

# Introduktion til Sprogteknologi

Forår 2020  
Lecture 3

# Dagens mål

- Læse, og forstå outputtet af diverse led af Kalaallisut-analysatoren
- Benytte analysatoren

# Dokumentation

- Morfologiske analysator
  - <https://giellalt.uit.no/lang/kal/>
  - <https://giellalt.uit.no/lang/kal/root-morphology.html>
- Syntaktiske tags
  - Kalaallisut Syntaktiske Tags.pdf i ~/lt-course/lecture03/ mappen.

# Kalaallisut-analysatoren

1/4

Findes i mappen `~/langtech`, som indeholder mapperne:

- `giella-core` Hjælpefiler fra Giellatekno - må ikke redigeres
- `giella-shared` ... ditto ...
- `kal` Kalaallisut-analysatoren
- `regression` Tests der checker om de ændringer man har lavet i `kal` gør noget uforventet

# Kalaallisut-analysatoren

2/4

Mappen `~/langtech/kal` indeholder mange ting, men vigtigst for jer er mappen `~/langtech/kal/src` som yderligere indeholder disse vigtige mapper:

- **morphology** Morfologisk analysator, lavet med LEXC-filer
- **syntax** Morfologisk disambiguator og syntaktisk tagger, lavet med Constraint Grammar-filer

# Kalaallisut-analysatoren

3/4

Den morfologiske analysator `~/langtech/kal/src/morphology` indeholder bl.a.:

- **root.lexc** Definition af alle tags og andre morfologiske stumper
- **stems** Mappe med rødder og leksikaliserede rod-lignende former
  - `abbreviations.lexc` `nouns.lexc` `particles.lexc` `propernouns.lexc` `acronyms.lexc` `numerals.lexc` `pronouns.lexc` `verbs.lexc`
- **affixes** Mappe med derivationer
  - `derivations-inflections.lexc` `numerals.lexc` `propernouns.lexc` `symbols.lexc`

# Kalaallisut-analysatoren

4/4

Den morfologiske disambiguator og syntaktiske tagger [~/langtech/kal/src/syntax](#) indeholder bl.a.:

- [kal-pre1.cg3](#) 1 regel, som fjerner paragraf-brud efter kolon
- [kal-pre2.cg3](#) 19 regler, som forbereder håndtering af direkte tale
- [disambiguator.cg3](#) 3390 regler, som udfører både morfologisk disambiguering og syntaktisk tagging