



Talegjenkjenningsteknologi - en kunnskapsinnsamling



Prosjekt: Tale-til-tekst (3T)

Tale-til-tekst-prosjektet skulle levere status for tale-til-tekst-teknologi for norsk, samisk og tegnspråk og tilby en lavterskel veiledning om hvilke muligheter som allerede finnes på markedet og hvordan bruke disse mulighetene. Denne rapporten er del av leveransen.

Veiledningen for bruk av transkriberingsteknologi finnes på hlf.no/taletiltekst

Finansiering: Barne-, ungdoms- og familiedirektorater gjennom tilskuddsordningen [Tilskudd til kunnskapsutvikling, kompetanseheving og informasjon innen universell utforming](#)

Dato: Desember 2022

Prosjektleder: Sigrid Skavlid, e-post: sigrid@universellutforming.no

Prosjektdeltakere Hørselshemmedes Landsforbund: Øydis Lebiko og Anita Hvambstahl



Universell Utforming AS

Sehesteds gate 6, 0164 OSLO

www.universellutforming.no

Innhold

1. Introduksjon	4
2. Metode	5
3. Hva er talegjenkjenningsteknologi?	6
Naturlig tale, diktering og stemmestyring.....	7
Språkmodell, trebank og domenespesifikke ordlister.....	7
Utviklet ved hjelp av kunstig intelligens	9
Lærende språkteknologiske systemer	10
4. Språk og digitalisering	12
Språk i verden	12
Språk i Norge	13
Digitalisering av norske språk	14
Norsk språkteknologi (nynorsk og bokmål).....	16
Samisk språkteknologi	17
Tegn-til-tale og tale-til-tegn-teknologi	18
Språkdød.....	19
5. Noen prosjekter innen talegjenkjenningsteknologi.....	22
6. utfordringer.....	25
Manglende språkdata	25
Manglende ressurser	25
Kommersielle hensyn	25
Spranget mellom naturlig, spontan tale og transkribert formalspråk	25
Bakgrunnsstøy	25
Domenespesifikke treningsdata	26
Stemmegjenkjenning.....	26
Banale avatarer og utgangspunkt i hørende sine behov.....	27
Personvern og opphavsrett	27
Bias i treningsdata	27
Dialekt.....	28
Valgfrihet og likestilling i skriftspråk.....	28
7. Oppsummering.....	29

1. Introduksjon

Taleteknologi er en gren innen kunstig intelligens som har skutt fart de siste årene. For hørselshemmede har den generelle teknologiske revolusjonen de siste ti-årene medført mer likestilling og mer frihet. E-post, SMS og skriftlig digital kommunikasjon i private og kommersielle sammenhenger har ført til mer skriftlig kommunikasjon som er svært nyttig for hørselshemmede. De nye digitale møteformene som har blitt den nye normalen både under og etter korona-nedstengingen, har ført til nye utfordringer for hørselshemmede. Direkte taletranskribering vil kunne hjelpe hørselshemmede i digitale møter og konsultasjoner hvis transkriberingen er god nok. Automatisk transkribering vil også kunne erstatte skrivetolker og tegnspråktolker når kvaliteten er tilfredsstillende.

Talegjenkjenning er også et verktøy for personer som har problemer med å bruke vanlig tastatur og mus. Denne gruppen kan bruke talegjenkjenning for å kunne styre datamaskinen eller mobiltelefonen og for å diktere tekst ved hjelp av stemmen.

Dyslektikere og andre med skrivevansker samt synshemmede kan bruke tale-til-tekst-teknologi for diktering. Det er også knyttet forventning til enklere tilgang til referatskriving, journaler, teksting av videoer, direkte TV-sendinger mv. Det er uten tvil stor nytte for mange å ha tilgang på god tale-til-tekst-teknologi.

Hvor langt er arbeidet med talegjenkjenningsteknologi kommet i Norge? Hva er status? Er dette et reelt hjelpemiddel for hørselshemmede i dag? Hvordan står det til med norsk, samisk og tegnspråk når det gjelder talegjenkjenning og transkribering?

Dette var bakgrunnen for søknaden til Bufdirs tilskuddsordning for kunnskapsutvikling, kompetanseheving og informasjon innen universell utforming. Sammen med Hørselshemmedes Landsforbund (HLF) fikk Universell Utforming AS midler til å undersøke status for tale-til-tekst-teknologi. Parallelt med denne kunnskapsinnsamlingen har vi gjennomført egentester og brukertester av tale-til-tekst-produkter. Som del av prosjektet er det laget en guide for hvordan bruke ulike tale-til-tekst-funksjoner. Denne guiden er publisert på HLFs nettsider. ¹

Rapporten er et forsøk på å si noe om hva situasjonen for tale-til-tekst er i en norsk kontekst anno 2022 og hvilke utfordringer man står overfor.

¹ <https://www.hlf.no/taletiltekst>

2. Metode

Måten vi har arbeidet med prosjektet og denne rapporten på, har vært tredelt.

