

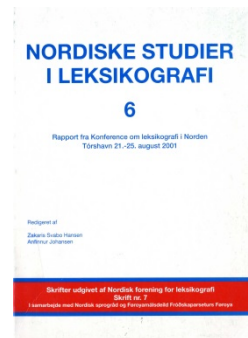
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Norsk Ordbok – om normeringa av Metaordboka

Forfatter: Terje Svardal

Kilde: Nordiske Studier i Leksikografi 6, 2003, s. 351-356
Rapport fra Konference om leksikografi i Norden, Tórshavn 21.-25. august 2001

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Norsk Ordbok – om normeringa av Metaordboka

In this lecture I will discuss problems related to my work with Metaordboka, a normalised register of all lemmas represented in the paper slip archives underlying “Norsk Ordbok” - the original slip archives showing considerable variety in the rendering of one and the same lemma.

My work on the Metadictionary has encountered mainly two types of problems. There are on one hand various problems related to the digitalisation of the archives, where f. i. poor paper quality makes the the digital archives difficult to read and interpret. On the other hand there are philological questions, varying from “spelling mistakes” to the normalisation of dialect material.

Innleiing

I dette føredraget vil eg presentere nokre typiske problem som har kome fram under arbeidet med normeringa av Metaordboka, eit prosjekt knytt til Norsk Ordbok. Metaordboka, som de har fått presentert frå ulike vinklar i dei føregåande føredraga, er tenkt som ein søkjeinngang til dei leksikalske databasane som Norsk Ordbok byggjer på. I første omgang normerer vi opp den digitaliserte utgåva av Setelarkivetarkivet til Norsk Ordbok. Metaordboka skal i første omgang lette arbeidet for redaktørane, då ein får ei grovsortering av materialet der variantar blir samla under eitt oppslagsord. Korleis dette skjer i praksis, vil eg vise etter føredraget då eg vil presentere Metaordboksapplikasjonen. I Metaordboka systematiserer vi materialet og grupperer setlar/ord under felles søkjeinngang. Det blir ikkje gjort homograf-separering. Så langt har eg gått igjennom oppslaga på i, j, k, o, p, q og delvis r. Vi starta arbeidet med bokstaven i/I, då det er den neste som står for tur for redigering i Norsk Ordbok.

I dette føredraget vil eg kort nemne nokre reint tekniske problem og deretter gå inn på normeringsproblematikken ved bruk av dialektmateriale. Eg vil prøve å belyse dette med eit tenkt døme på ordboksredigering, der eg ser på lågtysk i høve til høgtysk. Avslutningsvis vil eg demonstrere normeringsapplikasjonen, som er dataverktøyet eg arbeider med til dagleg.

Ordboksverket har som undertittel ”Ordbok over det norske folkemålet og det nynorske skriftspråket”. Dette skil det frå andre tilsvarande ordboksverk, som er reine skriftspråksbaserte verk. Eg vil gå inn på problemstillinga i høve til korleis ein normerer dialektmaterialet, som utgjer ein stor del av grunnlaget for redigeringa av Norsk Ordbok. Eg påstår at haldninga til dialekt/folkemål i høve til høgspåk er annleis i Noreg enn i andre europeiske land. Eg vil kome inn på denne problemstillinga i andre delen av føredraget.

Hovudarkivet til Norsk Ordbok består av ca. 3,2 millionar setlar, fordelte på i utgangspunktet ca. 500.000 oppslagsord. Setlane kan grovt delast inn i målføresetlar og setlar med litterære ekserpt. I tillegg finst det såkalla kombinerte setlar som inneheld både målføredøme og litterære ekserpt. Målføresetlane er materiale for ein stor del innsamla av dei om lag 600 frivillige medarbeidarane knytte til Norsk Ordbok. Den faglege bakgrunnen til desse frivillige medarbeidarane var svært ulik og setlane er til ein viss grad prega av det. For setlane viser dette igjen på det vis at rettskrivinga oppviser stor variasjon. Litteraturen

er samansett av originalverk og omsett litteratur. Original-litteraturen er for det meste skjønnlitteratur, medan omsetjingar er det mest av på fagbokområdet, då det ikkje er stor dekning av fagbøker skrivne på nynorsk.

Ved sida av dialektmaterialet er normeringa delvis inkonsekvent grunna dei ulike normalane som er brukt i arkivet. Norsk Ordbok er normert etter 1938-rettskrivinga, men i arkivet ser ein også spor etter nyare rettskriving. For Metaordboka har ein valt å følgje same norm som Norsk Ordbok rettar seg etter.

