

Strategi for kunstig intelligens – innspill fra Språkrådet

Språkrådet er statens fagorgan i språkspørsmål, nasjonalt samordningsorgan for terminologi og rådgir om språkteknologi.

Språkteknologi er et av de sentrale fagfeltene innen kunstig intelligens. Språk inngår som en komponent i teknologi som brukes i søkemotorer, roboter, samfunnsanalyse, informasjonsutveksling osv. Norsk språk og språkteknologi er en forutsetning for at ny teknologi og digitalisering skal virke – innad i Norge og i informasjonsutveksling over landegrensene.

Det er avgjørende at den nye strategien for kunstig intelligens i Norge inneholder en plan for utvikling og utnyttelse av norsk språkteknologi. Den bør peke på muligheter og utfordringer som sikrer at alle innbyggerne kan bruke sitt eget språk i møte med digitale løsninger. Danmarks [nasjonale strategi for kunstig intelligens](#) peker særskilt på språkteknologi som viktig bruksområde for kunstig intelligens. Det gjør også strategiene i flere andre europeiske land, som f.eks. Østerrike og Frankrike.

Et avgjørende moment for at språkteknologi skal virke, er tilgang til ferske språkdata. De færreste er klar over hva språkdata er, og hva språkdata kan brukes til. Strategien for kunstig intelligens bør derfor inneholde en plan for hvordan språkdata skal gjøres tilgjengelige, og hvilket ansvar hver enkelt etat har for å avlevere dem til Språkbanken (se eget avsnitt lenger ned), eller gjøre dem tilgjengelige på andre måter.

1. Hva er språkteknologi?

Språkteknologi er et samlebegrep for ulike teknologier som bruker naturlig menneskespråk (i motsetning til kunstige språk, som f. eks. programmeringsspråk) som data, enten som inntutt eller som utputt. Man skiller i hovedsak mellom to typer språkteknologi: taleteknologi og prosessering av naturlig språk (NLP). Taleteknologi er teknologi som genererer tale fra tekst eller omvendt. NLP dreier seg om tolkning av semantisk innhold og er mer omfattende og mangfoldig.

2. Kunstig intelligens og språkteknologi: Muligheter og gevinster

Kunstig intelligens er allerede i bruk i språkteknologiske løsninger som f.eks. maskinoversettelse mellom språk og målformer, samtaleroboter (*chatbots*), talegjenkjenning og talesyntese (kunstig tale). Språkteknologi brukes også til en rekke andre ting, som å kontrollere fil- og arkivsystemer, kalkulere sannsynligheter (for eksempel for at brukeren har søkt på ett uttrykk heller enn et annet), tolke ytringer (for eksempel i tolking av søkefelt hvor det står 'beskriv problemet du opplever' eller 'beskriv hva du ønsker informasjon om') osv. Bruksområdene er mange og vil i framtida trolig bli langt flere.

2.1 Talegjenkjenning og talesyntese – styr ved hjelp av stemmen

I media har mye oppmerksomhet blitt viet smarthøytalere fra f.eks. Google og Amazon. Talegjenkjenningen til assistenten til Google må læres opp ved bruk av store mengder lydfiler for å kunne forstå alle variantene av norsk talespråk. I dag er bruksområdene først og fremst rettet mot underholdning i hjemmet samt enklere aktiviteter som f.eks. å slå på lys og varme. Imidlertid er det på andre områder at man ser store gevinster ved å innføre løsninger basert på talegjenkjenning og

talesyntese. Det at man kan styre navigasjonsapplikasjoner med stemmen og få opplest kjøreanvisninger av en syntetisk stemme, øker sikkerheten på veiene. Sjåføren trenger ikke å rette oppmerksomheten mot en skjerm, men vil heller kunne holde blikket på veien. Dette er også et av de bruksområdene hvor nordmenn hyppigst benytter seg av slik språkteknologi.

2.2 Kommunikasjon og datautveksling på tvers av landegrenser: automatisk oversettelse

Europakommisjonen satser på å bygge opp en grenseoverskridende digital infrastruktur i ulike sektorer, *Connecting Europe Facility* (CEF Digital). Dette EU-programmet omfatter ulike «byggeklosser», blant annet prosjektet med å bygge opp *eTranslation*, en tjeneste for maskinoversettelse mellom språkene i EU i tillegg til norsk (begge målformer) og islandsk.

Maskinoversetteren *eTranslation* er integrert i den europeiske nettportalen for tvisteløsning i forbrukersaker (*Online Dispute Resolution* (ODR)). Mye tyder på at maskinoversettere vil bli integrert i andre europeiske tjenester og vil være i stand til å supplere og i noen grad erstatte den massive mengden interne tekstoversettelser og simultanoversettelser innenfor EU og EØS og mellom EU og EØS. Dette er et svært arbeidsintensivt område som medfører store utgifter siden profesjonelle oversettere som jobber på dette feltet, er høyt spesialiserte, og mengden manuelt arbeid er enorm. Maskinoversettelse kan dermed bli svært kostnadseffektivt.