I oppstartsfasen brukte vi nettsøk aktivt både for å undersøke hva som fantes av produkter og informasjon om hva talegjenkjenning er og hvordan det fungerer.

Dernest testet vi produkter selv for å komme fram til hvilke som kunne være egnete å gå videre med til å brukerteste med hørselshemmede. Dessverre viste våre egne tester at det ikke fantes egnete produkter for testing i alle de brukssituasjonene vi ønsket å undersøke og som vi anså som relevant for hørselshemmede.

Parallelt med nettsøk og offentlig tilgjengelig informasjon, har vi innhentet informasjon om talegjenkjenningsteknologi gjennom møter i Forum for taleteknologi som drives av Hørselhemmedes Landsforbund. Dette er en uformell gruppe som er åpen for alle som er interessert eller arbeider med talegjenkjenning eller generell språkteknologi i Norge. Møtene samler mellom 20 og 40 personer 3-4 ganger årlig. Ulike prosjekter presenteres og informasjon om tale-til-tekst på norsk utveksles.

I tillegg er kunnskap innhentet fra ulike podkaster om taleteknologi og kunstig intelligens, digitalt tilgjengelige universitetsforelesninger og konferanser om språkteknologi.

Informasjonsinnhenting har blitt supplert med semistrukturerte intervjuer, e-poster, telefoner og møter med en rekke fagpersoner innen ulike områder fra forskning til leverandører eller mulige leverandører av talegjenkjenningsprodukter. Confrere, Davvin-gruppen ved Universitetet i Tromsø, DeepSign, Dictus, HLF Briskeby utvikling, Lingit, NAV, NORCE, Norsk Regnesentral, NTNU, Omilon, Språkbanken/Nasjonalt biblioteket, Språkrådet, Stortinget, Teknologirådet, Universitetet i Oslo og Vitec er blant dem vi har vært i kontakt med i løpet av prosjektet.

Når det gjelder digitalisering og tegnspråk har vi tatt kontakt med regjeringens tegnspråkutvalg² som ble nedsatt i 2021 og som skal levere innstilling i 2023. Da det ikke framgår tydelig av utvalgets mandat om tegnspråktranskribering er noe utvalget skal ta opp, tok vi kontakt med dem, men vi har ikke fått avklaring på om de vil komme med innspill på tegntechnologi.

² <https://nettsteder.regjeringen.no/tegnsprakutvalget/>

3. Hva er talegjenkjenningsteknologi?

Talegjenkjenning er en gren av språkteknologien og er ofte basert på kunstig intelligens (KI).

Språkteknologi generelt kan deles inn i:

- a) transkribering, det vil si tekst produsert digitalt basert på tale
- b) maskinoversettelse, det vil si at et dataprogram kan oversette fra et språk til et annet som for eksempel Google Translate eller Bing Translate
- c) korrekturprogram og skrivestøtte som for eksempel i Microsoft Word
- d) informasjonsgjenfinning, det vil si søk i store datamengder som for eksempel Google søk
- e) interaktive systemer gjennom stemmestyring, for eksempel bruk av Siri på iPhone eller Amazon Alexa³

I dette prosjektet tar vi for oss det første punktet, det vil si prosessen fra tale til tekst.

Talegjenkjenning er en teknologi som gjennom å kutte opp talen i mindre biter, klarer å omgjøre denne til skriftlig tekst.⁴ Dette gjøres ved å overføre talelyden til et digitalt format hvor den deles opp i ulike språklyder eller foner. En fon korresponderer i det vesentligste med vokaler og konsonanter og det er identifisert om lag 100 ulike foner i de ulike språkene rundt i verden.

I talespråk er fonem en grunnleggende, teoretisk enhet – det minste lydsegmentet som kan endre betydningen av et ord.⁵ Når en fon, eller en gruppe foner, utgjør en meningsgivende forskjell i et språk, kalles det et fonem.⁶ I talegjenkjenning er det da viktig å skille ut de ulike lydene, som høres ulike ut, men som betyr det samme, som for eksempel rulle-r og skarre-r i norsk.

For å finne ut om språklyder har betydningskillende funksjon, forsøker man å finne såkalte minimale par. Dette er par av ord som bare skiller seg fra hverandre ved at ett fonem er forskjellig. For eksempel er ordparet lus og rus et minimalt par, som viser at l og r er to fonemer i norsk, siden vi ved å bytte ut den ene språklyden med den andre får et nytt ord med ny betydning.

I talegjenkjenning må maskinen kunne skille de ulike lydene, ulike måter å uttale disse på, og overføre det til et standard skriftspråk. Språkteknologi og talegjenkjenning er en del av utviklingen innen KI som har utviklet seg raskt de siste årene. Språkteknologi benevnes også som Natural Language Processing (NLP) eller datalingvistikk i faglitteratur.

