

Nordisk dialektkorpus: Oversettelse fra dialekt til bokmål

Av Signe Laake i samarbeid med Ingebjørg Før Gjermundsen, Astrid Grov, Kristin Hagen, Janne Bondi Johannessen, Kari Kinn, Alexander Lykke, Eirik Olsen

Tekstlab 2011

NORDISK DIALEKTKORPUS: OVERSETTELSE FRA DIALEKT TIL BOKMÅL	1
VEILEDNING FOR OVERSETTELSEN AV LYDNÆR TRANSKRIPSJON TIL BOKMÅL	2
<i>Tagger</i>	2
<i>Pronomen</i>	3
<i>Bøyningsformer</i>	4
<i>Kjønn og kongruens</i>	5
<i>Mer kongruens</i>	5
<i>Stedsnavn</i>	6
<i>Spørreord</i>	6
<i>Sitater, avbrudd</i>	7
<i>Preposisjoner</i>	7
<i>Noen hyppige "problemord":</i>	7

Veiledning for oversettelsen av lydner transkripsjon til bokmål

Hovedretningslinjer:

- Vi normaliserer hardt til bokmålsnormalen.
- Vi normaliserer ikke ordstilling/syntaks, men normerer konsekvent mht morfologi
- To normaliseringer markeres spesielt med lang=x og lang=o

Generelt om normeringsprinsippet vårt: Normeringa er den delen av oversettelsesarbeidet som kan være vanskeligst. Det verste er som oftest å gi slipp på morfologiske særegenheter som en føler bærer såpass semantisk vekt, at betydningen endres ved oversettelse. Grunnpoenget er likevel at vi har den lydner transkripsjonen med kopling til lydopptakene, og følgelig trenger ikke bokmålsoversettelsen morfologisk sett være dialektner.

Vi bruker Bokmålsordboka (BMO), med visse unntak. Vi skal i hovedsak bruke hele normalen, så vi kan bruke den varianten som ligger nærmest dialekten. Vi skal ikke bruke tid på å slå opp for å finne dialektner former fordi vi har en lydner transkripsjon allerede. Slå kun opp ord som du er i tvil om er normert i bokmålsnormalen.

Vi oversetter ord for ord. Hvert ord i den lydner transkripsjonen skal oversettes med ett ord - **alltid en til en**. Det er ikke mulig å legge til ord eller ta noen ord bort.

Diskusjon rundt bokmålsnormen (fra møtet 04.11.10):

Vanskelig å få oversettelsen konsekvent når bokmålsnormen er så vid. Skal man velge en ukjent variant fra BMO som ligger nærmest dialekten, eller skal man velge den mest kjente og utbredte varianten? (Eksempel: *meine* vs. *mene*). Vi skal ikke bruke tid på å slå opp i BMO. Bruk den formen du kjenner fra før av. Her vil også det transliteratoren allerede har lært, være til hjelp. Dette vil også gjøre oss mer konsekvente. Så ikke rett opp i forslagene som transliteratoren kommer med. (Foreslår transliteratoren *heim* så la det stå.)

Forslag:

- utarbeide en liste over de hyppigste ordene som varierer og så bestemme hvilke former vi skal bruke for hvert ord. Vi kan bruke frekvensordlister o.l. for å finne ut hvilken variant "folk flest" ville valgt.
- bruke Aftenposten-normalen? Vanskelig å gjennomføre konsekvent og vanskelig å forsvare faglig.
- "advarsel" i søkegrensesnittet om at vi bruker hele bokmålsnormalen og en lenke til BMO.

Tagger:

Vi bruker to tagger: lang=x og lang=o.

Lang=x bruker vi som i NoTa: Dersom ordet ikke finnes i BMO, merkes det med lang=x. Dette kan være dialektord eller ord fra andre språk. Dette vil først og fremst være leksikalske ord. Ord som markeres med lang=x forsøker vi å normalisere i oversettelsen.

Eksempel:

Informanten sier: *æulette*. Vi skriver *øyelette* +lang=x

Informanten sier: *brøytebærrmen*. Vi skriver: *brøytebermen* +lang=x.