2.3 Jamstilling av bokmål og nynorsk: automatisk oversettelse

Automatisk oversettelse basert på kunstig intelligens virker ikke bare mellom ulike språk, men også mellom de to norske målformene bokmål og nynorsk.

Statlige virksomheter plikter i henhold til målloven å svare innbyggerne på den målformen som vedkommende henvender seg på. I tillegg skal publiserte tekster, inkludert nettsider, inneholde minst 25 % på den ene målformen. Imidlertid viser resultatene fra mållovsrapportering at store deler av sektoren ikke er i stand til å oppfylle kravene i loven.

Deler av offentlig sektor skal nå teste ut automatisk oversettelse av tekster mellom målformene slik at innbyggerne kan møte nettsider hos statlige organisasjoner på sin egen målform. Lykkes dette arbeidet, vil det kunne bidra til å styrke både bokmål og nynorsk som likeverdige målformer i offentlig sektor.

2.4 Velferdsteknologi

Både [Nasjonalt senter for e-helseforskning \(2018, 16\)](#) og [Teknologirådet \(2018, 40\)](#) har pekt på gevinster ved å benytte kunstig intelligens på språkteknologi for å skape nye tjenester basert på velferdsteknologi. Det gjelder for eksempel effektiv tekstanalyse og rask uthenting av informasjon fra elektroniske pasientjournaler, som kan bestå av flere hundre dokumenter med ustrukturert tekst. Det sparer helsepersonell tid i kritiske situasjoner. I tillegg åpner teknologien opp for mer persontilpasset behandling.

Rapporten fra Teknologirådet peker også på språkteknologi basert på tale til tekst. Helsepersonell kan diktere informasjon direkte inn i pasientjournalen ved hjelp av taleteknologi som lærer å forstå brukerens egen dialekt og terminologi. Denne teknologien er allerede på markedet og i utstrakt bruk.

2.5 Kundestøtte

Et annet bruksområde for språkteknologi og kunstig intelligens er samtaleroboter (*chatbots*) som kan svare på spørsmål fra publikum døgnet rundt. Skatteetatens erfaringer med samtaleroboter er særlig

positive knyttet til enkle og hyppige spørsmål, ifølge Teknologirådet (2018, 39). I utviklingen av samtaleroboten benyttet etaten seg av maskinlæring på et treningssett med ulike typer spørsmål og intensjoner for å forstå det brukerne lurer på.

Også Statens lånekasse for utdanning har sett på mulighetene for å innføre samtaleroboter i virksomheten. En utprøving ble foretatt i 2017, da som et verktøy internt i organisasjonen. Rapporten peker på at det var et gjennomgående problem at norsk språk ikke var støttet.

2.6 Robotisering av saksbehandling

Enklere saksbehandling i offentlig og privat sektor kan effektiviseres gjennom robotisering, som kan overføre arbeidsprosesser fra mennesker til roboter. Det vil ofte kreve språkteknologiske verktøy til f.eks. tekstanalyse. Verktøy for tekstanalyse kan forstå tekst og kategorisere informasjonen i den. Slike verktøy kan derfor hente ut nødvendig og relevant informasjon fra fritekst som kundene har registrert i elektroniske skjemaer eller andre steder.

Robotisering av enkle saker vil friggi tid for saksbehandlere slik at de kan bruke den på mer kompliserte saker.

3. Hva skal til?

Alle disse tjenestene er i dag reelle, og mulighetene er langt flere. Men for å kunne forbedre kvaliteten og legge til rette for nye løsninger er utviklerne og innovasjonsbransjen helt avhengige av tilgang til store mengder språkdata, både skriftlige (bokmål og nynorsk) og muntlige (dialekter og uttalevariasjon). Automatisk oversettelse blir matet med parallelle tekster og fagspesifikk terminologi på de to målformene for at tekstene skal ha høy nok kvalitet og inneholde riktig og presis fagterminologi. Med andre ord trengs det flere språkdata for at tjenestene skal nå alle innbyggerne i Norge, uansett hvilken dialekt de snakker eller hvilken målform de skriver på.

3.1 Hvordan kan språkdataene samles inn?

Språkrådet samarbeider først og fremst med Nasjonalbiblioteket, men også Difi, om å samle inn språkdata og gjøre dem allment tilgjengelig i [Språkbanken ved Nasjonalbiblioteket](#). De som utvikler de språkteknologiske tjenestene og produktene, skal kunne hente ut nødvendige språkdata fra banken og kunne gjenbruke dem fritt for videreutvikling av språkteknologi basert på kunstig intelligens. En ekstrabevilling fra Kulturdepartementet på 10 millioner kroner har satt fart i dette arbeidet. Likevel er det nødvendig å se det sektorovergrepene aspektet ved språkteknologien. Hver sektor som vil ta i bruk språkteknologi, trenger områdespesifikke språkdata. Enkeltsektorene må derfor selv bli pålagt et ansvar for å samle inn egne data, det være seg tekst, oversettelsesminner eller fagtermer. Mange offentlige virksomheter er ikke klar over at språkdataene som de disponerer, er nødvendige for å utvikle god norsk språkteknologi, og bevisstheten om dette må økes. Offentlig sektor produserer i dag store mengder tekst, og offentlige virksomheter som f.eks. NRK og Stortinget sitter på lydopptak som utgjør en gullgrube for språkteknologisk utvikling.