Naturlige språk (Natural Language) er språk utviklet av mennesker. Det som skiller naturlige språk fra programmeringsspråk, er at det er flertydig og at mange ord kan bety mye forskjellig alt ut fra hvilken kontekst de står i. Programmeringsspråk er entydige, og det er hensikten med slike språk. Når du klikker på en lenke på en nettside, er det helt entydig i

³ Lison, Pierre (2020, september). *Dette er sannsynligvis viktig – Hva er språkteknologi (eller NLP)?* [Audio podkast] Hentet fra <https://open.spotify.com/episode/52DlcDED8GpZnoAqpTjT1Y>

⁴ Tørresen, Jim (2013) *Hva er kunstig intelligens*, Universitetsforlaget, s. 75

⁵ Fonem. (u.å.) *Wikipedia*, tilgjengelig på <https://no.wikipedia.org/wiki/Fonem>, lest 29.9.2021

⁶ Fon. (u.å.) *Wikipedia*, tilgjengelig på <https://no.wikipedia.org/wiki/Fon>, lest 11.10.2022

html-koden hva som skal skje, nemlig at en ny nettside skal åpnes eller at e-postklienten din skal åpnes hvis lenken er kodet med at dette skal skje.

Gjennom kontekst og erfaring lærer mennesker seg hva som er den riktige betydningen av et ord i en setning, helt uten at vi tenker over det. Som for eksempel det norske ordet «rett» som kan bety mye ulikt. Fra barnsben lærer vi de ulike betydningene av dette ordet og som regel tenker vi ikke på at ordet har ulik betydning. Hvis vi forsøker å oversette det til engelsk blir det veldig klart. Her kan vi velge mellom straight, correct, law, court, directly, meal, dish, m. fl. Vi forstår at når vi har en tallerken med mat foran oss er det en «rett», mens «rett og galt» har en annen betydning.⁷ For maskiner er ikke dette så enkelt da de ikke har menneskelig hjerne og erfaring og må derfor kode om ordene. For automatisk oversettelse er dette en utfordring og i tidlige versjoner av Google Translate kunne setningen «Han måtte møte i retten» bli oversatt til «He had to meet in the dish». Oversettelsen er for nordmenn komisk og vi forstår med en gang hva som har gått galt, fordi vi vet at ordet «rett» har flere betydninger, men for en engelsktalende er det ikke morsomt. Derfor blir det å forstå konteksten også viktig for maskiner, ellers blir oversettelsene eller ordgjettingen i transkribering feilaktige.

Naturlig tale, diktering og stemmestyring

Talegjenkjenningsteknologi er altså den delen av språkteknologien som prosesserer det talte språk. Talegjenkjenning må ikke forveksles med stemmegjenkjenning som er en form for biometri hvor maskinen gjenkjenner hvem som snakker og kan verifisere rett bruker på samme måte som ved et fingeravtrykk. Talegjenkjenning er å overføre hva som blir sagt til tekst (talertranskripsjon) eller å utføre kommandoer med stemmen. Det siste blir gjerne kalt stemmestyring som forutsetter talegjenkjenning.

Naturlig, spontan tale eller samtale mellom mennesker inneholder mange fyllord, gjentakelser og halve og ufullstendige setninger. Å få denne type tale overført til en korrekt tekst med hele setninger med tegnsetting er det foreløpig kun mennesker som kan få til. Mellommenneskelig kommunikasjon og språk er svært forskjellig fra tekster som vi finner i bøker eller aviser.

Diktering er når man leser inn teksten med tanke på transkribering. Noen transkriberingsløsninger setter inn tegn som komma eller punktum ved pauser, mens andre krever at man må lese det inn. Naturlig, spontan tale slik som vi gjør hele tiden med hverandre, er veldig ulik hvordan man dikterer der man leser inn setninger man har klare i hodet.

Språkmodell, trebank og domenespesifikke ordlister

Trebank er en type tekstkorpus som er annotert med syntaktisk informasjon. De syntaktiske analysene er representert som en trestruktur, derav navnet trebank.⁸ En lingvist tar for seg

⁷ Øvreid, Lilja (26.8.2019) *IN1140 Introduksjon til språkteknologi*. Hentet fra <https://www.youtube.com/watch?v=dHzweP3xyA0>

⁸ Trebank (u.å) *Wikipedia*. Tilgjengelig via <https://nn.wikipedia.org/wiki/Trebank>. Lest 14.11.2022

et datamateriale og merker opp hva som er subjekt, verb, objekt mv i setningen og hvordan disse ordene forholder seg til hverandre.

Et tekstkorpus eller korpus er en samling av maskinelt lesbare tekster som tjener et bestemt formål, samt overholder regler for innhold og format. For språkforskere er tekstkorpus et stort og strukturert sett av tekster (vanligvis elektronisk lagret og prosessert), som benyttes som utgangspunkt for analyser av språk. Til støtte for analysen er tekstenes ord ofte annoterte, det vil si forsynt med opplysning om ordklasse, ordstamme eller faste vendinger. På grunnlag av dette analyseres setningers grammatiske struktur. Dette brukes til oversettelse og talegjenkjenning.⁹

Mens man fram til 2000-tallet baserte seg på kodete, regelbaserte systemer slik som trebanker tok etter hvert maskinlæring over. Det vi si at systemet tilegner seg kunnskap om språk automatisk ved å gjenkjenne statistiske mønstre i store datamengder.

