Språkrådet

Kommunal- og distriktsdepartementet
Postboks 8112 Dep
0032 Oslo

Deres ref.	Vår ref.	Saksbehandler	Dato
	23/1013-1	Kristine Eide	29.11.2023

Innspill fra Språkrådet til ny nasjonal digitaliseringsstrategi

Innledning

Språkrådet er statens forvaltningsorgan i språkspørsmål og følger opp språkpolitikken slik den er formulert i Prop. 108 L (2019–2020) *Lov om språk.*
Språkrådet vil understreke at en nasjonal digitaliseringsstrategi må være en strategi for digitalisering på norsk. Det fremste målet i den norske språkpolitikken er at norsk, med skriftspråkene nynorsk og bokmål, skal være i bruk og brukelig på alle områder i samfunnet. Den nye digitaliseringsstrategien må beskrive hvordan vi som samfunn kan sikre at norsk språk blir en integrert del av digitaliseringen i både offentlig og privat sektor.
I det følgende innspillet kommer Språkrådet med konkrete forslag til hva som bør gjøres for å sikre norsk språk og språkbrukernes rettigheter i fremtidig digitalisering.

Språk og språkteknologi i ny digitaliseringsstrategi

Digitaliseringsstrategien må sørge for at brukerne av offentlige tjenester i fremtiden møter godt og klart bokmål, nynorsk og samisk. Strategien må ta utgangspunkt i at norsk er det nasjonale hovedspråket i Norge (jf. lov om språk § 4), og legge til rette for en digitalisering som støtter opp om dette, samtidig som den legger til rette for digitalisering på samiske språk, som er forvaltningsspråk i den samiske forvaltningsregionen, jf. samelovens kapittel 3, og på de nasjonale minoritetsspråkene og norsk tegnspråk.
Hittil har en stor del av digitaliseringen i Norge foregått kun på bokmål. For eksempel har statlige virksomheter anskaffet prateroboter som bare skriver bokmål, og fagtermer som er sentrale for den enkelte sektors digitalisering, er ofte utviklet bare på bokmål.
Talegjenkjenning og talesyntese fungerer best på engelsk,[1] og i den grad slik teknologi finnes for norsk, har den vært tilpasset talt standard østnorsk og bokmål.
Internasjonalt skjer utviklingen i språkteknologi og kunstig intelligens i stor grad på engelsk, og små og mellomstore språk taper terreng. Vi ser de samme tendensene over hele

[1] Taleteknologi med kunstig intelligens. Rapport fra teknologirådet, desember 2022 (https://media.wpd.digital/teknologiradet/uploads/2022/12/Endelig-versjon.pdf).

Postadresse
Postboks 1573 Vika
NO-0118 Oslo

Besøksadresse
Observatoriegata 1 B
NO-0254 Oslo

Sentralbord
+47 22 54 19 50

E-post
post@sprakradet.no
Internett
www.språkrådet.no

Organisasjonsnummer
971 527 404

verden. Denne utviklingen går på tvers av hovedformålene som ligger i den norske språklovgivningen, og får konsekvenser for demokrati, rettssikkerhet og innbyggernes tillit til det offentlige.
Store språkmodeller, som ChatGPT og Whisper, kan i prinsippet brukes på alle språk, men virker best på store språk med god tilgang på data. Språkmodeller er en helt nødvendig komponent i kunstig intelligens, og vi antar at de vil utgjøre en sentral del av en ny digitaliseringsstrategi. Strategien må sikre både at disse modellene behersker det norske språket, og at de generative modellene reflekterer norske samfunnsforhold. Modellene er ikke tillitvekkende så lenge de ikke skriver sant om norske forhold.

Oppsummert:

- Digitaliseringsstrategien må legge til grunn at utviklingen av digitale tjenester skal støtte opp om språklovens bestemmelser om at norsk er nasjonalt hovedspråk og forvaltningsspråk i Norge (jf. lov om språk §§ 1, 4).
- Digitaliseringsstrategien må gi en retning som motvirker den utviklingen som er beskrevet ovenfor, ettersom denne vil være til skade både for de språklige rettighetene til innbyggerne og for det norske språkmangfoldet.
- Digitaliseringsstrategien må beskrive hvordan norsk språk og norske samfunnsforhold skal tas vare på i store språkmodeller.
- Digitaliseringsstrategien må ha som mål at digitale løsninger og verktøy som skal brukes til kommunikasjon i Norge, må fungere på godt og riktig norsk og på både nynorsk og bokmål.