Informanten sier: *sjitrævvvel*. Vi skriver: *skitrael* +lang=x

Sett på lang=x i sammensetninger både hvis forleddet eller etterleddet er et ord som ikke finnes i BMO:

Mainstreamholdningen+lang=x

Lang=o bruker vi når må omskrive/oversette et funksjonsord. Dette er i tilfeller der dialekten har funksjonsord som ikke finnes i BMO i den samme betydningen. Her velger vi å bytte ut funksjonsordet med et standardisert bokmålsord og markerer ordet med lang=o.

Eksempel:

Informanten sier: *de va meir sny då så no* Vi skriver: *det var mer snø da **enn** +lang=o nå*

Informanten sier: *ette korrt så varrt de jo bærr å stå å prate* Vi skriver: *etter hvert så **ble** +lang=o det jo bare å stå og prate*

Derre blir ikke markert med lang=x. (Det er en glipp at dette ordet ikke er med i BMO.)

Huse va laga tå mur →

huset var laga av _<Event_desc="o" _type="language" _extent="previous"/> *mur*

”hun blei toucha _<Event_desc="x" _type="language" _extent="previous"/> veldig hardt”

Leksikalske ord brukt i en annen betydning, lar vi stå umarkert.

Diskusjon rundt lang=o (fra møtet 04.11.10):

Hvor nødvendig er lang=o-taggen? Ved å søke på lang=o, vil man få veldig mange treff på pronomen.

Dette vet man jo allerede varierer i norske dialekter. De interessante treffene som f.eks *så* for *enn*, vil bli borte i mengden av treff. Dessuten vil man ved å bruke kartfunksjonen få opp alle lyd nære varianter av det ortografiske ordet.

Men det kan virke meningsløst å fjerne alle lang=o-tagger på dette tidspunktet fordi det er noe vi har brukt mye tid på til nå.

Det kan være interessant å søke på lang=o for bestemte steder.

Pronomen

Pronomen blir oversatt til standardformene som finnes i BMO. Hvis informanten bruker et pronomen som finnes i BMO, men på ”feil” sted, blir dette pronomenet oversatt til riktig pronomen i følge BMO. Dette gjelder for eksempel kasus (jf. hovedrettningslinje nr. 2: morfologi normeres konsekvent). Se eksempel 1 og 2.

Marker pronomen som er sterkt fonologisk avikende med lang=o!

NB! For å unngå at hele korpuset fylles opp med [lang=o]-tagger, har vi sett oss tvunget til å ikke markere pronomen 3. p. sg./pl. Jf. Praksisen i eksemplene under, der eks. 1 og 2 ikke har [lang=o]-tag, mens eks. 3, som er 1. p. pl. har det.

Eksempel 1:

Informanten sier: *ka **dæmm** e fire stykkja?* Vi oversetter: *hva **de** er fire stykker?*

Eksempel 2:

Informanten sier: *æ blir me **ho**.* Vi oversetter: *jeg blir med **henne**.*

Eksempel 3:

Informanten sier: *me likar **konn** i Selljorr*. Vi oversetter: *vi+ lang=o liker oss+ lang=o i Seljord*.

Trykklett **n** og **a** blir oversatt til standard bokmål. Disse blir ikke markert med lang=o, i likhet med andre pronomen i 3. person. Trykklett **n** kan bli oversatt til **en**, **han** eller **den**. Vi prøver så godt vi kan å bruke dialektens eget system når vi avgjør hvilket pronomen vi oversetter til, men dette er vanskelig.

Vi oversetter ikke referensielt (anaforisk) pronomen *han/ho* til *den* som i eksempel 4:

Eksempel 4:

Informanten sier: *æ hadd fræmm sykkærn i hæLLga # **hann** e bra fin*.

Vi oversetter: jeg hadde frem sykkelen i helga # **han** er bra fin.