Språkmodellene læres opp til å predikere hva som er neste ord i en sammenheng enten neste, de to neste osv. opptil fem for de mest avanserte. Når man snakker og man har et talegjenkjenningsprogram aktivert, vil programmet gjette på hvilke ord som sies. Hvilke ord som foreslås er avhengig av sammenheng og frekvens og hva som da er mest sannsynlig.

I noen fag brukes ord veldig spesifikt og annerledes enn i alminnelig språkbruk som for eksempel i medisinsk kontekst. Da har det vist seg å være nyttig å utvikle såkalte domenespesifikke modeller hvor systemet trenes på tekster fra dette fagfeltet slik at det skal fungere best mulig for bruken innenfor dette området.

Da Stortinget tok i bruk talegjenkjenningsteknologi i 2021 for å lage førsteutkast av møtereferat¹⁰, fikk de laget en domenespesifikk løsning slik at systemet ble trent på tekster spesifikke for bruk på Stortinget.

En test av automatisk transkribering av et kommunestyremøte i Lillestrøm i forbindelse med dette prosjektet, ble utfordringen med frekvente ord og domenespesifikke ord tydelig. På kommunestyremøtet ble: «De som er for, viser det med stemmetegn» transkribert til «De som er for, viser det med stjernetegn», «gå til votering» ble til «gå til hotellet» og «hovedutvalg» ble transkribert til «hodeplagg».

Stortingets transkriberingsløsning er trent på materiale fra Stortinget og har derfor bedre forutsetninger for å gjenkjenne ordene som er typiske fra en politisk møtesammenheng. Transkriberingsløsningen som ble brukt på møtet på Lillestrøm, var knyttet til Googles talegjenkjenning. Den hadde da ikke samme forutsetning for å forstå konteksten og transkribere korrekt, siden ord som «stemmetegn» og «votering» ikke er så mye i alminnelig bruk i de sammenhengene Googles talegjenkjenning har erfaring fra.

⁹ Tekstkorpus (u.å) Wikipedia. Tilgjengelig på <https://no.wikipedia.org/wiki/Tekstkorpus>. Lest 2.12.2022

¹⁰ Kunstig intelligens hjelper Stortinget med referatskriving (2022) Bufdir.no, tilgjengelig på https://www.bufdir.no/uu/Nytt/kunstig_intelligens_hjelper_stortinget_med_referatskriving/

Utviklet ved hjelp av kunstig intelligens

Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål. Enkelte KI-systemer kan også tilpasse seg gjennom å analysere og ta hensyn til hvordan tidligere handlinger har påvirket omgivelsene. Slik definerer Nasjonal strategi for kunstig intelligens begrepet.¹¹

Videre kan vi skille mellom kunstig intelligens med ulike tilnærminger og teknikker:

- maskinlæring (inkludert eksempelvis dyplæring og forsterkende læring),
- maskinresonnering (inkludert planlegging, søk og optimering)
- robotikk (som kontroll, sensorer og integrasjon med andre teknologier i cyber-fysiske systemer).

Maskinlæring er en gruppe matematiske og statistiske teknikker som ved hjelp av eksempler «trener» maskiner til å løse en oppgave, og som gjør at maskinen blir bedre med erfaring. Dette skiller seg fra tradisjonell programmering, også kalt regelstyrte systemer, hvor maskinen løser en oppgave ved hjelp av et sett med presise instruksjoner.¹²

På 1950-tallet definerte matematikeren Alan Turing hva som må til for å kalle en maskin intelligent. Testen for systemet er senere blitt hetende Turing-testen. Testen går ut på at et menneske (dommer) kommuniserer med en datamaskin eller et annet menneske ved hjelp av kun tastatur og skjerm. Kommunikasjonen kan være om alle mulige tema, og vare mange timer. Hvis dommeren ikke kan avgjøre om det er et menneske eller en maskin han kommuniserer med, er testen for intelligent system bestått.¹³

Ifølge Jim Tørresen i «Hva er kunstig intelligens»¹⁴, må en maskin ha visse egenskaper for å kunne etterligne et menneske:

«Den må kunne:

- gjenkjenne og generere *naturlig språk* for å kommunisere med et menneske
- *lagre informasjon* for å representere kunnskap den har fått eller får
- *resonner* basert på lagret informasjon og trekke nye konklusjoner
- kunne *lære* for å tilpasse seg nye omstendigheter og trekke ut mønstre[....]