Vi følgjer som nemnt nynorsk rettskriving slik ho vart vedteken i 1938 i arbeidet med Metaordboka. Som rettleiing for normeringa bruker eg Matias Skards *Nynorsk ordbok* (Oslo, 1954) så langt ho er til hjelp. Dette er ei relativt lita ordbok med ca. 20.000 oppslagsord. Oppslagsord i Metaordboka som ikkje finst hos Skard blir normerte etter dei reglane ein kan lese ut av føreordet og av oppslagsorda i ordlista til Skard. Dessutan er dei allereie utgjevne delane av Norsk Ordbok (A-grøntanlegg), Grunnmanuskriptet (GM), som har relativt mange målførevarianter i døma, og Nynorskordboka (Oslo, 1993) til hjelp i normeringsarbeidet. Grunnmanuskriptet er det første utkastet til Norsk Ordbok. Det er ei samanstøyping av eldre ordbøker, m.a. av Aasen og Ross.

Problematisk i høve til normeringa er at ikkje nokon av dei eksisterande ordbøkene er fullstendige og dekkjer det komplette ordforrådet i Setelarkivet. Dette er for så vidt sjølvsagt, då Setelarkivet i alt har ca. 500.000 oppslagsord, medan Skard som nemnt har ca. 20.000.

Tekniske problem

Reint tekniske tilhøve utgjer ein relativt liten del av normeringsproblema. I hovudsak kan ein føre desse problema tilbake til setlar med dårleg kontrast som skannaren har hatt problem med å tyde. Av om lag tre millionar setlar i arkivet til Norsk Ordbok, blir færre enn ein prosent av dei definerte som ikkje lesbare. Dette utgjer ca 30.000 setlar som ikkje er lesbare i digital versjon. Dei eksisterer fysisk i arkivet. Dei har ikkje opplysningar om oppslagsord, ordklasse eller kjelde. Desse setlane har i første omgang mindre verdi i ordboksredigeringa.

Vidare manglar om lag 1,5% av setlane, om lag 45.000 setlar, opplysningar om ordklasse. Den grammatiske merkinga kan ein til dels finne igjen ved å samanlikne setlane i dataarkivet med Setelarkivet til Norsk Ordbok, eller ein kan utifrå døma på setlane avgjere kva ordklasse og ved substantiva kva kjønn oppslagsorda skal førast under. Er det ført opp kjelde, vil ein ved å sjekke denne, kunne finne ut rett ordklasse for oppslagsordet. Setlane i arkivet er delvis også retta eller normerte med blyant og sorterte etter desse tilføyningane i arkivet. Desse rettingane eller normeringane er i varierende grad komne med på dei digitale setlane, og tidlegare omsortering av arkivet er dermed til dels gått tapt. Men, som det allereie er peika på, utgjer desse setlane ein relativt liten del av arkivtilfanget.

Av setelarkivet er m.a.o. ca. 75.000 setlar vanskelege å normere frå ein reint teknisk synsvinkel, men på andre sida er faktisk 97,5% av setlane lesbare utan større problem!

Normeringsproblem

Det største problemfeltet går på ortografi og ordklasse eller grammatiske spørsmål. Eg vil prøve å illustrere dette med typiske døme henta frå arbeidet mitt så langt.

Ei forholdsvis enkel avgjerd normeringsmessig og eit typisk døme har ein ved dei oppslaga som inneheld *kaffi-/kaffe-*, som enkeltoppslag og i samansetjingar. Her har eg valt

kaffi- som norm, då dette eintydig er den korrekte forma, som er oppført i Skards ordbok og også i andre nynorske ordlister og -bøker. Forma *kaffe* må ein definere som bokmåls/riksmåls/dialektvariant av *kaffi*. Dette er eintydig det same ordet, og difor ikkje problematisk å føre under same oppslaget.

På same måten har eg handsama *kross-/kors*-variasjonen i tilfanget. Forma *kross* er den tilletne forma etter 1938-rettskrivinga og må difor også brukast i Metaordboka. Einaste tvilen er at *kors* er teke inn som sideform i nynorsk etter seinare rettskrivingsbrigde, men ettersom Metaordboka berre skal innehalde hovudformer, må oppslagordet eintydig vere *kross* i dette tilfellet.

Som nemnt har eg også gjort endringar i ordklassetilvisingane der det har vore naudsynt. Feila fordeler seg på fire ulike typar. Det er «feil», der ein må peike på variasjonen i norske dialektar i høve til nynorsk normalspråk, eller avvikande grammatisk kjønn for substantiv, feil ordklassetilvising og manglande ordklasse som kan rettast på grunn av seteldøme eller kjeldetilvising.

