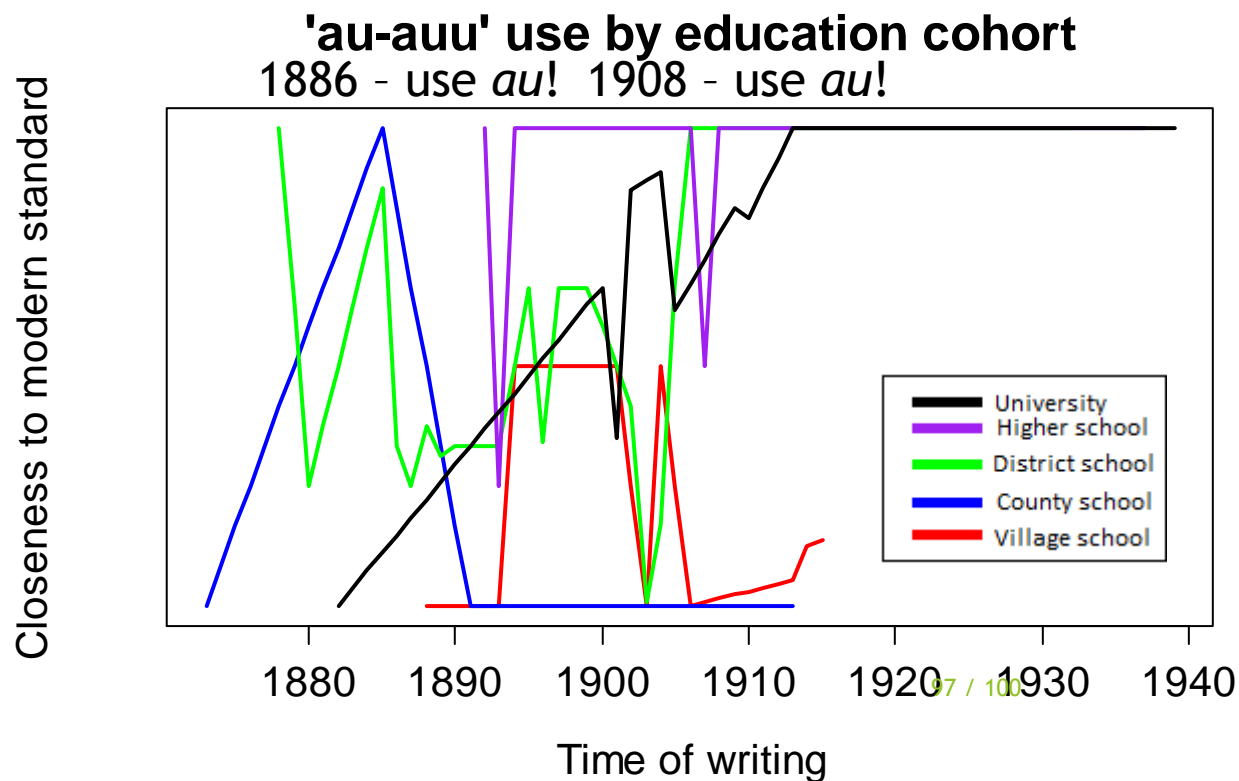
au-auu, by chosen domicile

▶ Here the domicile seems to noticably impact the average use of the new variant, but it may just be the corpus.

'au-auu' use by domicile

1886 - use *au*!1908 - use *au*!

Closeness to modern standard

Time of writing

Large city
Medium city
Small City
Rural area

1880   1890   1900   1910   1920   1930   1940

Estonian context in 1880-1920

- Standard language culture in transformation
  - emerges in time
  - different in some details
- The given variables show some trends in the pathways of variables through larger cities and higher education, but not very clear

Thank you!

- I didn't think we would get this far. ☺