For å passere denne testen må systemet i tillegg:

- ha datasyn for å oppfatte objekter
- bestå av en robot som kan manipulere objekter og flytte seg omkring»

¹¹ Kommunal- og moderniseringsdepartementet (2020). Nasjonal strategi for kunstig intelligens, (2020), tilgjengelig på <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>

¹² Teknologirådet (2017) *Saken forklart - Kunstig intelligens: smart eller skremmende?* Tilgjengelig på <https://teknologiradet.no/kunstig-intelligens-smart-eller-skremmende/>

¹³ Turing test (u.å) – *Wikipedia*, tilgjengelig på <https://no.wikipedia.org/wiki/Turingtest>, lest 9.11.2022

¹⁴ Tørresen, Jim (2013) *Hva er kunstig intelligens*, Universitetsforlaget, s. 17

Tørresen mener at disse seks egenskapene rommer det meste som kalles KI i dag, men forskere har i liten grad prioritert å lage slike menneskelignende systemer. Menneskelig intelligens er særskilt ved at den kan trekke veksler på et område det kjenner, til nye områder. Dette har kunstig intelligens større problemer med. Dette gjelder også innen tale-til-tekst-teknologi.

Vi kan skille mellom enkel og kompleks KI, eller spesifikk eller generell kunstig intelligens. Med generell eller kompleks kunstig intelligens menes intelligens som ligner på menneskelig intelligens, som kan kombinere en rekke sanseinntrykk, erfaring og ny informasjon, og utføre komplekse handlinger på bakgrunn av dette i nye og ukjente situasjoner slik som Turing-testen foreskriver. Generell kunstig intelligens er en stund frem i tid, om den noensinne kommer.

Det som vi snakker om i dag når det gjelder kunstig intelligens, er spesifikk KI, det vil si KI som skal løse spesifikke oppgaver som for eksempel finne den beste nettsiden ut fra din søke-term basert på din søkehistorikk (interesser) og lokasjon, gjenkjenne ansikter ved fotografering og fokusere eller å transkribere det du sier til tekst.¹⁵

Lærende språkteknologiske systemer

På samme måte som mennesker, må programmet lære. Systemet må trene på å gjenkjenne ord og vendinger og de mange ulike måtene dette blir uttalt på. Systemet må også bli korrigert for å bli bedre. Barn lærer ved å høre på andre, gjennom peking og lyd (for eksempel peke på hund og si «voff-voff», etter hvert blir dette til «se, hund») og bli korrigert.

Datamaskiner må læres opp på en lignende måte. Innen taleteknologi må systemet trenes på det som kalles *korpus*. Dette er språklige ytringer enten skriftlig eller muntlig som har framkommet gjennom naturlig tale/skrift og som digitaliseres.¹⁶ Systemet lærer også hva som er det mest sannsynlige ordet i setningen ut fra sammenhengen og hvor vanlig ordet er i sammenheng med de andre ordene i setningen. Over tid lærer maskinen at den vanligste sammenhengen av ordet «retten» i setningen «han måtte møte i retten» er i betydningen domstol. Da blir maskinoversettelse enklere og Google oversetter til «He had to meet in court».

I 2022 har Google Translate 500 millioner daglige brukere.¹⁷ Dette gir en enorm datamengde. Systemene blir da også flinkere og flinkere til å gjette riktige ord ut fra hva som er vanlig i en kontekst på samme måte som voksne mennesker kan gjette seg til hva som blir sagt ut fra hvem man snakker med og hva samtalen dreier seg om, selv om man ikke får med seg alle ordene.

¹⁵ Skavlid, Sigrud (2022) Kunstig intelligens og sensorteknologi – kunnskapsinnsamling. Tilgjengelig på https://www.medialt.no/pub/Statusrapport_kunstig_intelligens_og_sensorteknologi.docx

¹⁶ Andersen, Gisle: korpus - språkvitenskap i Store norske leksikon på snl.no. Hentet 12. oktober 2022 fra https://snl.no/korpus_-_spr%C3%A5kvitenskap

¹⁷ Wise, Jason (2022) *How many people use Goggle Translate in 2022?* Earthweb. Tilgjengelig på <https://earthweb.com/how-many-people-use-google-translate/>

Mens barn blir korrigert av andre mennesker og lærer gjennom dette, lærer språkteknologiske modeller gjennom bruk og læringsalgoritmer. Først må man ha et kodet datamaterialet gjennom et korpus som en grunnopplæring. En algoritme er en oppskrift på hvordan systemet skal lære og hva det skal lære videre gjennom bruk.

Ved å mate systemet med språkmateriale vil det lære gjennom de algoritmene det er satt opp med og hvordan systemet blir brukt.

