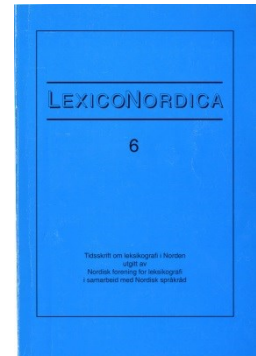


LexicoNordica

Titel: Nyttige normer i et ”normløst” sprog
Forfatter: Torbjørn Nordgård
Kilde: LexicoNordica 6, 1999, s. 91-98
URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

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 LexicoNordica (1-16) 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.

Torbjørn Nordgård

Nyttige normer i et "normløst" språk

This article discusses certain controversial issues in connection to Norwegian writing standards: the apparent freedom of choice with respect to spelling and the lack of style definitions (radical or conservative) and pronunciation information in Norwegian dictionaries. It is argued that such information should indeed be included in dictionaries written for immigrants, and methods for adding such information based on the project *NorKompLeks* (Norwegian Computational Lexicon) are outlined.

Innledning

Denne artikkelen tar opp visse normeringsemner som på grunn av sin kontroversielle status i norsk språkpolitikk ikke er nedfelt i ordbøker som Norsk språkråd har vært med på å utarbeide (f.eks. Bokmålsordboka, jf. Landrø & Wangensteen 1993): stilnormering (radikalt og moderat bokmål eller nynorsk), herunder stavemåter av oppslagsformer og bøyingsparadigmer, og ikke minst uttalevariasjoner. For innvandrere er det et problem at stilnormer ikke er definert, fordi valgfriheten i norsk skriftspråk er vanskelig å forholde seg til for en som ikke har norsk som morsmål. Det er naturligvis også et problem for innvandrere at moderne norske ordbøker ikke har uttaleinformasjon, når vi ser bort fra visse fremmedord.

Vi skal nedenfor gå nærmere inn på slike problemstillinger og an-vise noen forslag til løsninger der prosjektet *NorKompLeks* vil stå sentralt. Det er kun bokmål som blir behandlet her, men argumentene og konklusjonene som fremføres nedenfor, har også gyldighet for nynorsk.

Stilnormer i norsk

Det finnes ulike grader av skriftspråksnormering. En "streng" normering tillater få eller ingen valg med hensyn til stavemåter og bøyingsalternativer. Dette er situasjonen i f.eks. fransk og engelsk skriftspråk. På den andre siden av skalaen finner det vi kan kalle "løs" normering, dvs. at språkbrukerne selv kan velge mellom ulike stavemåter. Norsk skriftspråk er et godt eksempel i så måte.

Dette betyr ikke at det ikke finnes skriftspråksnormering i norsk. I tilfeller der det finnes alternativer, skiller man mellom sidestilte varianter (f.eks. *ørret* og *aure*), sideformer (f.eks. normalformen *røyke* og sideformen *røke*) og observerte, men likevel ulovlige stavemåter (f.eks. *flyver*). Dersom man kun benytter normalformer, holder man seg i praksis til den såkalte læreboknormalen. Denne normen skal følges i lærebøker for skoleverket og i offentlige dokumenter.

Dagens norske skriftspråksituasjon kan karakteriseres som et kompromiss mellom ulike rettskrivingsreformer i dette århundret. Etter 1981-reformen har det egentlig ikke vært mye strid om prinsippene for norsk rettskriving, selv om Språkrådets forslag til nye stavemåter av ord av utenlandsk opprinnelse alltid skaper en viss debatt, men dette er ikke noe særnorsk fenomen.

I bokmål kan vi skille ut tre skriftspråksstandarder: læreboknormalen, radikalt bokmål og moderat bokmål. Litt forenklet kan man si at i radikalt bokmål foretrekkes dialektnære og nynorsklignende varianter, mens moderat bokmål følger motsatt strategi. Vi kan i tillegg se for oss en "liberal" norm der alle sideformer aksepteres og kanskje også en "mest brukt" norm basert på frekvenstillinger i løpende tekst.

