

Kommentar til Korpuseminaret

Internt notat til deltakarar på seminaret

Takk for eit interessant og innhaldsrikt seminar!

Eg vil gjerne få lov til å presisere og utdjupe synspunkta mine i innlegget mot slutten av seminaret om parallellkorpus. På ein sein ettermiddag og i ein situasjon med eksterne forstyrringar er eg redd mine synspunkt kom ut litt unyansert og upresist.

Det var interessant å få meir innføring i det leksikografiske arbeidet. Eg verdset mykje arbeidet med Norsk ordbok og anna leksikografisk arbeid og støttar tiltak som kan vere til hjelp her. Vi har mellom anna hatt stor nytte av data i Ordbanken og ordbøkene på nett, og eg håper den planlagde revisjonen av desse blir realisert.

Når det så blir reist spørsmål i kva grad eit felles korpus for bokmål og nynorsk også kan brukast av språkteknologien, meiner eg dette:

1. Korpus som kan vere nyttig for språkteknologisk FOU bør planleggast ut frå det som er dagens og eit tenkt framtidig behov i den teknologiske utviklinga. Derfor bør også potensielle brukarar av slike korpus trekkast med for å konkretisere behova, og Språkbanken må stå sentralt.
2. Generelt er det eit stort behov for tilgang til nynorskkorpus. Alt som kan auke talet på slike korpus er positivt. Språkindustrien er særleg retta inn mot verktøy og prosessar som er nyttig for dagens språkbrukarar. Talekorpus er viktig her. Når det gjeld tekstkorpus, har tekstar med dagens normering størst verdi, og korpus bør primært avgrensast bakover til tidleg 1980-tal. Korpusa må vere lagt til rette med nødvendige metadata for maskinell behandling.
3. Med basis i desse kriteria ville vere fint om dei delane av Nynorskkorpuset som ikkje alt er tilgjengeleg i Språkbanken, kunne bli det. Men då som einspråklege korpus.
4. Generelt til spørsmålet om parallellkorpus: Parallelliteten bør som minimum ligge på tekstnivå, med kopling mellom bokmålstekst og korresponderande nynorsktekst. Lærebøker er eit eksempel. I den grad ein har ressursar til parallellføring på setningsnivå og tagging, er dette ein stor fordel.
5. Eg er spørjande til om eit Norsk nasjonalkorpus bygd på dei to aktuelle korpussamlingane vil oppfylle kriterium for parallellitet og tidsavgrensing og slik kunne fungere som parallellkorpus for industrien. Eg trur oppretting av eit parallellkorpus på friare grunnlag med utgangspunkt i dei språkteknologiske behova vil vere lettare å forsvare og få finansiert.
6. Ut frå etablert og lovfesta sambruk av bokmål og nynorsk i det norske samfunnet vil språkteknologi for norsk ha stor nytte av parallellkorpus på bokmål og nynorsk. Denne type korpus bør prioriterast i vidare korpusutvikling.

Bjørn Seljebotn, Nynodata

bjorn@nynodata.no