Du har sikkert sett at du ved automatisk oversettelse kan få spørsmål om «Hvordan fungerte denne oversettelsen». Når du gir tilbakemelding, lærer systemet. Hvis man ikke retter når man har diktet, lærer også systemet. Systemet lærer også gjennom statistikk. Systemet vil anta at et ord er det riktige, jo oftere det brukes i en gitt sammenheng. Slik sett er språkteknologi et fag med mange grener som lingvistikk, informatikk, statistikk og logikk. Gjennom programmering av algoritmer som er selvlærende, lærer systemet hva som sannsynligvis er riktig. Akkurat som et menneske ville kunne lært et annet språk ved å se millioner av timer med filmer på det språket hvis man hadde hatt tid til det. For maskiner med enorm datakraft derimot, tar det mindre tid å skanne tilsvarende materiale og sette det sammen til et språk som et menneske kanskje hadde trengt 100 år på. Derfor lærer maskinene eller språkmodellene språkene, å gjenkjenne ordene og å skrive dem ut riktig. Kunstig intelligente systemer er gode til å finne mønstre i datamateriale.

4. Språk og digitalisering

Språk i verden

Det finnes mange ulike språk i verden. Anslagene ligger på mellom 6000 og 7500. Siden det kan være glidende overgang mellom en dialekt og det som kan defineres som et eget språk, finnes det ikke et helt nøyaktig antall. Fordelingen mellom språkene på verdensbasis er svært ujevn. Nesten halvparten av jordens befolkning har et av de fem vanligste språkene i verden som morsmål. Dette er mandarinkinesisk, spansk, engelsk, hindi og arabisk. Utvider man antallet til 100 språk, omfatter det om lag 95% av menneskene på jorden i dag.¹⁸

Databasen Ethnologue oppgir at det i 2022 finnes 7151 språk i verden. Omtrent 40% av språkene er definert som truet. Ofte har disse språkene færre enn 1000 gjenværende mennesker som snakker språket.¹⁹

Amazon Transcribe kan i dag gjenkjenne automatisk 37 av verdens språk.²⁰ Dette gjør at man kan bygge talegjenkjenningsteknologi ved hjelp Amazon Transcribe uten at noen mennesker trenger å definere for systemet hvilket språk det er.

Meta (moderselskapet til Facebook) har satt seg som mål å kunne oversette tale-til-tale for 100 av verdens språk i sanntid. De arbeider også med å kunne oversette i sanntid språk uten fast skriftspråk.²¹ Ikke alle muntlige språk har et skriftlig uttrykk som for eksempel ulike plystrespråk. Selskapet har tidligere satt seg som mål at deres system med kunstig intelligens skal oversette alle skriftlige språk med maskinoversettelse (tekst til tekst). Blant andre Meta og Massachusetts Institute of Technology arbeider med å lage språkmodeller som kan lære av mindre mengde annoterte data og også lære å gjenkjenne språk uten skriftspråk.²²

¹⁸ Simonsen, Hanne Gram; Kjøll, Georg; Faarlund, Jan Terje: språk i Store norske leksikon på snl.no. Hentet 9. desember 2022 fra <https://snl.no/spr%C3%A5k>

¹⁹ How many languages are there in the world? (2022) Hentet fra 9. desember 2022 fra <https://www.ethnologue.com/guides/how-many-languages>

²⁰ Amazon (2022) Amazon Transcribe now supports automatic language identification for multi-lingual audio, pressemelding 1.7.2022, tilgjengelig på <https://aws.amazon.com/about-aws/whats-new/2022/06/amazon-transcribe-supports-automatic-language-identification-multi-lingual-audio/>

²¹ Edunov, Sergey, Paco Guzman, Juan Pino., Angela Fan (2022) Teaching AI to translate 100s of spoken and written languages in real time. ai.facebook.com, tilgjengelig på <https://ai.facebook.com/blog/teaching-ai-to-translate-100s-of-spoken-and-written-languages-in-real-time/>

²² Massachusetts Institute for Technology (2021) Toward speech recognition for uncommon spoken languages. Pressemelding. Tilgjengelig på <https://news.mit.edu/2021/speech-recognition-uncommon-languages-1104>

Språk i Norge

Norsk er det nasjonale hovedspråket i Norge. Norsk har to ulike skriftspråk – nynorsk og bokmål. Samiske språk og norsk er likeverdige språk og likestilte etter sameloven kapittel 3. De samiske språkene som er mest brukt i Norge i dag er nordsamisk, sørsamisk og lulesamisk.²³

I henhold til språklova er kvensk, romani og romanes nasjonale minoritetsspråk i Norge. Som språklige og kulturelle uttrykk er kvensk, romani og romanes likeverdige med norsk. Romani er språket til romanifolket (taterne), mens romanes snakkes av romene (tidligere kalt sigøynere).

Norsk tegnspråk er det nasjonale tegnspråket, og som de andre minoritetsspråkene er tegnspråk likeverdig med norsk språklig og kulturelt uttrykk.²⁴

I dette prosjektet har vi hentet inn kunnskap om status for talegjenkjenning for norsk, samisk og tegnspråk.

Språk har, som vi har vært inne på, som regel både et muntlig og et skriftlig uttrykk.

