

Resumé:

Korpusbaseret profilering af arvesprogstaleres sproglige kunnen

Gert Foget Hansen
Københavns Universitet

Det forholdsvis nye forskningsfelt 'heritage linguistics' fokuserer på talere der er opvokset med et (minoritets-)sprog, men som i hverdagen hovedsageligt anvender et omgivende majoritetssprog. Forskningen tager bl.a. sigte på at beskrive arvesprog og at afdække talernes sproglige kunnen på arvesproget.

Vurderingen af talernes sproglige kunnen hviler typisk på kvalitative beskrivelser og eventuelt optælling af visse prægnante træk, men kan også være en rent intuitiv og subjektiv vurdering. Vi vil i oplægget forsøge at gøre op med den intuitive tilgang og i stedet anvise måder at operationalisere spørgsmålet om fluency.

I projektet "Danske Stemmer i USA og Argentina" (KU) er der skabt et korpus af arvesprogssdansk bestående af et argentinadanske delkorpus på 542.000 ord med 50 talere og et amerikadanske delkorpus på 417.000 ord med 174 talere. Vi ønsker at benytte os af de muligheder som dette (i arvesprogssammenhæng) relativt store korpus giver for at udvikle kvantitative måder at karakterisere talernes sprogbrug og kunnen.

I begge korpora indgår basale baggrundsdata om informanterne som køn, fødeår, tidspunkt for emmigration eller udvandrer-generation, hjemby, mm. Optagelserne er transkriberet efter samme skabelon som LANCHART korpusset og derudover er alle ord annoteret for

sprog. Ord der ikke entydigt kan henregnes til ét sprog på grund af lydlig tvetydighed (fx *school* vs. *skole*) eller fordi de er hybrider (fx *milkfabrik*), er udskilt i selvstændige kategorier. Det samme gælder for tøvolyde, selvafbrydelser og egennavne.

I oplægget vil vi vise hvordan annotationerne og baggrundsdata giver mulighed for at tegne individuelle sproglige profiler for talerne som viser fx den præcise andel af dansk vs. engelsk, hyppigheden af hybride og ambige ord, og forhold der har med fluency at gøre som tøven, selvafbrydelser og talehastighed samt mulighed for at undersøge korrelationer mellem de forskellige aspekter af sprogproduktionen og/eller baggrundsvariable.