

# Becoming a State Language: Finnish Public Debate and Modal Grammar 1820–1917

Antti Kanner, Tuuli Tahko, Jani Marjanen

University of Helsinki  
antti.kanner@helsinki.fi

**Abstract.** This paper explores the development of Finnish into a standardized language of politics, science and culture in the nineteenth century. For contemporaries, this meant that Finnish could be regarded as a language that supported Finland as a state. We assumed that this expansion of the domains of use of written Finnish would have necessitated the development of more nuanced ways of expressing opinions and attitudes. We studied this by charting the overall frequency of modal expressions as well as the share of epistemic/evidential adverb types among modal adverbs. We found that the share of modal expressions increased in conjunction with the expansion of the Finnish-language press in the 1840s and again in the period after 1880. More importantly, we found that the amount of very frequently used epistemic/evidential adverb types increased in the mid-1800s, meaning that on average, Finnish newspaper writers had more means for nuanced expression in the latter half of the century.

**Keywords:** Historical Linguistics, Modal Grammar, Standardization, State Building, Vernacularization.

## 1 Introduction: From Vernacular to Standard

In the early 1800s, Finnish as a written language was underdeveloped and underprivileged. By the early 1900s, through active promotion and development of written standards, Finnish had become a language of politics, science, culture and administration alongside Swedish (Engman 2016; Huumo, Laitinen & Paloposki 2004). This is an example of *vernacularization*, a process in which vernacular languages or language variants shift in status and usage in society. Vernacularization is typically linked to language standardization when vernaculars – local, mainly spoken languages with relatively low status – start to be used in a written form for literary purposes (Pollock 2006; Burke 2004), although it can also signify a kind of destandardization when low-prestige forms of language gain acceptability in society (Coupland 2016). The process and its connections to societal conditions vary across time and space – the relationship between Sanskrit and local languages in South Asia in the first and early second millennium or Latin and romance languages in late medieval and early modern period portray different dynamics than the creation of national languages in the nineteenth century or the revitalization of minority languages in the present. Nevertheless, vernacularization is always related to shifts in how co-existing languages are perceived and used. In Europe,

the early modern and modern periods entailed a transformation in linguistic geography where diglossic systems (Ferguson 1959; Fishman 1967; Hudson 2002) started to erode and local vernaculars were adopted and promoted as *national languages* that should be used in administration, law, politics, learning and the arts. In Finland, vernacularization happened in two steps, first with Swedish being adopted more and more as a language of administration and science in the eighteenth century (Lindberg 2006), and second, in a more powerful and rapid transformation, when Finnish developed into a national language with state and nation bearing functions during the course of the nineteenth century (Huumo, Laitinen & Paloposki 2004).

Vernacularization can be studied from a number of linguistic and historical perspectives ranging from changes in the publication landscape and the transformation of how people valued different languages to the development of the languages themselves. This paper focuses on the latter perspective by analyzing changes in the Finnish language in a diachronic corpus consisting of digitized newspapers from 1820–1917. We argue that as Finnish became the language of politics and cultural criticism, authors needed more nuanced literary rhetorical conventions to address a new reading public. An increase in printed material (Marjanen et al. 2019) happened parallel to an increase in readership (Laitinen, Mikkola & Salmi-Nikander 2013) and participation in the form of letters sent to newspapers (Sorvali 2020). This all led to a more elaborate notion of the public and, concretely, the emergence of a readership that was largely unknown to the author. An early nineteenth-century author could assume the position of a learned educator presenting information to more or less passive recipients. As the readership grew and began to take part in the discussion, language had to develop means to sustain such civil conversation. This was reflected in new vocabulary relating to the public (Pietilä 2006), but it also required new ways of expressing epistemic modality.

We propose that the strengthened status of Finnish, and the expansion of its domains of use, is reflected in changes in Finnish modal grammar. Once Finnish became more readily used in public debate, more nuanced and complicated structures emerged to fulfil the newly arisen rhetorical need to evaluate the certainty and reliability of knowledge. In this study, we focus on the emergence, frequency, and distribution of epistemic modal expressions. We claim that these specific linguistic resources, used to align authors' views with those of their perceived audiences, are robust markers of larger linguistic change that took place when language transcended from mostly agrarian spoken language to a literary and administrative language.

## 2 Materials and Methods

Our data set consists of newspapers published in Finnish between the years 1820 and 1917. These have been digitized and made available for data and text mining by the National Library of Finland, with morphological, syntactic and NER annotations provided by FIN-CLARIN. They consist of 5.2B token words and provide a nearly complete record of newspapers and periodicals in the country published in this period. The newspaper corpus as such cannot be regarded as representative of the Finnish language in general, but it is the best historical corpus available, as newspapers covered a wide

range of topics and recorded new features in language by publishing everything from poetry to reports on political events and reflections on academic texts.