I Norge har vi tradisjon for at alle snakker sin egen dialekt. NRK har riktignok som krav at det i nyhetsendinger og programinformasjon skal det som hovedregel snakkes normert bokmål eller nynorsk, men ellers er ikke normert talespråk utbredt i Norge. Tvert imot er det å holde på dialekten sin når man flytter til et nytt sted, i alle fall de siste 30-40 årene, blitt en hedersbetegnelse, selv om man i tidligere tider gjerne la vekk «landsmålet» når man flyttet til byen.

Innenfor skriftlig bokmål og nynorsk finnes det også stor variasjon i hvordan man kan uttrykke seg. Det varierer med grader av svært konservativt bokmål (riksmål) og konservativt nynorsk, opp mot tilnærmelesvis samnorsk i midten av skalaen. Dette er en utfordring for tale-til-tekst-teknologi da variasjon i både inndata og utdata som er korrekte, er store og vanskelige å håndtere for språkteknologiske systemer. Det danske selskapet Dictus²⁵, som leverer talegjenkjenningen til Stortinget, transkriberer både nynorsk og bokmål, har løst dette ved å kode hvordan hver enkelt representant ønsker å bli gjengitt, for eksempel om de vil ha a eller en-endeelse i bestemt form substantiv (boka vs. boka). I debatter presenteres taleren alltid med navn slik at det kan knyttes til den representantens transkriberingsønsker.

11,6 prosent av elevene i grunnskolen har nynorsk som hovedmål. Bokmål utgjør 87,3 prosent. Den resterende prosenten har et annet språk som hovedmål.²⁶ Nå er nynorsk og

²³ Regjeringen.no (29.04.2022). Fakta om samiske språk, tilgjengelig på <https://www.regjeringen.no/no/tema/urfolk-og-minoriteter/samepolitikk/samiske-sprak/fakta-om-samiske-sprak/id633131/>

²⁴ Jf. Språklova §4-7, tilgjengelig på <https://lovdata.no/dokument/NL/lov/2021-05-21-42>

²⁵ <https://dictus.dk/>

²⁶ Foss, Emilie Stolpe (13.2.2022) 1 av 9 har nynorsk som hovudmål i skolen, ssb.no, tilgjengelig på <https://www.ssb.no/utdanning/grunnskoler/statistikk/elev-ar-i-grunnskolen/artikler/1-av-10-har-nynorsk-som-hovudmal-i-skolen>

bokmål skriftspråk og ikke talte språk, og dialekter kan ligge nærmere den ene målformen enn den andre.

Det finnes ikke oversikt over hvor mange som har samisk som morsmål i Norge, men antall elever som får undervisning i samisk gir en indikasjon. I skoleåret 2016/17 var det til sammen 814 elever som mottok opplæring i et av de samiske språkene av et samlet elevtall på 629 275. For skoleåret 2015/16 var det 203 elever registrert med standpunkt- eller eksamenskarakter i et av de samiske språkene, ved avgang i 10. klasse, mens det for skoleåret 2016/17 var 219 elever som avla nasjonal prøve på samisk.²⁷

Ifølge oppslag på Språkrådets nettsider er det registrert mellom 20 000 og 30 000 samer som snakker de ulike samiske språkene i Norge, Sverige, Russland og Finland. Man regner med at det er om lag 100 000 samer, og av disse kan rundt 30 % snakke et samisk språk, og om lag 15 % bruker et samisk skriftspråk.²⁸

Nordsamisk, sørsamisk og lulesamisk har vern etter Europarådets minoritetsspråkpakt. Ifølge UNESCO er nordsamisk regnet som et truet språk, mens lulesamisk og sørsamisk er regnet som alvorlig truede språk.²⁹

Verken norsk eller samisk kommer inn blant de 100 mest talte språkene i verden.³⁰

Når det gjelder tegnspråk finnes det heller ikke nøyaktige tall. Ifølge et anslag fra Norges Døveforbund finnes det ca. 16 500 tegnspråkbrukere i Norge, mens Døvekirken oppgir 25 000. Blant disse finner vi både døve, hørselshemmede og hørende.³¹

Digitalisering av norske språk

I 2022 fikk Norge ny språklav. Loven har som formål å styrke norsk språk og fremme likestilling mellom bokmål og nynorsk. Loven skal også sikre vern og status for de andre språkene som staten har ansvar for: kvensk, romani, romanes, samiske språk og norsk tegnspråk. Sentrale statsorgan skal over tid bruke minst 25 prosent av både bokmål og nynorsk i allment tilgjengelige dokument, noe som ofte betyr 25 % nynorsk. I § 14 går det frem at «Statsorgan skal gjere skjema og andre sjølvbeteningsstenester tilgjengelege på bokmål og nynorsk samtidig.» Skjema på både nynorsk og bokmål er vi vant til, men selvbetjeningsstenester er ikke lenger bare digitale skjema, men også tekstbaserte

