

Transcriber effects in the Icelandic parliament corpus

Lilja Björk Stefánsdóttir
Anton Karl Ingason

Digital Parliamentary Data in Action (DiPaDA 2024)
28th May 2024





Introduction

- Goal:
 - Understand the effect of transcribers in parliament speeches when measuring style-shift in lifespan change.
- Using the Icelandic Parliament Corpus for sociolinguistics
- Studying sociolinguistic lifespan change with a high-definition method
- The case of MP Ásmundur Einar Daðason
- Implications for ongoing work
- Conclusion and future work



Style shift and lifespan change

- Sociolinguistic style shift:
 - People can speak in a more formal or informal way depending on context and life situation
 - Attention-paid-to-speech, linguistic marketplace etc.
- Marker of style: Icelandic Stylistic Fronting
 - We measure the rate of this word order over the career of Icelandic politicians and relate findings to their personal histories.
- The Icelandic Parliament Corpus has thousands of speeches to study this way.

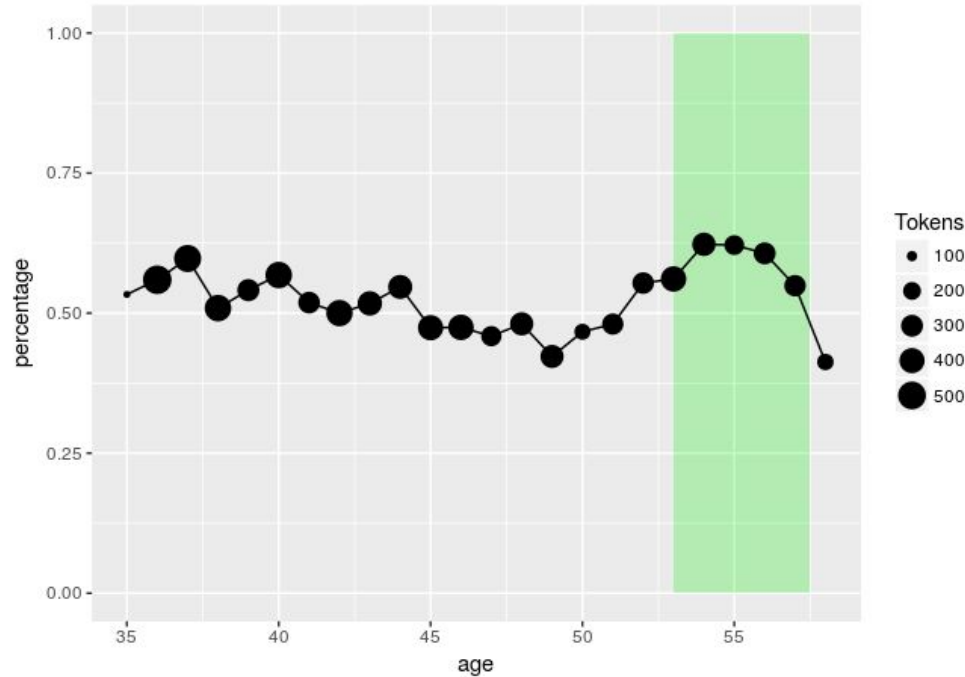


Rate of Stylistic Fronting

- (1) *Bækur* [_{CP} *sem* {*lesnar eru* / *eru lesnar* } *til skemmtunar*] *eru bestar.*
books [_{CP} that {**read** are / are **read** } for entertainment] are best
'Books that are read for entertainment are the best ones.'

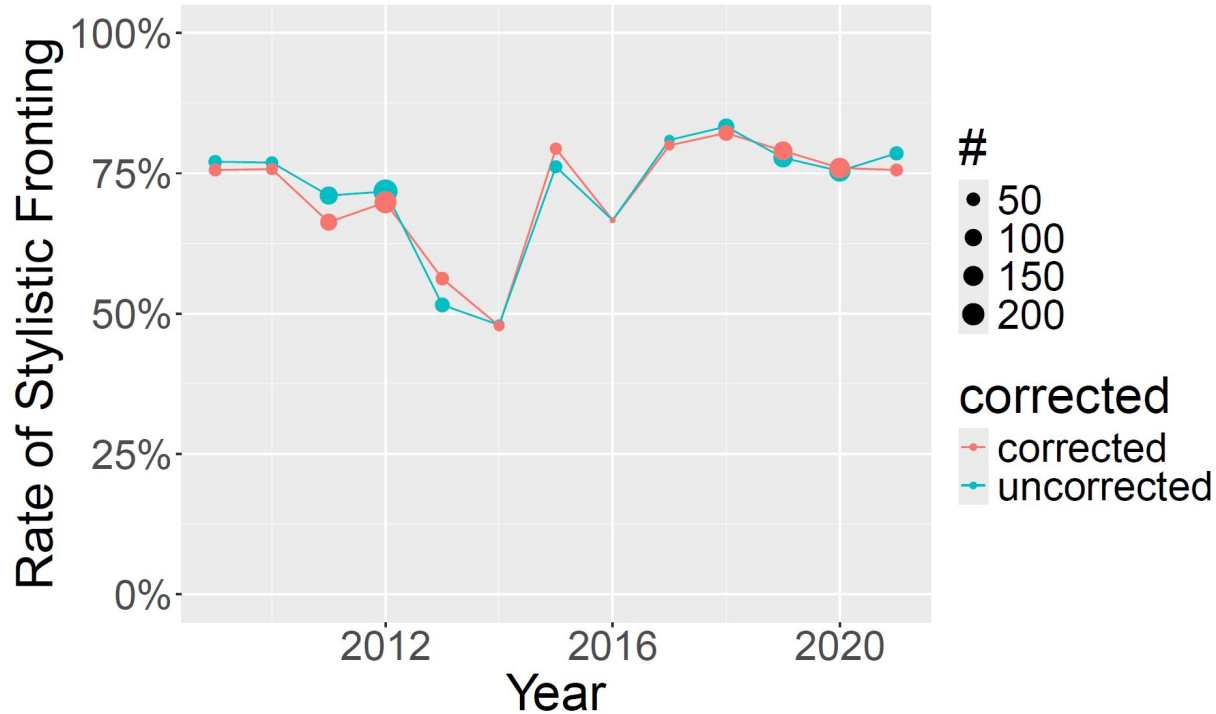
Our high-definition study on Sigfússon

Peak Stylistic Fronting following economic crisis



Einarsson data: Corrected transcription

Not a significantly different rate of SF





Implications for ongoing work

- We continue to study more MPs
- Still checking more MPs manually
- Also running experiments with more annotators
 - Verify inter-annotator agreement
- Error rate is not identical between MPs
- For analysis that must rely on automatic coding, understanding the transcriber error is crucial.



Conclusion

- Automatic methods can be used to study style shift in the Icelandic Parliament corpus but manual checking of data results in even greater accuracy.
- Open access parliament data is important for studying style shift across the lifespan.
- We look forward to reporting on more findings from the project in the future.
 - 1.5M euro ERC starting grant.
 - 5 year project, just got started.