Verken den radikale eller den moderate bokmålsvarianten inkluderer læreboknormalen i sin helhet. Dette kan illustreres med de sidestilte formene *ørret* og *aure*. I moderat bokmål finner man ikke *aure*, selv om *aure* ikke er noen sideform. I radikalt bokmål er derimot *aure* helt kurant. Dette betyr at man ikke kan identifisere radikalt og moderat bokmål ved å kartlegge hvordan de velger sideformer.

Bildet kompliseres ytterligere ved at brukere av radikalt og moderat bokmål tillater seg å anvende ulovlige former. Den ulovlige formen *sne* kan forekomme i moderat bokmål istedenfor den definerte formen *snø*. I radikalt bokmål kan man finne dialektuttrykk som ikke er anerkjent, f.eks. *åsså* (også) og *tjangs* (sjanse).

Vi har så langt bare sett på variasjoner i oppslagsformer. Stilnormene dukker også opp i bøyninger der det finnes bøyingsklasser som kan karakteriseres som henholdsvis moderate og radikale. Demarkasjonsprinsippene er de samme som ved oppslagsformene, – radikalt bokmål foretrekker nynorsklignende og dialektnære bøyninger, mens moderat bokmål unngår slike bøyingsvarianter. Noen eksempler på radikale bøyninger er *aska*, *kasta* og *rustfarga*, mens moderat bokmål har *asken*, *kastet* og *rustfarget*. Adjektivisk bruk som involverer *kaste*, varierer også: *de utkasta pøbbgjestene* vs. *de utkastede pubgjestene*.

Når stilnivåer på grunnforms- og bøyingsnivå kombineres, blir bildet nokså komplisert. *Aske* og *oske* er sidestilte former, det samme er

slokke og *slukke*, mens *sløkke* er en (radikal) sideform. Dette gir opphav til muligheter i bestemt form entall av *aske/oske*:

asken
aska
osken
oska

Formen *asken* brukes i moderat bokmål, mens *aska* og *oska* er gangbare i radikalt bokmål. *Osken* er en merkelig hybrid som er tillatt, dog som sideform. Man vil antagelig aldri finne *osken* i bruk. For *slokke* – *slukke* – *sløkke* dukker det opp mange varianter. Vi ser nærmere på adjektivisk bruk i flertall og bestemt form entall:

slokka (den sløkka brannen)
slokkede (den sløkkede brannen)
slokkete (den sløkkete brannen)
sløkte (den sløkte brannen)

slukka (den slukka brannen)
slukkede (den slukkede brannen)
slukkete (den slukkete brannen)
slukte (den slukte brannen)

sløkka (den sløkka brannen)
sløkkede (den sløkkede brannen)
sløkkete (den sløkkete brannen)
sløkte (den sløkte brannen)

Det er altså tolv tillatte varianter å velge mellom. Læreboknormalen akspeterer seks av dem. Formene som ender på *-ede*, er sideformer, sammen med alle *sløkke*-tilfellene. *Slukte* og *sløkkede* er merkelige hybrider. Vi ser ved hjelp av dette eksempelet at bokmål egentlig ikke kan sies å ha noen normering i ordets intuitive forstand. Vi observerer også at læreboknormalen ikke skjærer igjennom valgfriheten på noen effektiv måte.

Stilnivåer bør ikke blandes. Stavemåten *pøbb* for *pub* hører avgjort ikke hjemme i moderat bokmål, og *utkastede* hører ikke hjemme i radikalt bokmål. Derfor blir sekvenser som *de utkastede pøbbgjestene hadde tømt osken fra sigarettene på gulvet* skjærende hybrider som er gangbare etter norsk rettskrivning. Moderat bokmål skulle tilsi *de*

utkastede pubgjestene hadde tømt asken fra sigarettene på gulvet, mens radikalt bokmål gir de utkasta pøbbgjestene hadde tømt aska/oska fra sigarettene på gulvet. Andre kombinasjonsmuligheter kan kanskje også rubriseres som radikale.