3.2 Ordlistor og termlister

I tillegg til store tekstmengder trengs det kvalitetssikrede ord- og termlister for å utvikle gode språkteknologiske tjenester. Det gjelder særlig om språkteknologien skal anvendes på et særskilt, faglig område, for eksempel i helsevesenet. Da vil man trenge fagspesifikke termlister for helse- og medisinbegreper. Med andre ord må språkteknologien beherske fagområdetets fagterminologi.

I dag pågår det et omfattende digitaliseringsarbeid i offentlig sektor der begreper i datasett skal beskrives og defineres når de blir gjort tilgjengelige. Dette arbeidet er forankret i St.meld. 27 (2015–2016) *Digital agenda* og andre styringsdokumenter. Det blir nå utviklet termlister med begrepsdefinisjoner for de ulike forvaltningsområdene som datasettene gjelder, og disse termlistene blir tilgjengeliggjort i [Felles begrepskatalog](#), som er utviklet av Brønnøysundregistrene. Om termlistene er flerspråklige, kan de tjene som innputt i maskinlæring og gjenbrukes i språkteknologiske løsninger for de aktuelle forvaltningsområdene.

Det samme gjelder termlister som blir utviklet i andre samfunnssektorer, f.eks. termlister for studenter i universitets- og høyskolesektoren og termlister generert fra tekniske standarder på en rekke fagområder. Det er da helt avgjørende at det ikke knyttes juridiske vilkår til termlistene som hindrer språkteknologisk gjenbruk.

3.3. Utdanning

Spesialisert utdanning av lingvister ved norske utdanningsinstitusjoner er nødvendig for utvikling av norsk språkteknologi av høy kvalitet. Datalingvister med norsk som morsmål og med høy kompetanse på nynorsk og bokmål, er høyt etterspurte i bransjen, og dessverre setter mangelen på kvalifiserte lingvister en stopper for utvikling av produkter på norsk hos store internasjonale utviklere.

Det er derfor avgjørende for norsk språkteknologi at det eksisterer et godt nok utdanningstilbud i faget ved minst ett av universitetene i landet. Kunnskapsdepartementet bør derfor ta initiativ til å opprette et slikt fagtilbud for at kompetansebehovet blir dekt i framtida. Det er også vesentlig at det finnes utdanningstilbud i fag som leverer språkdata til språkteknologi, f.eks. terminologilære. I dag finnes det ikke lenger et undervisningstilbud i dette faget, og det hemmer tilgangen til terminologilister til språkteknologisk utvikling.

4 Avsluttende kommentar

Språkrådet er svært glad for at digitaliseringsministeren og Kommunal- og moderniseringsdepartementet har tatt initiativ til en nasjonal strategi for kunstig intelligens. Språkteknologi på norsk, begge målformer, er avgjørende for at kunstig intelligens skal komme til nytte for hele det norske samfunnslivet.

Litteratur

Dansk sprognævn (2019) *Dansk sprogteknologi i verdensklasse rapport om språkteknologi*.
<https://dsn.dk/udgivelser/sproгнаevnets-udgivelser/sproгнаevnets-rapporter/sprogteknologi-i-verdensklasse>

Den danske regjeringen (2019) *National strategi for kunstig intelligens*.
https://digst.dk/media/19302/national_strategi_for_kunstig_intelligens_final.pdf

Europakommisjonen (2019) *eTranslation*.
<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>

Nasjonalt senter for e-helseforskning (2018) *Health analytics. Kunstig intelligens – nye muligheter for helsetjenesten*. https://ehealthresearch.no/files/documents/Prosjektrapporter/NSE-rapport_2018-03_Health_analytics_Kunstig_intelligens.pdf

Språkrådet (2018) *Språk i Norge – kultur og infrastruktur*.
https://www.sprakradet.no/globalassets/diverse/sprak-i-norge_web.pdf

Statens lånekasse for utdanning (2017) *Sluttrapport fra konseptutredning av muligheter for effektivisering ved bruk av kunstig intelligens*.
<https://lanekassen.no/Global/Om%20organisasjonen/Sluttrapport%20konseptutredning%20kunstig%20intelligens%20L%C3%A5nekassen%202017.pdf>

Teknologirådet (2018) *Kunstig intelligens – Muligheter, utfordringer og en plan for Norge*.
<https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>