

## Digitalt perspektiv på høyringsforslaget

### Bakgrunn

Nynodata utviklar verktøy for omsetting, korrektur og søk på bokmål og nynorsk og har så langt særleg vektlagt nynorsk. Vi har registrert alle valfrie ord- og bøyingsformer, kategorisert desse og organisert dei i ulike stilmalar som sikrar eit konsekvent språk gjennom heile dokumentet, uavhengig av forfattar. Eit eigenutvikla bøyingsssystem er kopla opp mot stilmalane.

Våre kommentarar til høyringsforslaget har dette som bakgrunn, i tillegg til vårt kjennskap til utviklingstrendar innanfor elektronisk behandling av ordlister og språk.

Etter at høyringsforslaget blei lagt fram, har vi dessutan testa det mot våre språkressursar og lagt endringane inn i den nynorske ordlista *Ordrik*. Resultatet ligg fritt tilgjengeleg for nedlasting frå heimesidene våre, slik at den som vil kan teste ut forslaget på konkrete ord. Der ligg også enkel statistikk over ord som blir endra. Ordtilfang: ca 100 000 nynorskord.

Det undrar oss litt at det digitale perspektivet er så lite synleg i høyringsforslaget. Dei fleste skriv i dag med digitale verktøy, anten det gjeld på skule, privat eller i arbeidslivet. Denne trenden blir truleg endå sterkare i tida som kjem og effekten for arbeidet med digitale verktøy burde vore vurdert og vektlagt i utgreiinga.

Ut frå eit digitalt perspektiv har vi desse generelle kommentarane til forslaget:

### Bøyingsystemet

Vi støttar nemnda si opprydding i bøyingsystemet når det gjeld verbbøyinga, særleg dei partisippformene som i dag er sideformer. Her har vi sett at mange av sideformene fell saman med hovudformer av andre ord og lagar *homografar*. Dei blir stava likt, men har ulik mening, og dataverktøy har problem med å skilje dei frå kvarandre. Det same gjeld valfri form i ubestemt eintal av svake hokjønnsord, *ei kvinna*, som er identisk med den bestemte forma i eintal. Vi er altså einige i å ta bort desse formene. Dei andre endringsforslaga for bøying betyr lite for digitale verktøy.

### Valfridom

Det er positivt at ein tek bort skiljet mellom hovudformer og sideformer. Dei vanlege stovekontrollane, f.eks. i Word, godkjenner begge deler og har såleis avgrensa nytteverdi i statsforvaltninga slik dei fungerer i dag. For vanlege stovekontrollar fører forslaget til betring.

Generell valfridom ut over dette er ikkje noko stort problem for digitale verktøy.

Stilkontrollar, f.eks. slik Nynodata utviklar, korrigerer teksten *etter ei trongare brukardefinert norm* innanfor rettskrivingsnorma. At der då blir former som ikkje er i aktiv bruk hos kvar enkelt person eller organisasjon, er ikkje noko problem for maskina.

Dette samsvarer for så vidt med det dei fleste erfarer. Ein vel sin eigen stil og bruker aktivt berre ein del av dei tillatne formene, f.eks. konsekvent e-infinitiv eller dobbel m. Det samsvarer også med ein veksande trend innan ulike organisasjonar, som lagar sine brukarstyrte ordlister med internt tilrådde eller obligatoriske former.

Problemet med valfridom blir derfor i framtida eit underordna spørsmål for dei som bruker slike ordlister eller digitale verktøy til korrektur. Stilkontrollen korrigerer formene i tråd med

den brukardefinerte norma, innanfor ramma av rettskrivinga, anten valfridommen er stor eller liten.

### **Systemrett**

Eit viktig kriterium for å lage gode digitale verktøy er at språket er mest mogleg systemrett. Unntak frå systemet er fullt mogleg å registrere, men aukar utviklingskostnadene. Det aukar også feilraten, særleg der avvika fell saman med former frå andre ord.