Et forsøk på å definere stilnormer i norsk er gjort i EU-prosjektet SCARRIE, som er en pilotapplikasjon for korrekturlesing i dansk, svensk og norsk, jf. de Smedt & Rosén (1998).¹

Problemer

For godt skriveføre nordmenn er det ikke noe problem at det ikke finnes offentlig definerte normer som beskriver konsekvente stilnivå. Heller tvert imot. Språkføringen kan varieres, og man trenger ikke tenke særlig på hva som er rett stavemåte, siden man uansett klarer å holde seg innenfor det som er tillatt. Samtidig har de fleste nordmenn klare intuisjoner om hvordan stilnormene er, selv om de ikke er satt på trykk noe sted.

Men for innvandrere og nordmenn som har skrivevegringer av ulik grad, fører denne mangfoldigheten til problemer. Det er ikke lett å forklare en utlending hvorfor hybridsetninger som dem ovenfor bør unngås, når man ikke kan vise til ordbøker og grammatikker med uttømmende definisjoner av stilnivåer.

Fraværet av offisielt definerte stilnormer fører videre til at data-maskinelle verktøy for korrekturlesing ikke kan kontrollere stilbrudd slik man med rimelighet kan anta at brukerne ønsker. Så lenge offisielle språkorganer ikke engasjerer seg i dette arbeidet, vil andre gjøre det, og resultatet kan etter hvert bli at vilkårlige private aktører overtar forvaltningen av norsk rettskriving.²

Hvordan kan man definere stilnormer?

Man kan i hovedsak se for seg to måter å definere stilnormer på. Den ene er ved systematisk gjennomgang av de offisielle ordlistene med sikte på stilnormeringer, f.eks. nøytral, moderat og radikal norm. Dette er en arbeidskrevende oppgave, men den kan delvis automatiseres

¹ Enkelte av eksemplene i denne artikkelen er hentet fra SCARRIE-relatert leksikonarbeid ved NTNU Trondheim.

² Her er SCARRIE-prosjektet et eksempel på hva som kommer til å skje i årene som kommer. Vi kan risikere at IT-industrien gjør arbeid som Norsk Språkråd vegrer seg for å gjøre.

dersom man benytter ordlistene fra prosjektet *Norsk komputasjonelt leksikon* (NorKompLeks) ved NTNU i Trondheim³, sammen med det leksikografiske materialet ved Seksjon for leksikografi og målføregransking ved Universitetet i Oslo. I *NorKompLeks*-materialet kan bøyingskodene tilpasses stilnormer, f.eks. bøyningen av *aske* og andre svake substantiv som enten er maskuline eller feminine: "feminin" bøyning hører til den radikale normen, mens den "maskuline" bøyningen hører til den moderate. Dette er et generelt mønster: dersom et oppslagsord kan bøyes "feminint" eller "maskulint", er alltid den feminine bøyningen radikal, mens den maskuline er moderat. Et lignende mønster finner vi for verbmorfologien, der dialektlignende varianter typisk hører hjemme i radikalt bokmål. Denne strategien virker for alle oppslagsord som kan bøyes på mer enn en måte, men man må naturligvis samle opp og analysere alle kodekombinasjonene (ca. 400 stykker).

De tilfellene der det kun finnes en bøyingskode som er spesifisert som sideform, må gjennomgås manuelt. Et eksempel er *granskau*:

granskau : [m1]

Antallet slike sideformer der det ikke er valgmuligheter, er ikke plagsomt stort (ca. 450). Vi ser altså at det er ikke så vanskelig eller tidkrevende å definere radikalt og moderat bokmål, dersom det er interesse for det. Siden det er såpass enkelt, er det viktig at offisielle språkorganer tar på seg denne oppgaven snarest.

