

Becoming a state language

Finnish public debate and modal grammar 1820–1917

Antti Kanner, Hege Roivainen, Tuuli Tahko, Jani Marjanen

Data

- Finnish-language newspapers published in Finland 1820–1917
- 5.2 billion words in Finnish
- Available through the National Library of Finland & FIN-CLARIN
- Includes morphological, syntactic and NER annotations

Markers

- A complex historical process is studied with linguistic markers that are robust to OCR noise and variation typical for historical corpora
- **Modal adverbs** and **modal verbs** are suitable targets, as they form a closed set and have limited morphology
- From these, **epistemic** and **evidential expressions** have a direct connection to pragmatic functions in emerging domains of discourse

Modal expressions

- 19 most frequent epistemic and evidential modal adverbs, including the Finnish equivalents of e.g. *presumably, allegedly, maybe, certainly* etc.
- Four modal verbs (*can, may, shall, must*) with stable epistemic use in Finnish

Methods

Frequency counts are measured against:

- metadata categories
- frequency counts obtained from syntactically comparable categories

Contact us!

- antti.kanner@helsinki.fi
- Project: *Vernacularization and Nation Building: Historical, Linguistic and Computational Perspectives* (2019–2021) at the University of Helsinki



IMAGE SOURCE: NATIONAL LIBRARY OF FINLAND

From vernacular to standard

In the early 1800s, Finnish as a written language was underdeveloped and underprivileged. By the early 1900s, through active promotion, Finnish had become a language of politics, culture and administration alongside Swedish.

This is an example of *vernacularization*, a process in which local, mostly spoken languages are standardized and given a more prominent status in society. It happened in several places in Europe in the early modern and modern periods.

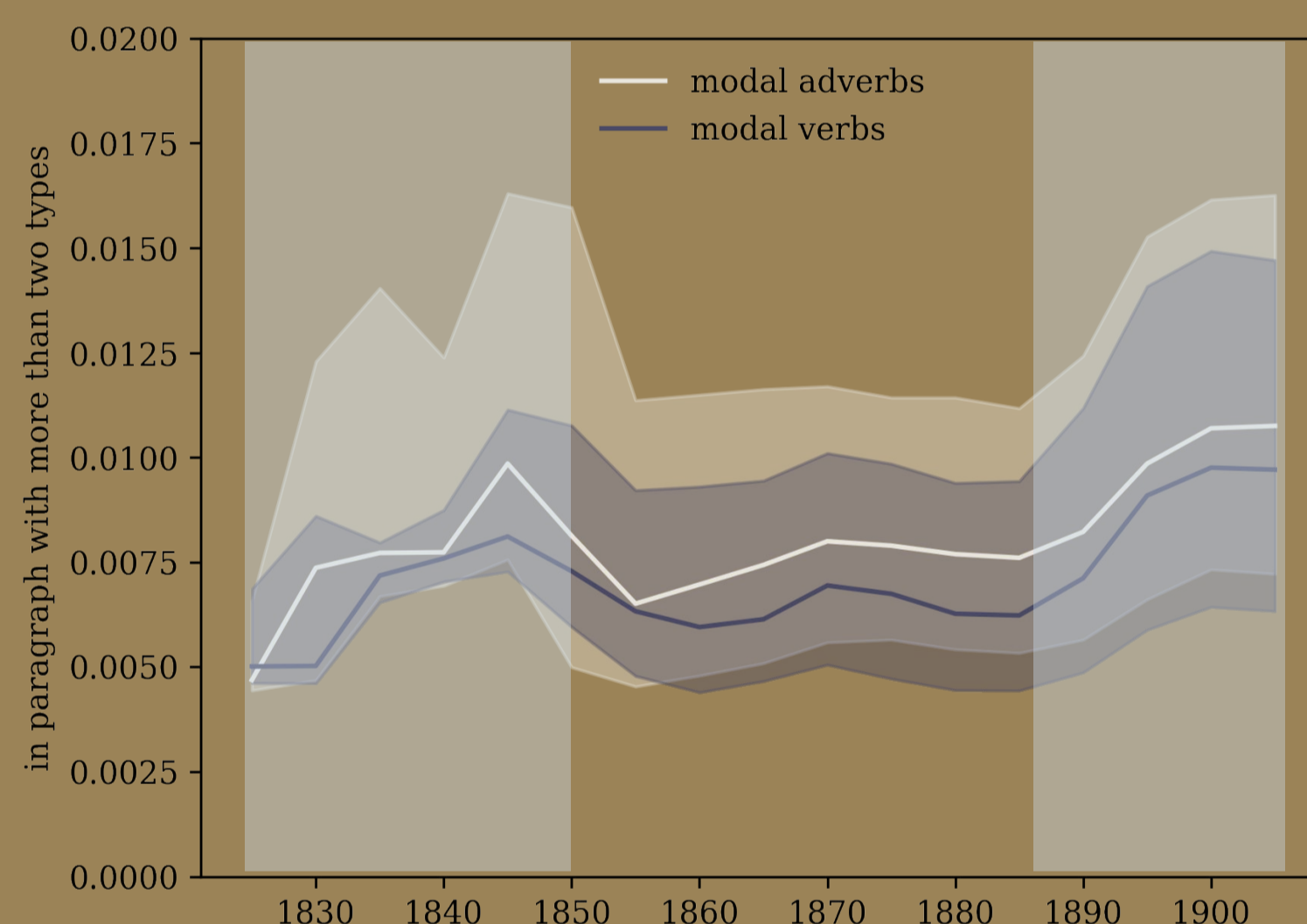
Vernacularization was related to radical shifts in how different languages were perceived and used. In the case of Finland, Finnish became a state-bearing language.