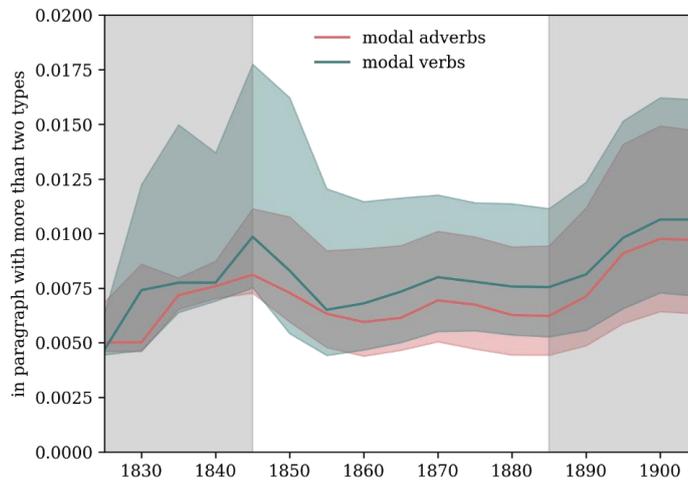
We study a complex historical process with linguistic markers that are robust to OCR noise and variation typical for historical corpora. *Modal adverbs* and *modal verbs* are suitable targets for our purposes because they form a closed set and have limited morphology; furthermore, epistemic and evidential expressions have a direct connection to pragmatic functions in emerging domains of discourse. In this paper, we focus on the 19 most frequent epistemic and evidential modal adverbs, including the Finnish equivalents of e.g. *olettavasti* ('presumably'), *ehkä* ('maybe'), and *varmasti* ('certainly'), as well as four modal verbs – *voida* ('can'), *pitää* ('have to'), *saattaa* ('may'), *täytyy* ('must') – with stable epistemic use in Finnish.

The sources of newly developed expressions fall roughly under three categories: 1) expressions that already existed in the language, but which undergo a semantic or syntactic change; 2) translations and calques from languages that are further ahead in the vernacularization process (in the case of Finnish, most calques are from Swedish) and thus provide models for what kind of expressions are presumably required to fulfil literary, administrative and public functions; 3) previously unused formations based on the language's own resources with no obvious outside models.

In analyzing the development of modal expressions, a key object of interest besides general frequency patterns is determining the composition of modal expressions as a grammatical subsystem. This requires comparing modal expressions against general tendencies in their wider grammatical categories. Further, as we are interested in the emerging distinctions among these expressions, we look at contexts where several of them are used. These contexts are interesting because the distinctions between two expressions with similar meanings are often evoked by their juxtapositions. An increase in the density of these contexts could thus plausibly point towards the emergence of more fine-grained communicative resources. We anticipated that there would be considerable changes in the ecology of epistemic modal expressions, these changes mainly taking the shape of the specialization of functions for a number of expressions. The most robust signals for these changes are perhaps the expressions' relative frequencies, the mapping of which is a relatively trivial task. These markers will be the focus of this paper.

### 3 Results

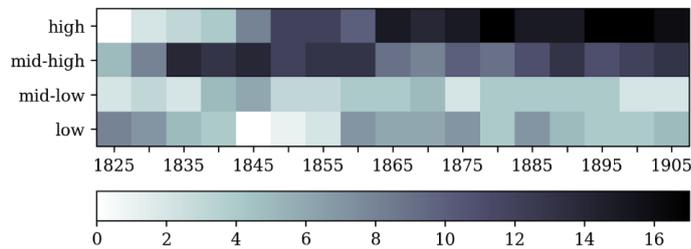
In our first experiment we analyzed the shares of modal expressions in individual paragraphs. The emergence of wider contexts of use for written discourse not only required new grammatical constructions functioning as rhetorical devices, but also the emergence of more definitive functional differences between these constructions. We thus presumed that semantic and pragmatic distinctions between the different modal expressions became entrenched during the period studied. In order to quantitatively analyze contexts where these distinctions could plausibly surface, we looked at paragraphs where at least two different modal expressions were used. From these, we calculated the number of modal expressions relative to the full word count of the paragraph.



**Fig. 1.** Type count of 19 modal expressions compared to paragraph word count, 1820–1910. Only paragraphs with two or more modal expressions were included.

Figure 1 presents the development of relative frequencies of modal verbs (blue) and modal adverbs (red) in time bins of ten years, overlapping by five years. The thick lines mark the mean frequency of each group, and the coloured area the 90% confidence band. The first data point is thus 1820–1829, the second 1825–1834 and so on. The temporal dynamic fits well with the development of publication activity in Finland. While there is an overall rising trend in both groups (a linear regression model shows a positive slope), there are two periods of rapid rise in the density of modal expressions. The first one (grey shadow on the left) corresponds with early enthusiasm for developing the Finnish-language press. This phase is cut short by the strict and wide censorship legislation imposed by the Russian authorities in the 1850s, which caused stagnation in the diversification of the press even after the legislation itself was revoked. The second period (grey shadow on the right) matches the full, wide-scale expansion of the use of Finnish as a vehicle of written communication in a rapidly modernizing society (Marjanen et al. 2019). Thus, the material development of the press measured in printing volumes was paralleled by qualitative changes in linguistic features.