En annen måte å definere stilnormer på er ved bruk av maskinlesegelige korpora der hvert oppslagsord er tilordnet informasjon om ordklasse og bøyingsinformasjon. Slik ordmerking i løpende tekst kan i dag gjøres automatisk, jf. f.eks. Johannessen (1998) og Nøklestad (1998). Dersom man klassifiserer tekstene etter stilnivå, kan man automatisk identifisere ordformer som hører hjemme i moderat og radikalt bokmål. Denne metoden har den åpenbare svakhet at tekstutvalget langt på vei bestemmer svarene, og det kan vise seg problematisk å få tak i tilstrekkelige tekstmengder. Det beste er antagelig en kombinasjon av ordlistegjennomgang og verifikasjon i løpende tekst.

Uttalebeskrivelser i ordbøker

³ NorKompLeks er nærmere beskrevet i Nordgård (1998a,b).

Det er generelt liten forståelse for at talenormer er nødvendige i norsk, noe som kommer av at de fleste nordmenn, inklusive undertegnede, aksepterer dialektbruk på en helt annen måte enn det som er vanlig i andre land. Men dette betyr ikke at man i ordbøker dermed skal avstå fra å opplyse om hvordan ord faktisk uttales i norsk. Slike beskrivelser er ikke det samme som en talenorm. En talenorm skal fortelle hvordan ord *bør* uttales, og det er noe helt annet enn å beskrive hvordan ord faktisk uttales i norsk. Likevel, offisiell norsk språkpolitikk har lite å si om norsk uttale.

Diskusjonen om hvorvidt det bør etableres en norsk talenorm, skal ikke berøres her. Men en uheldig konsekvens av at det knapt finnes uttalebeskrivelser for norsk (et unntak er Berulfsen 1969), er at innvandrere får unødig store problemer med å lære seg norsk. Når man skal lære seg et nytt språk, har man selvfølgelig behov for å vite hvordan ordene skal uttales. En ordbok uten uttaleinformasjon er simpelthen ikke særlig interessant for en som ikke har det aktuelle språket som morsmål. Fremtidige ordbøker for norsk bør inkludere informasjon om hvordan ord uttales i ulike regioner av landet.

Uttaleinformasjonen må inkludere vanlig fonemisk transkripsjon, inklusive trykkplassering og tonelagsmarkering på leksikalske oppslagsformer. Der dialektgrupper plasserer trykket på ulike steder, må det markeres. Det samme gjelder distribusjonen av "tjukk l", palatalisering osv. I tillegg er det av og til behov for å opplyse om tonelagsendring i bøyingsparadigmene, f.eks. bøyningen av monosyllabiske hankjønnsord som har tonem 1 i entall, men tonem 2 i flertall (jf. bøyningen av *bil*). Dette er oftest unødvendig informasjon for dem som har norsk som morsmål, men for utlendinger er den viktig.

Nyere uttalebeskrivelser for norsk er etablert i tilknytning til *Nor-KompLeks*-prosjektet. I dette materialet er alle grunnformer i bokmål og nynorsk beskrevet. Man har i første omgang lagt til grunn de vanligste uttalevarianter i sørøstnorsk. Beskrivelsene dekker variasjoner i trykkplassering, som i *ba1nan* (tonem 1 og trykk på andre stavelse) og *2banan* (tonem 2 og trykk på første stavelse). Det er også mulig å maskinenerere fonemiske beskrivelser av bøyingsparadigmene der feilraten er minimal. Et eksempel er listet opp nedenfor. Transkripsjonen er SAMPA, dvs. 7-bits redusert versjon av IPA. Tegnet =n er stavelsesbærende n, @ er trykklett pregløs "e" eller schwa, "markerer trykkplassering og tonem 1, "'markerer trykkplassering og tonem 2:

<i>Fonemisk transkripsjon i SAMPA</i>	<i>Grammatiske egenskaper</i>
" "bAnAn	subst hankj entall ubestemt

" "bAnAn=n	subst hankj entall bestemt
" "bAnAn@n	subst hankj entall bestemt
" "bAnAn@r	subst hankj flertall ubestemt
" "bAnAn@n@	subst hankj flertall bestemt
" "bAnAn=n@	subst hankj flertall bestemt
bA"nA:n	subst hankj entall ubestemt
bA"nA:n@n	subst hankj entall bestemt
bA"nA:n=n	subst hankj entall bestemt
bA"nA:n@r	subst hankj flertall ubestemt
bA"nA:n@n@	subst hankj flertall bestemt
bA"nA:n=n@	subst hankj flertall bestemt

Vi ser av dette eksempelet at *NorKompLeks*-beskrivelsene også inneholder grammatisk informasjon, og man kan generere grafemiske bøyingsparadigmer på samme måte som de fonologiske paradigmenes. Den grammatiske informasjonen er i det alt vesentlige i overensstemmelse med *Norsk Referansegrammatikk*, jf. Faarlund, Lie og Vannebo (1997).

