

# Introduktion til Sprogteknologi

Forår 2020  
Lecture 5

# Dagens mål

- Kopier alle corpora og scripts til **work**-mappen
- Bryd en tekst op i sætninger
- Find alle forkortelser
- Find alle proprier
- Find mest frekvente ord i en bestemt kontekst
- Sammenligne Per Langgård's rettelser med outputtet af nuværende analyse

# Nye Kommandoer

- **sed** Stream Editor - ændrer inputtet via regular expressions
- **perl -pe** Fordi **sed** har problemer med nogle tegn, er perl -pe en god erstatning
- **diff** Difference - viser forskelle på 2 filer