More nuanced expressions were needed

As Finnish became the language of politics and cultural criticism, authors needed more nuanced literary rhetorical conventions to address a new reading public. We hypothesize that the strengthened status of Finnish, and the expansion of its domains of use, is reflected in changes in Finnish modal grammar.

To explore this, we measured modal expressions and analysed their usage in a relatively large corpus of Finnish newspapers.

Modal expressions correlate with expanding areas of language use



The chart shows the type count of modal expressions compared to paragraph word counts. Only paragraphs with two or more modal expressions were included.

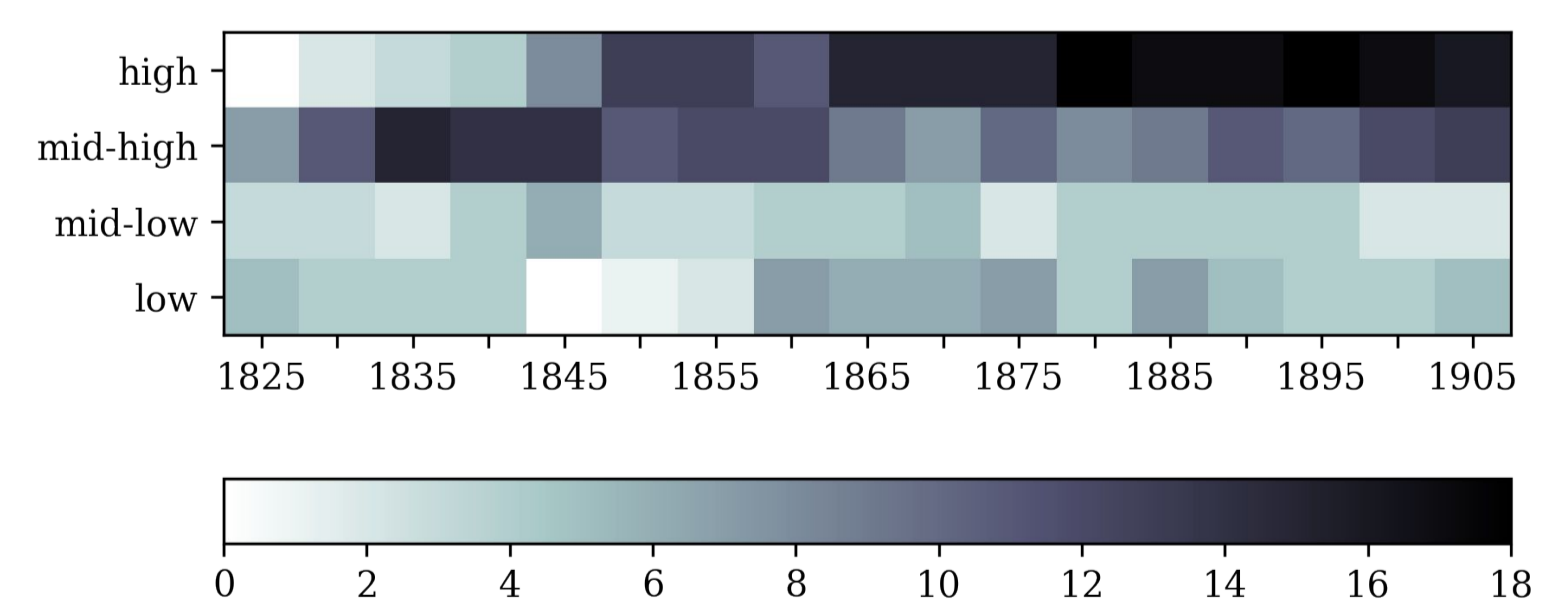
There are two periods of rapid rise in the density of modal expressions: one immediately preceding the wide censorship acts in the 1850s and another one from the mid-1880s onwards.

The first period corresponds with early enthusiasm in the Finnish-language press, the second with full, wide-scale expansion of the use of Finnish as a vehicle of written communication in a rapidly modernizing society.

Expansion of modal lexicon

We grouped the studied modal adverbs into four categories according to their frequency. The darker the color, the more epistemic/evidential adverb types there are in that frequency band.

In 1825, there were hardly any highly frequent modal adverbs. By 1865, the selection of the most often used adverbs had grown. They became a more prominent linguistic feature, manifesting growth in the expressive potential of Finnish.



Politics and language change

Our hypothesis of an increase in modal expressions seems to hold on a general level.

However, this development was less straightforward than expected due to political circumstances. The censorship decree of 1850 hampered reporting on politics and culture. This also reversed some of the earlier diversification in the use of modal expressions.

Tighter censorship in the 1890s and the abolishment of censorship in 1905 did not have similar effects. By this time written Finnish was established enough to maintain these expressions.

Further study

Our analyses indicate that the normalized frequency of modal expressions was higher for journals than newspapers. This supports our hypothesis, but requires further comparison of curated sub-corpora.

If the growth of modal expressions is a feature that reflects the writers' increasing need to take into account an abstract notion of the general public, we should be able to detect this also in other languages with heavy expansion in the nineteenth century, such as Estonian, Latvian, Lithuanian and Norwegian nynorsk.