I dei tilfella der Skard eller dei andre rettskrivingskjeldene oppgir berre eitt grammatisk kjønn, har eg normert andre variantar i arkivet til dette kjønnnet. Eit døme er oppslagsordet *jolle m*, som det i Setelarkivet finst setlar merka med **m** og setlar med merkinga **f/m**(femininum og maskulinum). Eit anna døme er *bane* som etter norma frå 1938 er maskulinum, men som har ein relativt høg frekvens av døme der ordet er hokjønn. Eg kan peike på at seinare normeringsvedtak for nynorsk har teke omsyn til den høge frekvensen av *bane f* i nynorsk tekst, og teke det inn som alternativ ved sida av maskulinum.

Eit større problem har vi ved substantiv som etter nynorsk rettskriving kan ha fleire grammatiske kjønn, som t.d. *lov m* og *f*, eller *klasse m* og *f*. Her har eg, i samråd med redaktørane for Norsk Ordbok, valt eitt kjønn ut frå kva variant som har høgst bruksfrekvens i moderne nynorsk. Argumentet for denne løysinga er at dei grammatiske opplysingane i Setelarkivet ikkje vil gå tapt, sjølv om vi vel å normere t.d. *lov* til hokjønn i Metaordboka. Til kvart oppslagsord eller artikkel i Metaordboka er det knytt x tal på setlar. Kvar einskild setel inneheld eigne opplysningar mellom anna om ordklasse. Eg vil vise dette ved demonstrasjonen av normeringsapplikasjonen som er arbeidsverktøyet mitt. Metaordboka er ein inngangsportale til Setelarkivet som skal hjelpe brukarane til å orientere seg i materialet, og ho bør difor vere så eintydig og lett å finne fram i som råd er.

Felles for alle desse normeringsproblema og løysingane på dei, som eg har gått inn på her, er at dei får konsekvensar for heile basen. Norsk er eit språk med stort potensial for to- eller fleirledda samansetjingar. Setelarkivet til Norsk Ordbok inneheld mange samansetjingar, og det tyder dermed at når vi normerer enkeltord, som t.d. *kaffi*, *klasse* og *lov*, så får dette konsekvensar for heile basen. Har vi valt *lov f*, i tydinga *lov* og *rett*, må vi vere konsekvente i heile basen, og t.d. normere *trygdlov* og *skattelov* til hokjønn.

Dialekttilfang

Eg vil få innleie delen som omhandlar handsaminga av målføretilfang i NO samanlikna med andre tilsvarande ordbøker med to temmeleg ulike sitat. Det eine er frå redatøren av Språkvård si innleiing til Prof. Karl-Hampus Dahlstedts artikkel "Dahlstedts hundra ord" henta i frå Språkvård (4/1986). Der heiter det: "Av de drygt 250 ord som han då noterade har han här valt ut 100 ord som han anser värda att tas upp i riksspråket".

Det andre sitatet er henta frå redaktør Alf Helleviks føreord til Norsk Ordbok. Ordlyden

er: "Når det gjeld bruksområdet og stilverdet åt orda, korkje kan eller bør ein derfor i denne ordboka draga noko skarpt skil mellom målføre (dialekt) på den eine sida og skriftmål (med vyrdsla talemål) på den andre. Sjølv sagt er det mange målføreord som ikkje godt høver inn i riksnorma av nynorsken, men det er då jamnast av andre grunnar enn den eine at dei er "dialektiske". I *prinsippet* kan ein såleis seia at alle gode målføreord er gode skriftmålsord." (NO, s. VII)

Eit sitat frå Peter v. Polenz Tyske språkhistorie viser ei vanleg, eller den vanlege haldninga til dialektilfang i høve til høgspråk: "I dag kan ein berre handsame lågtysken som ein del av den tyske dialektologien."¹

Som nemnt før i består arkivet av mykje tilfang frå dei norske dialektane og det er og eit mål at Norsk Ordbok ikkje berre skal dokumentere det nynorske skriftspråket, men også dei norske målføra. Haldninga til folkemål er i hovudsak annleis i Noreg enn i andre land, der skiljet mellom folkemål og høgspråk er skarpt, eller i alle høve skarpere enn i Noreg. Dette er ei heilt anna haldning til folkemålet enn i NO, der mykje av grunnlaget for redigeringa nettopp er målføretilfang og dette tilfanget frå blir handsama som ein sjølvsgd, integrert del. Skulle ein tenkje seg ei ordbok utarbeidd for tysk etter same prinsipp som NO, reiser det problemet korleis ein ville normere dialektvariasjon i høve til den høgtytske norma. Dette både med omsyn til ordforråd og uttale/ortografi.

I andre lands nasjonalordbøker, som t.d. DWB, WDT, OED, SAOB og ODS, som vi kan samanlikne Norsk Ordbok med, tek ein som nemnt lite eller inkje omsyn til dialektane. Ved sida av dei store, nasjonale ordboksverka vil ein finne ordbøker over ulike dialektar eller regionale variantar i språkområdet.