Språkdata og andre grunnlagsressurser

En ny nasjonal digitaliseringsstrategi må sørge for store, tilgjengelige grunnlagsressurser til språkteknologi. I denne sammenhengen forstår vi grunnlagsressurser både som språkdata i form av tekst, tale og leksikon (ord- og termlister) og som språkmodeller og andre verktøy til bruk i språkteknologi. Strategien må også konkretisere hvordan enkeltvirksomheter kan ta ansvar for og dele de grunnlagsressursene de har, slik at de kommer hele språksamfunnet til nytte.
Det finnes ingen teknisk eller språkvitenskapelig grunn til at norsk språkteknologi skal henge etter for eksempel engelsk språkteknologi. Kvaliteten på norsk språkteknologi er i stor grad begrenset av datamangel. Dette er ikke et særnorsk fenomen. Både i Norge og internasjonalt rapporteres det om «datatørke» i arbeidet med å utvikle språkmodeller. Det trengs mer tekstdata til de store, generelle språkmodellene, og samtidig trengs det sett med fagspesifikke språkdata til å finjustere modellene slik at de fungerer også innenfor spesifikke fag- og forvaltningsområder.
Både offentlig og privat sektor er avhengige av at det finnes gode grunnlagsressurser til utvikling av norsk språkteknologi. Mye av selve teknologien utvikles av private, men det bør være et nasjonalt ansvar å tilrettelegge grunnlagsressurser for denne utviklingen.
Enkeltvirksomheter sitter på språkdata som de ikke har delt. Det kan være fordi de ikke kan deles, på grunn av opphavsrett eller av personvernhensyn. Det kan også være fordi virksomhetene ikke vet hva språkdata er, hvor store verdier som ligger i dataene, eller hvordan de skal dele. Likevel vil alle virksomheter som planlegger å utvikle lokaltilpassede KI-verktøy, trenge språkdata fra eget domene, og de vil ha bruk for kuraterte datasett til lokal bruk, f.eks. til å finjustere generelle språkmodeller til egen virksomhet.

I hovedsak ligger to juridiske hindre i veien for å utnytte språkdataene som finnes: regelverk for opphavsrett og personvernregelverket. Opphavsretten står i veien for å utnytte både litterære tekster og sakprosa, som vi kan anta at vil være det aller beste datagrunnlaget for kunstig intelligens som skriver godt norsk. Den har også vist seg å være i veien for utnyttelse av taleressurser. Det at dataene er samlet og lagret i Norge, samt at teknologien er utviklet her, gir et bedre utgangspunkt for å løse disse utfordringene.
Den forrige digitaliseringsstrategien pekte også på deling av data som en forutsetning for digitalisering, og Språkbanken ved Nasjonalbiblioteket ble nevnt som et ressurssenter for utvikling av språkteknologi på norsk. På grunnlag av Nasjonalbibliotekets egne og innsamlede ressurser, er det i løpet av det siste året blitt laget språkmodeller som fungerer på både bokmål og nynorsk og som forstår norske dialekter. Uten denne nasjonale innsatsen ville norsk språkteknologi ha ligget enda lenger etter enn den gjør i dag. Det er viktig at dette arbeidet fortsetter.

Oppsummert:

- Arbeidet med innsamling av språkdata må fortsette. Flere typer språkdata fra flere fag, sektorer og næringer må samles inn, og en digitaliseringsstrategi bør inneholde en plan for hvordan nye, ferske språkdata kan strømme til Språkbanken.
- Digitaliseringsstrategien må inneholde tiltak for å sikre god håndtering og økt deling av språkdata og grunnlagsressurser for språkteknologi. Strategien må også beskrive hvordan offentlige og private virksomheter kan dra nytte av slike data og ressurser.
- Strategien må beskrive hvordan vi kan sikre nok språkdata fra alle domener uten å gå på akkord med personvern og opphavsrett.

IT-arkitektur

Språkrådet har gjennom mange år registrert at språklige hensyn ikke ligger systematisk til grunn for arkitekturutforming i datasystemer i offentlig sektor. Et eksempel kan være datasystemer som bare kan brukes på ett språk. Dette får konsekvenser for kommunikasjonen mellom det offentlige og brukerne av offentlige tjenester, og det fører til at det offentlige bryter statens egen språklovgivning og Stortingets vedtatte språkpolitikk.
Språkrådet erfarer at offentlige organer ikke tar høyde for krav til bruk av bokmål og nynorsk i utforming av digitale systemer. Ofte blir kostnader pekt på som årsak til at det ikke blir lagt til rette for løsninger som tar høyde for flere språk. Et typisk eksempel er statlige organer som bruker prateroboter som bare kan svare på bokmål.

