

Eit fullformsystem for analyse av eldre tekst på tidlig nynorsk, bygt på Aasen-normalen.

Christian-Emil Smith Ore, Oddrun Grønvik & Trond Minde

Universitetet i Oslo, Universitetet i Bergen

Ordbøker er vindauge inn i den tida dei blei skrivne. Definisjonane reflekterer den kulturforståinga som rådde i samtida. Slik sett er eldre ordbøker ein verdifull ressurs for kulturstudium. Men kan dei og hjelpa oss til å lage verktøy for maskinell analyse av eldre tekst?

For analyse av eldre norsk tekst er dette eit akutt behov. Dei norske standardskriftmåla har vore gjennom store endringar dei siste 150 åra. Bokmålet har sidan 1907-reformen fjerna seg frå dansk, og Aasen-normalen frå 1864-1873 skil seg vesentleg frå moderne nynorsk. Dess eldre tekst som skal analyserast ved hjelp av eksisterande verktøy, dess dårlegare blir derfor resultatet.

Det er bygt opp ein basis av eldre rettskrivingsdokument gjennom to komponentar i Språksamlingane (UiB 2021), Metaordboka og Ordbokshotellet (Ore&Grønvik 2020 og Ore 2020). Ordtilfanget er registrert med grunnform og ordklasse, og gjev interessante jamføringsresultat i høve til tidsakse og ordbokssjanger. Men til analyse av laupande tekst trengst dei bøygde formene. Ein fullformgenerator bygd på ei stor attenhundretalsordbok, med paradigme frå ein tilsvarende grammatikk, er verd å prøva.

Det langsiktige prosjektet for norsk språk vil vera å laga fullformregister for utvalde ortografiske målepunkt, t.d. åra for dei store rettskrivingsrevisjonane. For nynorsk er første målepunkt dokumentert i oppslagsord frå *Norsk Ordbog* (Aasen 1873) med bøyingsmønster (paradigme) frå *Norsk Grammatik* (Aasen 1864), og i 2020-2021 er det derfor laga en fullformsgenerator for Aasen 1873.

Ein ekstravinst er at sambandet mellom grammatikk og ordbok (Aasen 1864 og Aasen 1873) no er kartlagt og dokumentert. Aasens føremål med ordboka og grammatikken kan underkastast meir informert drøfting i høve til den tids vitskaplege standard for talemålsdokumentasjon og for bygging av standardspråk.

Med grunnlag i analysatoren for tidleg nynorsktekst kan ein studera i kva mon samtidstekstar fylgjer Aasens rettskriving. Fullformgeneratoren er førebels testa på tekstane til A. O. Vinje. Resultat vil bli framlagde i innlegget.

Referanser:

Ore, Christian-Emil Smith & Grønvik, Oddrun (2020). Studying language change through indexed and interlinked dictionaries. In Gavriilidou, Zoe; Mitsiaki, Maria & Fliatouras, Asmakis (Ed.), *Lexicography for Inclusion, Euralex 2021*. Euralex. ISSN 978-618-85138-1-5. p. 321–330.

Ore, Christian-Emil Smith (2020). Å ta Hans Ross på ordet: Ross' ordbok i relasjon til Aasens med Metaordboka som verktøy, I: Caroline Sandström; Ulla-Maija Forsberg; Charlotta af Hällström-Reijonen; Maria Lehtonen & Klaas Ruppel (red.), Nordiska studier i lexicografi, Rapport från 15 konferensen om lexicografi i Norden. Nordisk föreningen för lexicografi. ISBN 978-952-7359-03-7. kapittel. s 253 – 264

UIB 2021: <https://www.uib.no/ub/101277/norsk-kulturarv-i-skrift-og-tale>

UNIFOR2021 : <https://unifor.no/stiftelser/programredaktor-andor-birkeland-og-hustru-halinas-legat/> og <https://unifor.no/stiftelser/professor-didrik-arup-seips-fond/>

Aasen, Ivar (1965; 1. utg. 1864): *Norsk Grammatik*. Etter fyrsteutgåva frå 1864. Oslo. Universitetsforlaget. XXII, 371 s. (https://urn.nb.no/URN:NBN:no-nb_digibok_2006081600044)

Aasen, Ivar (1873): *Norsk Ordbog: med dansk Forklaring*. Christiania. Mallings Boghandel, XIV, 976 s. (https://urn.nb.no/URN:NBN:no-nb_digibok_2013122026001).

Nøkkelord: ordtilfang; tekstanalyse; fullformsyste; ordbøker og eldre tekst