In the second experiment (Fig. 2) we further grouped the studied modal adverbs into four categories (low, mid-low, mid-high, high) according to their frequency. Because the frequencies are subject to changes in corpus composition for each time segment, the categories were based on frequencies of all adverbs, a much larger and more varied group than its modal subgroup.



**Fig. 2.** The amount of modal adverbs per time bin in four different frequency categories derived from all adverbs, 1820–1910.

From the frequency distributions of all adverbs we obtained frequency categories for top 5% (“high”), top 5–50 % (“mid-high”), bottom 5–50 % (“mid-low”) and bottom 5% (“low”). The darker the colour, the more epistemic/evidential adverb types there are in that frequency category. In the 1820s, there were hardly any highly frequent modal adverbs. By the 1860s, the selection of the most often used adverbs had grown (Fig. 2). They became a more prominent linguistic feature, manifesting growth in the expressive potential of Finnish when conveying (un)certainities and stances towards reported events. Contrary to the previous analysis, the increase in epistemic/evidential adverb types in the highest frequency category suggests a permanent change in language use already from the mid-1850s onward. This development seems to be to a lesser degree associated with the material development of newspapers. Overall, the analyses nonetheless point towards a similar development.

## 4 Conclusions and Further Study

Our study traces historical changes in linguistic features and relates them to the development of the public sphere in Finland. Detailed analysis, based on large-scale historical corpora, of when modal expressions have come to be used in written Finnish is a major contribution to the study of the Finnish language.

We predicted an increase and diversification in the use of modal expressions. This seems to hold on a general level. However, the 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 retain these expressions.

We found that changes in the relative frequencies of a selection of the modal expressions increase in conjunction with key turning points in the Finnish press. However, a full analysis, including multivariate statistical tests, is still needed for making this argument in full. This study forms a segment of a more ambitious research programme where we will track modal expressions’ grammatical, contextual and semantic features simultaneously and in correspondence to each other. This requires further comparison and curation of dedicated sub corpora for analysis. The overall approach will be akin

to behaviour profile analysis, where occurrences of the studied linguistic items are examined across a wide range of variables and then subjected to scrutiny by univariate, bivariate and multivariate statistical tests (e.g. Divjak & Gries 2006; Arppe 2008).

We believe that our observations on the Finnish case, which is rather straightforward with a relatively rapid vernacularization process, may lead to discovering similar trends in other languages as well. 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.

## References

1. Arppe, A.: Univariate, bivariate and multivariate methods in corpus-based lexicography. University of Helsinki, Helsinki (2008).
2. Burke, P.: Languages and communities in early modern Europe. Cambridge University Press, Cambridge (2004).
3. Coupland, N.: Labov, vernacularity and sociolinguistic change. *J Sociolinguistics*. 20, 409–430 (2016). <https://doi.org/10.1111/josl.12191>.
4. Divjak, D., Gries, S.T.: Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*. 2, (2006). <https://doi.org/10.1515/CLLT.2006.002>.
5. Engman, M.: Språkfrågan: Finlandssvenskhetens uppkomst 1812–1922. Svenska litteratursällskapet i Finland, Helsingfors (2016).
6. Ferguson, C.A.: Diglossia. *Word*. 15, 325–340 (1959).
7. Fishman, J.A.: Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues*. 23, 29–38 (1967).
8. Hudson, A.: Outline of a theory of diglossia. *International Journal of the Sociology of Language*. 2002, 1–48 (2002). <https://doi.org/10.1515/ijsl.2002.039>.
9. Huumo, K., Laitinen, L., Paloposki, O. eds: Yhteistä kieltä tekemässä: Näkökulmia suomen kirjakielen kehitykseen 1800-luvulla. Suomalaisen Kirjallisuuden Seura, Helsinki (2004).
10. Laitinen, L., Mikkola, K., Salmi-Niklander, K. eds: Kynällä kyntäjät: Kansan kirjallistuminen 1800-luvun Suomessa. Suomalaisen Kirjallisuuden Seura, Helsinki (2013).
11. Lindberg, B.: Den antika skevheten: Politiska ord och begrepp i det tidig-moderna Sverige. Kungl. Vitterhets historie och antikvitets akademien, Stockholm (2006).
12. Marjanen, J., Vaara, V., Kanner, A., Roivainen, H., Mäkelä, E., Lahti, L., Tolonen, M.: A National public sphere? Analyzing the language, location, and form of newspapers in Finland, 1771–1917. *Journal of European Periodical Studies*. 4, 54–77 (2019). <https://doi.org/10.21825/jeps.v4i1.10483>.
13. Pietilä, V.: "Matti Matalaisen" julkea ehdotus ja vähän muutakin. *Tiedotustutkimus*. 29, 41–57 (2006).
14. Pollock, S.I.: The language of the gods in the world of men: Sanskrit, culture, and power in premodern India. University of California Press, Berkeley (2006).
15. Sorvali, S.: "Pyydän nöyrimmästi sijaa seuraavalle" – Yleisönosaston synty, vakiintuminen ja merkitys autonomian ajan Suomen lehdistössä. *Historiallinen Aikakauskirja*. 118, 324–339 (2020).