

Arbete med att samla in översättningsdata i Norge

Jon Arild Olsen
seniorrådgivare

Norska nationalbiblioteket

jon.olsen@nb.no

1. Den språkpolitiske situasjonen i Norge

To offisielle språk: norsk og samisk

To offisielle skriftformer for norsk: bokmål og nynorsk

Bokmål anvendes av ca. 87% av befolkningen, nynorsk av 13%.

Kulturdepartementet har besluttet at eTranslation skal fungere for både bokmål og nynorsk.

Betydelige leksikalske og morfologiske skillnader mellom bokmål og nynorsk.

Behøves separate språkdata.

2. Status for norsk språk i eTranslation

Ca. 1 million oversettingselement, væsentligen bokmål – engelsk.

Nåværende data kommer væsentligen fra oversettinger av den europeiske unionens regelverk («acquis communautaire»).

Ytterligere ca. 300 000 oversettelsesminner fra anonymiserte klagomål for å forbedre funksjonaliteten til Online Dispute Resolution Portal («Tvistløsning på nettet»).

Nasjonalbiblioteket fikk i 2017 ansvaret for å koordinere arbeidet med innsamling av språkdata til eTranslation.

To utmaningar:

- Øke antallet oversettelsesminner bokmål – engelsk.
- Erbjuda nynorsk i eTranslation.

3. Oversetting i offentlig sektor

Bare Utenriksdepartementet har «in-house» oversettingstjeneste:

- Seksjon for oversettelse av EU-lover (ca. 19 stillinger)
- Seksjon for generelle oversettelser (ca. 6,5 stillinger).

Noen offentlige virksomheter utfører oversettelse internt, men anvender ikke profesjonelle oversettelsesverktøy.

Majoriteten kjøper tjenestene av private oversettelsesbyråer.

4. Särskilda utmaningar for nynorsk

I dag ca. 100 000 oversettelsesminner mellom nynorsk og engelsk.

Lite oversettelse til og fra nynorsk i offentlige virksomheter.

Forslag fra eTranslation: gjøre bokmål til «pivot language» for nynorsk.

Behøves i så fall et stort parallelkorpus for bokmål og nynorsk.

5. Arbeide med å samle inn oversettingsminner

- a) **Avtale med Amesto Translations, nå Semantix, om oversettingsminner fra alle oppdrag for offentlige virksomheter.**

Ca. 1 millioner oversettingsminner.

Nasjonalbiblioteket har juridisk ansvar for all videre bruk, inkludert håndtering av personlige opplysninger og intellektuelle rettigheter.

Deler av materialet må anonymiseres.

b) Avtale med Standard Norge om mottak av oversettelsesminner fra perioden 2011-17.

Ca. 200 000 oversettelsesminner.

Anvendingen skal ikke skade Standard Norges kommersielle interesser.

Løsning: «Scramble» oversettelsesminnene.

c) Uni Computing har produsert bokmål/nynorsk – engelsk korpora basert på offentlige internettsider (NAV, Skatteetaten og Ny i Norge).

Ca. 100 000 elementer

6. Fremtidsplaner

- a) Sammen med Irland, Island og Kroatien har Norge søkt CEF-programmet om støtte til tre prosjekter:
- Fjerne konfidensielt materiale fra minner fra Utenriksdepartementets seksjon for generelle oversettelser. Tre arbeids måneder. Ca. 400 000 oversettingsminner.
 - Kvalitetssikre oversettingsminner fra EØS-sekretariatet i Brussel. En arbeidsmånad. Ca. 100 000 oversettingsminner.
 - Etablere et bokmål-nynorsk parallelkorpus på ca. to millioner elementer basert på lærebøker i gymnasiet.

7. For å lære mer

- ***Kartlegging av behovet for automatisk oversettelse i statlig sektor*** (Oslo Economics, 2016):
<https://www.regjeringen.no/no/dokumenter/>
- ***Norsk i den digitale tidsalderen/Norwegian Language in the Digital Age*** (MetaNet, 2012):
<http://www.meta-net.eu/whitepapers/e-book/norwegian-bokmaal.pdf>

Tack för uppmärksamheten!