Potensial for innvandrersordbøker

Dersom det er interesse for å utnytte informasjonstypene beskrevet ovenfor i fremtidige innvandrersordbøker i Norge, vil ordbøkene kunne tilføres dimensjoner som innvandrerne i lang tid har etterlyst: uttaleinformasjon, stilinformasjon og fullstendig informasjon om bøyninger, både grafemisk og fonemisk. Uttaleinformasjonen vil ikke være normativ, men deskriptiv. Brukerne kan selv velge hva slags uttalevarianter de vil holde seg til, og de vil formodentlig velge dem som er mest brukt der de bor. Stilinformasjonen vil kunne fjerne usikkerhet med hensyn til de valgfrie formene. Bøyningsinformasjonen hjelper dem til å bruke de riktige grammatiske formene.

Ordbøker som inkluderer all denne informasjonen, blir svært omfangsrike dersom de trykkes opp på vanlig måte. Kanskje kan man trykke ordbøker med ulike typer informasjon, – en vanlig type med tradisjonell leksikografisk informasjon, en med uttaleinformasjon, en med stilinformasjon, osv. Likevel, den beste løsningen vil nok være et hypertextformat, f.eks. på CD-ROM eller rett og slett på Internett. Hypertextformatet tillater at informasjonen modulariseres: brukeren kan ved enkle operasjoner få fram uttaleinformasjon, stilinformasjon eller bøyingsparadigmer. Dette formatet gjør det overkommelig å lage

nyutgivelser i takt med økning av ordforrådet og eventuelle grammatiske og fonologiske endringer (f.eks. den for enkelte formastelige utbredelsen av "sj"-lyden i talt norsk).

Referanser

- Berulfsen, Bjarne 1969: *Norsk uttaleordbok*. Oslo: Aschehoug.
- Faarlund, Jan Terje/Svein Lie/Kjell Ivar Vannebo 1997: *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Hagen, Kirstin/Janne Bondi Johannessen 1998: Disambiguering uten syntaks. I: Faarlund, Jan Terje/Brit Mæhlum/Torbjørn Nordgård (red.): *MONS 7. Utvalde artiklar frå det 7. Møtet Om Norsk Språk*. Oslo: Novus.
- Landrø, Mærit Ingebjørg/Bøye Wangensteen 1993: *Bokmålsordboka: definisjons- og rettskrivningsordbok*. Oslo: Universitetsforlaget.
- Nordgård, Torbjørn 1998a: Norwegian Computational Lexicon (NorKompLeks). I: *Proceedings of the 11th Nordic Conference on Computational Linguistics*. Center for Sprogteknologi, Københavns Universitet.
- Nordgård, Torbjørn 1998: Nasjonal leksikografisk database. Status og potensial. I: Torp, Arne (red.): *NORSkrift*. Universitetet i Oslo.
- Nøklestad, Anders 1998: Statistisk disambiguerende tagging av norsk. I: Faarlund, Jan Terje/Brit Mæhlum/Torbjørn Nordgård (red.): *MONS 7. Utvalde artiklar frå det 7. Møtet Om Norsk Språk*. Oslo: Novus.
- Rosén, Victoria/Koenrad de Smedt 1998: SCARRIE. Automatisk korrekturlesing for skandinaviske språk. I: Faarlund, Jan Terje/Brit Mæhlum/Torbjørn Nordgård (red.): *MONS 7. Utvalde artiklar frå det 7. Møtet Om Norsk Språk*. Oslo: Novus.