Oppsummert:

- Digitaliseringsstrategien må beskrive hvordan hensynet til språklovgivningen og den norske språksituasjonen kan bli med helt fra oppstartsfasen av alle offentlige digitaliseringsprosjekter.
- Digitaliseringsstrategien må legge opp til en satsing på kvalitetssikrede språkteknologiske fellesløsninger på tvers av virksomhetene som er i tråd med kravene i språkloven og annet relevant lovverk.
- Digitaliseringsstrategien må også beskrive hvordan vi kan sikre en digitalisering som bidrar til bruk, utvikling og styrking også av de samiske språkene, og vern og fremme av de andre språkene som er omfattet av språkloven (norsk tegnspråk, kvensk, romani og romanes) (jf. lov om språk §1).

Fagspråk og terminologi

Fagspråk og terminologi er sentralt i digitaliseringen, både i grensesnittet mellom ulike datasystemer og når mennesker skal kommunisere med maskiner. For at språkteknologien skal virke innen et fagområde, må fagspråket være på plass. For eksempel produserer ikke de nye store KI-modellene korrekt helsespråk og sjøfartsterminologi av seg selv. De trenger å bli trent på store mengder tekst med godt fagspråk for å kunne produsere gode resultater. Det gjenstår mye arbeid med felles begrepsapparat. Språkrådet har sett at flere offentlige virksomheter ikke utvikler parallelle begrepsapparat på bokmål og nynorsk. Uten slike parallelle begrepsapparat kommer ikke språkteknologien til å virke like godt på nynorsk som på bokmål.

Oppsummert:

- Digitaliseringsstrategien må konkretisere hva som må gjøres for at offentlige virksomheter i større grad skal utvikle terminologi på bokmål og nynorsk, som et grunnlag for god digitalisering på alle fagområder.

Internasjonalt arbeid

Alle temaene som er tatt opp ovenfor, er problemområder også for de andre europeiske språkene. Det gjelder forholdet mellom nasjonalspråket og engelsk, datatørke, integrering av terminologi og fagspråk, arkitektur så vel som kvalitetssikring, opphavsrett og personvern. Internasjonale plattformer, digitale løsninger og flerspråklige modeller tar ofte ikke hensyn til at det er to norske skriftspråk. I den grad norsk er inkludert som et mulig språkvalg, vil det i praksis si bokmål. Dette kan både skyldes både mangelen på nynorske språkdata og utilstrekkelig kunnskap blant internasjonale utviklere om den norske språksituasjonen og den norske språklovgivningen. For at bokmål og nynorsk skal integreres som separate skriftspråk i flerspråklige digitale løsninger, plattformer og språkmodeller, må begge skriftspråkene synliggjøres på internasjonale plattformer, som for eksempel det nylig opprettede Language Data Space.

Det er også dokumentert at små språk ikke har tilgang til internasjonale plattformer på samme måte som store språk har, blant annet i ELE-rapporten *Report on the Nordic Minority Languages*[2] (ELE (European Language Equality), 2022). Rapporten inneholder anbefalte tiltak for bedre språkteknologi og bedre tilgang til internasjonale plattformer for små språk. Dette vil ha relevans for en nasjonal digitaliseringsstrategi som også skal gjelde samiske språk og kvensk. Det finnes en tilsvarende rapport for tegnspråk.[3]

Oppsummert:

- Digitaliseringsstrategien bør konkretisere hvordan norsk skal integreres i europeiske språkteknologiske fellesløsninger som for eksempel eTranslation.

---

[2] ELE (2022) *Report on the Nordic Minority Languages* (https://european-language-equality.eu/wp-content/uploads/2022/06/ELE___Deliverable_D1_38__Language_Reports_nordic_languages_-2.pdf).

[3] ELE (2023) *Report on Europe's Sign Languages* (https://european-language-equality.eu/wp-content/uploads/2023/06/ELE___Deliverable_D1_40__Europe_s_Sign_Languages_.pdf).

- Digitaliseringsstrategien må konkretisere hvordan Norge kan dra nytte av og bidra til arbeidet for digitalt språklig likeverd (Digital Language Equality[4]) som pågår i EU fram mot 2030.

Språkrådet stiller gjerne sin kompetanse på dette feltet til disposisjon i det videre arbeidet med å utarbeide en ny digitaliseringsstrategi.

Vennlig hilsen

Åse Wetås	Lars Ivar Nordal
direktør	avdelingsdirektør

Brevet er elektronisk godkjent og sendes uten underskrifter.

Mottakere:

Kommunal- og distriktsdepartementet	Postboks 8112 Dep 0032 Oslo

[4] Rehm, G. and Way, A (Eds) *European Language Equality,* Springer, 2023 (https://doi.org/10.1007/978-3-031-28819-7).