Korleis handsamar så vi målføretilfang i Metaordboka, og for den del i NO og NOB? Eit døme på normering av målføretilfang har vi ved oppslagsordet *kartoffel* i Metaordboka. Til dette oppslagsordet er det knytta i alt 25 setlar med i alt 6 variantar. Vi har følgjande variantar: *kantaføl*, *kanteføl*, *kanteföll*, *kantefölle*, *kantoffel*, *kartoffel* og *kartoflar*. Desse har vi valt å samle under *kartoffel*, då dette er normeringa på både dansk og tysk. Det må eintydig vere same ordet.

Eit anna døme har vi ved *kappelaup*, som er eit dialektnamn på *løvetann*. Under *kappelaup* har vi samla *kappelau*, *kappilaup*, *kappilaav*, *kapplau*, *kapplauv*, *koppelaup*, *koppeloppe*. Inngangen *kappelaav* fører brukarane rett inn i dialektvariasjonen i materialet.

I innleinga sa eg at eg ville ta føre meg eit tenkt døme og diskutere korleis ein ville handsame dialektilfang i eit ordboksverk over det tyske språket der ein i tillegg til døme frå litteraturspråket tek omsyn til dei tyske dialektane. Høgtytsken byggjer i hovudsak på Mitteldeutsch (mellomtysk) med innslag av oberdeutsche (høgtytsk) dialektar. Lågtysk har ikkje sett spor etter seg i høgtytsken. Problema vil etter mi meining vise seg i to hovudtypar, geografisk betinga ordforråd og lydmessige skilnader mellom dei tyske dialektane og korleis ein i tilfelle normerer dei. I denne samanhengen kunne ein også diskutere om lågtysk skulle handsamast som eige språk, i og med at den høgtytske norma byggjer på mellom- og høgtytske dialektar. Men den diskusjonen vel eg å sjå bort bort frå her.

Eit nokså likt døme har vi ved sørlandske dialketar og grensa mellom lågtysk og høgtytsk. Den bløde kyststriben, med b d g for p t k, normerer vi i Metaordboka til harde konsonantar. Den bløde kyststriben blir på normert norsk til den blaute kyststripa. Tilsvarende måtte ein for lågtysk, som ikkje blei omfatta av den 2. høgtytske, eller gammalhøgtytske lydforstykinga måtte normere open til offen, perd til pferd, appel til apfel o.s.b. Tek ein utgangspunkt i

historisk lydutvikling for dei ulike dialektane, er det råd å ta dei inn i ei normert ordbok.

Ordgeografiske tilhøve er også interessante å sjå på i høve til om ord er "ordboksverdige" for å vri litt på redaktøren si innleiing til prof. Dahlstedt sin artikkel. Ved utarbeidinga av NOB, blei det ved sida av bruksførekomst (frekvens) teke omsyn til geografisk utbreiing. Eit kriterium var at eit ord måtte vere belagt i minst fire fylke for at det skulle takast med i NOB. For tysken sin del må ein gå ut frå at ordforrådet oppviser ein viss variasjon frå Alpane til Nordsjøen. Men ved ordgeografiske undersøkjingar, vil ein ut frå eit krav om geografisk spreining, velje målføreord som skal takast inn i eit slikt verk.

Konklusjon

Ut frå det reint praktiske arbeidet med Metaordboka og erfaringane vi har har gjort der, for ikkje å gløyme tidlegare systematiseringsarbeid i Setelarkivet, må ein ved det tenkte tyske ordboksdømet kunne seie at det finst ei løysing på normering av folkemål, som gjer at regionalt prega språk, eller dialekttilfang kan takast med i ei allmennordbok. For å gjere Helleviks ord til mine, I *prinsippet* kan ein (såleis) seia at alle gode målføreord er gode skrifmålsord.

¹ "Heute kann das Niederdeutsche nur noch in der dt. Dialektologie behandelt werden." Peter von Polenz, Geschichte der deutschen Sprache, de Gruyter, 1972 (Mi omsetjing).

Litteratur

Ore, Chr.-E. *Metaordboken – et rammeverk for Norsk Ordbok* i , Gøteborg 1999.

Grønvik, O. *Om kjeldegrunlaget for Norsk Ordbok* i Norsk Ordbok – nynorsken leksikografiske kanon?, Oslo 1996.

Rommetveit, M. *Stilmarkørar* i Ord og Mål , Oslo 1989.

Polenz, P. v, Geschichte der deutschen Sprache, de Gruyter, 1972.

Norsk Ordbok, Oslo 1966.

Dahlstedt, K.-H. "*Dahlstedts hundra ord*" i Språkvård 4/1986.

Schmidt, W. Geschichte der deutschen Sprache, Stuttgart/Leipzig 1993.