Liste over pronomen som vi markerer med lang=o:

2. person entall og 1. og 2. person flertall:

du oversetter vi *deg+lang=o*

åss/me/mi/konn oversetter vi *vi+lang=o*

dåkker/dikkån/dåkk/di oversetter vi *dere+lang=o*

Ikke marker varianter av 3. person flertall (døm, dæm, røm, rem, dom)

Bøyningsformer

I det transkriberte materialet varierer bøyningsformene mye. Vi normerer morfologi, og bruker derfor bare bøyningsformer som finnes i BMO. Et problem er når en godkjent ordform er brukt på feil sted:

Eksempel:

*alle var like **fattig** # og alle var like **rik** omtrent*

*datter hannes **ha** jo hæssta*

Fattig, *rik* og *ha* finnes riktignok i BMO, men *fattig* og *rik* er ikke godkjente flertallsformer og *ha* er ikke godkjent presensform. Å godkjenne godkjent bøyningsform på "feil" sted vil favorisere de bokmålsnære dialektene. I andre dialekter som er fjernere fra bokmål, vil ingen ord i BMO "tilsvare" uttalevariasjonen av bøyningsuffixs. Vi tror derfor det er riktigst å følge bokmålsnormalen når det gjelder bøying, selv om ordet likevel finnes i BMO med den lokale uttalevariasjonen av bøyningsuffixset. Dette stemmer også overens med en av hovedreglene i NoTa: *Legg merke til at vi skal bruke bokmålsordboka som norm både når det gjelder fonologisk stammevariasjon og når det gjelder uttalevariasjoner i bøyningsuffixs.*

Eksempelet over ville blitt oversatt til:

*alle var like **fattige** # og alle var like **rike** omtrent*

*datter hans **har** jo hester*

Noen ganger er det også korrekt å normere tempus i dialekter som avviker slik som i eksemplet under. Dette gjør det også lettere å normere dialekter som har preteritumsformer som ligner eller er like presensformer (for eksempel ...).

Eksempel:

*ikkje sekkert så mannge # så kannsje **ha** tatt denn ee # gådd denn veien vesst # di flesste måtte reise borrt*

(' ... som kanskje ***har** tatt den ... ')

Vi oversetter: (...) som kanskje **hadde** tatt den (...)

Kjønn og kongruens

Vi bruker kjønn etter normen i BMO og retter hvis informanten bruker ”feil” kjønn.

Eksempel:

æ brukkte ei massjin # de var denn massjina somm ...

Oversatt: *jeg brukte **en** maskin # det var den **maskinen** som...*

For å få riktigst mulig tagging, retter vi også kongruens:

*æ brukkte **ei lita** massjin*

Oversatt: *Jeg brukte **en liten** maskin*

Når vi retter kjønn på denne måten, må vi også rette eventuelle pronomenen som refererer til substantivet (legg forøvrig merke til at det her ikke rettes til *den*, som er det vanlige i bokmål, jf. eksempel 4 over, under overskriften pronomen):

*æ brukkte **ei** massjin # **ho** va effektiv ...*

Oversatt: *jeg brukte **en** maskin # **han** var effektiv ...*

Mer kongruens

Vi opererer muligens med et litt vidt kongruensbegrep, og oversetter altså slik:

Eksempel 1:

***de** derre dannseoppvisningan*

('***det** derre danseoppvisningene')

Vi oversetter: *de derre danseoppvisningene*

Eksempel 2:

*å så en å **anna dråssja** da*

('og så en og annen ***drosjen** da')

Vi oversetter: *og så en og annen **drosje** da*

Stedsnavn

Det nevnes mange stedsnavn i materialet. Ofte er dette små, lokale steder det ikke er mulig å finne på et kart. Ikke bruk tid på å finne den godkjente skrivemåten. Kanskje finne det ingen offisiell skrivemåte i det hele tatt. Normaliser navnet etter rettskrivningsregler. Husk stor bokstav!

