

En förhandstitt på verktyget för normalisering av andraspråkstexter

Dan Rosén

Språkbankens höstworkshop 2017



Språk-
BANKEN

SweLL: Forskningsinfrastruktur för svenska som andraspråk

- ▶ GU: Elena Volodina, Julia Prentice, Monica Reichenberg
 - ▶ SU: Mats Wirén, Gunlög Sundberg
 - ▶ UU: Beata Megyesi
 - ▶ UmU: Lena Granstedt
 - ▶ Riksbankens Jubileumsfond under 2017-2019
-
- ▶ insamling av ≥ 600 inlärartexter
 - ▶ normalisera: transformera texten m.a.p. “normspråket”
 - ▶ annotera efter en avvikelsetaxonomi

Exempel på inlärartext och normalisering

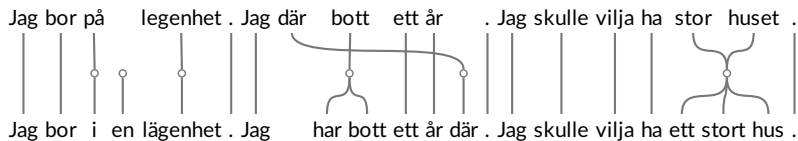
Jag bor på legenhet . Jag där bott ett år . Jag skulle vilja ha stor huset .

Exempel på inlärartext och normalisering

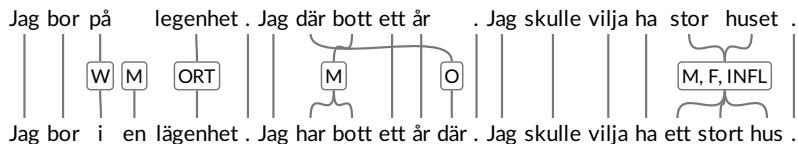
Jag bor på lägenhet . Jag där bott ett år . Jag skulle vilja ha stor huset .

Jag bor i en lägenhet . Jag har bott ett år där . Jag skulle vilja ha ett stort hus .

Exempel som parallellkorpus



Exempel som annoterad parallellkorpus



Demo

<http://demo.spraakdata.gu.se/dan/swell-editor>

Sammanfattning

- ▶ Verktyg för normalisering och annotering av inlärttexter
- ▶ Separation mellan dessa två aktiviteter
- ▶ Skapar parallellt en parallellkorpus

Nästa steg:

- ▶ Utveckling av avvikelsetaxonomi
- ▶ Internt pilotannoteringsprojekt

<http://demo.spraakdata.gu.se/dan/swell-editor>