Moderne ordlisteproduksjon registrerer ikkje ord med bøyingsformer som ei linje med tekst, men ord kopla til system som produserer alle bøyingsformer etter behov. Det kan også finnast andre typar koplingar, som ord mot ord eller grupper av ord, eller ord kopla mot semantisk informasjon, alt registrert i strukturerte databasar. Alle ord som fell utanfor systemet må behandlast og kategoriserast for seg. Om kategoriar i systemet blir kopla til eitt eller tusen ord, kostar arbeidet til utforming og vedlikehald like mykje.

Eksempel frå høyringsforslaget som etter vår meining er unødvendige og kompliserande unntak: lang eller kort infinitivsform av ord som *be/bede*, *gle/glede*. Vi ser ikkje nokon gode grunnar for å splitte desse orda opp i to nye grupper av ord. Vi ser heller ingen gode grunnar for å skilje ut enkeltord med valfrie bøyingsformer, f.eks. *teater*. Dersom slik valfridom av ulike årsaker tvingar seg fram, bør det gjelde heile gruppa av orda, ettersom valfridommen ikkje treng representere noko problem for digitale verktøy, slik vi har vist over.

Eit anna problem i høyringsforslaget er tendensen til normering av enkeltord i staden for systematisk normering. Kva reglar skal gjelde for ord som ikkje er nemnde, men som fell inn under same kategori som ord som er lista opp? Eit tilfeldig valt eksempel: *slåstkjempe* [*slåsskjempe*] -> *slåstkjempe*. Kva då med *slåsshanske* og *slåssbikkje* og andre ord på *slåss*? Og kva med ord som er avleidde av eit ord som har blitt endra?

### **Frekvensmålingar**

Nemnda legg opp til å bruke frekvensmålingar som kriterium i normeringsarbeidet. Vi stiller spørsmål ved denne tilnæringsmåten. Rett nok er frekvens mykje brukt i digitale verktøy, særleg i søkjeverktøy, men også stavekontrollar. Fokus her er å hjelpe flest mogleg til fort å finne det dei er ute etter, og frekvente ord og former treffer naturleg nok oftare enn mindre vanlege former. Men verktøya *hindrar ikkje* brukaren å nytte dei alternative orda, som når Google høfleg spør: "Meinte du...?"

Å bruke frekvens som normeringskriterium, slik at lite brukte former blir *ulovlege*, kan etter vår meining derfor ikkje forsvarast ut frå digital tankegang. Skulle ein slik bruk eventuelt vere føremålstenleg, ville det stille store krav til korpus som frekvensen blei bygd opp frå, med eit svært breitt utval av tekstar og stor tekstmengde.

Vi opnar likevel opp for bruk av frekvens i eitt tilfelle, *homografar*, dvs ord som blir skrivne likt, men har ulik meining. Men dette er eigentleg ikkje så mykje eit normeringsspørsmål som eit problem som bør handterast av ordlisteproducentar. Der uvanlege ord som ligg i ordlista fell saman med feilskrivning av vanlege ord, bør ein kunne stille spørsmålet om ikkje lite frekvente ord bør ut, i alle fall av ordlista.

### **Ressursspørsmål**

Nemnda har ikkje trekt inn kostnader som kriterium i arbeidet, og kanskje er det rett. Men for alle oss som produserer ordlister og språkverktøy, er kostnad eit sentralt spørsmål. Alle

endringar dreg med seg ein kostnad, og mangel på ressursar kan føre til dårlegare kvalitet eller verktøy med redusert funksjonalitet. Vi meiner naturleg nok derfor at alle endringar som ikkje er nødvendige eller føremålstenlege, bør utelatast. Ein bør ikkje ta bort ord eller former berre for å redusere totalmengda. Det vil dessutan gjere språket fattigare.

Men kostnader er også eit samfunnsspørsmål. Kor mykje vil det f.eks. koste alle høgskulane våre å måtte operere med to namn, dersom *skole* går ut av rettskrivinga? Dei fleste nyttar i dag ordet *høgskole*, som er felles for bokmål og nynorsk.

For Nynodata AS

Bjørn Seljebotn  
(dagleg leiar)