Spørreord

Spørreord er en utfordring. De varierer mye og brukes på ulike måter. Vi velger å oversette dem til bokmål hvis det er en fonologisk og semantisk likhet jf. eksempel 1. Hvis spørreordet er fonologisk og/(eller?) semantisk avvikende fra spørreordene i BMO, oversetter vi dem til spørreordet med lignende betydning i bokmål og markerer dem med lang=o slik som i eksempel 2:

Eksempel 1:

Informanten sier: *ka du trur ho* vi oversetter til: *hva du tror hun*

Eksempel 2:

Informanten sier: *korr du trur de går me Vålerænnga te hællga?* vi oversetter til: *hvordan+lang=o du trur det går med Vålerenga til helga?*

I bokmålsnormalen kan det finnes flere former av samme spørreord (*hvordan/åssen*), og det er ikke alltid like lett å avgjøre hvilken form som ligger nærmest (for eksempel: *kossj'n*). I tillegg har bruken av *åssen* andre konnotasjoner enn *hvordan*. Om en dialekt bruker et annet spørreord enn *hvordan*, blir det feil å automatisk oversette til den mest radikale formen i bokmål. Vi må vurdere i hvert enkelt tilfelle om spørreordet ligner med på *åssen* eller *hvordan*.

KOMMENTAR TIL SENERE:

Dette er et av tilfellene vi bør vurdere å normalisere hardt til en eller den andre. "Hvorledes" er faktisk også normert bokmål, og følgelig det nærmeste når noen sier "kolles" for "hvordan/åssen". Om vi skal følge tidligere argumentasjon, er det ikke noen som helst grunn for å i noe tilfelle velge den mest dialektnære forma når vi normerer.

Eksempel 1: informanten sier: *kådannt jore oss det?* Vi oversetter til: *hvordan gjorde oss det?*

Eksempel 2: informanten sier: *håssen var konserrt'n?* vi oversetter til: *åssen var konserten?*

Utfordrende spørreord:

Å ti oversettes til *hva* <Event_desc="o"_type="language"_extent="previous"/> *tid*.

Ka ti oversettes til *hva tid*

Ko ti oversettes til *hvor tid*

Katti oversettes til *når+ lang=o*

Hen oversettes til *hvor + lang=o*

Ka oversettes til *hva* (uten lang=o)

Korr oversettes til *hvor* (uten lang=o)
(også i uttrykk som *korr hann skreik*: *hvor han skreik*)

Kass?

Sitater, avbrudd

Sitater og metaspråk markert med hermetegn skal også oversettes. Disse blir uansett automatisk oversatt. Marker ord som ikke finnes i BMO med lang=x.

Vi oversetter det som står inni anførselstegn til standard bokmål.

Avbrudd blir ikke oversatt.

Preposisjoner

Preposisjonsbruk har stor variasjonsbredde, og hver preposisjon har et stort semantisk virkefelt. "Feil" bruk av en preposisjon lar vi stå. Sammensatte preposisjoner som ikke står i BMO, oversetter vi, men vi markerer dem ikke.

Om ferien -> *om ferien* (ikke *i ferien*)

Innigjønnå -> *innigjennom*

Utforbi -> *utforbi*

Diskusjon fra møtet 04.11.10: Hva med preposisjoner som ikke finnes i BMO og som ikke heller er sammensatte preposisjoner (eks. noLa/norda). Vi kom frem til at det mest fornuftige er å sette lang=x på slike preposisjoner selv om denne preposisjonen/adverbet kan kalles et funksjonsord.

Noen hyppige "problemord":

vorte → blitt_<Event_desc="o" _type="language" _extent="previous"/>

veke → uke

hjá → hos_<Event_desc="o" _type="language" _extent="previous"/>

ta/tå → av_<Event_desc="o" _type="language" _extent="previous"/>

så → som (pass på at det her ikke skal være tagg, bare "fonetisk" variasjon)

ka → hva (uten lang=o)

korr → hvor (uten lang=o)

Sammensatte preposisjoner som ikke fins i BMO? Det er lov å lage sammensatte preposisjoner som ikke står i BMO

lyt -> oversettes til et passende modalverb som finnes i BMO og markeres med lang=o
Der og her som formelt subjekt → der/her (skal med andre ord stå uten tag).

Kass? ->