²⁷ Holth, Bjørn Are og Magnar Lillegård (2017) Statistikk over samiske språkbrukere i Norge En kartlegging av eksisterende datakilder og vurdering av fremgangsmåter for statistikk. SSB Notater 2017/34. Tilgjengelig på <https://www.ssb.no/kultur-og-fritid/artikler-og-publikasjoner/attachment/327486?ts=15f6cbddc98>

²⁸ Norsk språkråd (2016) Samisk. Språkradet.no, tilgjengelig på <https://www.sprakradet.no/Spraka-vare/Spraka-i-Norden/Samisk/>

²⁹ Kommunal- og distriktsdepartementet (2022) Fakta om samiske språk. Tilgjengelig på <https://www.regjeringen.no/no/tema/urfolk-og-minoriteter/samepolitikk/samiske-sprak/fakta-om-samiske-sprak/id633131/>

³⁰ Ghosh, Iman (2020) MISCRanked: The 100 Most Spoken Languages Around the World, visualcapitalist.com, tilgjengelig på <https://www.visualcapitalist.com/100-most-spoken-languages/>

³¹ Språkrådet (udat.) Hvor mange snakker norsk tegnspråk? Språkradet.no, tilgjengelig på <https://www.sprakradet.no/Spraka-vare/Tegnsprakteiknsprak/Ofte-stilte-spor-smal-om-tegnsprak/hvor-mange-snakker-norsk-tegnsprak/>

snakkeroboter. Kommune-Kari er en tekstbasert snakkerobot som er i bruk i over 100 kommuner.³² Roboten svarer foreløpig kun på bokmål. Med stadig flere stemmebaserte tjenester vil også taleteknologi sannsynligvis gjøre sitt inntog i slike svartjenester og med de utfordringene det medfører med hensyn til likestilling av målformer og språk.

«Opplæringsloven krever at alle læremidler i skolen, også de digitale, skal foreligge på bokmål og nynorsk samtidig, og til samme pris. Dette kravet blir ikke alltid oppfylt, og digitaliseringen forsterker problemet», skriver Språkrådets direktør Åse Wetås i en kronikk i Utdanningsnytt i forbindelse med arbeidet med ny digitaliseringsstrategi for grunnopplæringen.³³

Hun skriver videre at «Nynorskversjonene av digitale læremiddel er ofte ufullstendige, og de kommer senere enn bokmålsversjonene. Landssamanslutninga av nynorsk kommunar (LNK) og Nynorsksenteret har nylig gjennomført en undersøkelse der 110 skoler med nynorsk som opplæringspråk har vurdert digitale læremidler. Undersøkelsen viser at færre enn 25 prosent av skolene vurderer kvaliteten på de nynorske digitale læremidlene som jevn god med kvaliteten på tilsvarende læremidler på bokmål.» Wetås konkluderer med at som følge av digitalisering vil nynorskelever få et dårligere tilbud. Når det gjelder samisk og tegnspråk står det enda dårligere til.

2022-2032 er UNESCOs Internasjonale tiår for urfolksspråk som UNESCO leder sammen med urfolk og stater. Målet med språktiåret er å sette søkelys på språksituasjonen til urfolk, samt mobilisere myndigheter og andre til å bevare, vitalisere og fremme urfolksspråk. Sametingene i Norge, Sverige og Finland har valgt å følge opp språktiåret sammen, og har levert en felles plenumssak for alle sametingene: Mijá tekno - Sametingenes felles flaggskipsprosjekt i forbindelse med tiåret for urfolksspråk.³⁴ Målet med flaggskipsprosjektet er å styrke samiske språk i den digitale verdenen. I saksfremlegget heter det blant annet at det er et stort behov for å styrke samisk språkteknologi, slik at det lages flere samiske språkteknologiske verktøy som gjør det lettere å bruke samiske språk i den digitale verdenen. For å utvikle samisk språkteknologi trengs det store mengder data i et korpus. For samiske språk er mangelen på data stor og mye større enn for eksempel for nynorsk.

«For alle samiske språk er mangelen på tekst akutt, og det er nødvendig å samle inn alle mulige tekster, slik at det er mulig å få til et bra korpusmateriale. Korpusmateriale er digitale tekstsamlinger som fungerer som grunnlagsmateriale når man skal lage språkteknologiske verktøy, og som er også viktig i samisk språkforskning. Hvis det skal være mulig å lage samiske språkteknologiske verktøy, som fungerer og som er til nytte for

³² Prokom (udat.) Kommune-Kari, tilgjengelig på <https://prokom.no/kari/>

³³ Wetås, Åse (2022) Digitaliseringen forsterker den språklige forskjellsbehandlingen, utdanningsnytt.no, tilgjengelig på: <https://www.utdanningsnytt.no/digital-opplaering-digitalisering-nynorsk/digitaliseringen-forsterker-den-spraklige-forskjellsbehandlingen/335032>

³⁴ Sametinget (5.9.2022) Saksfremlegg Mijá tekno - Sametingenes felles flaggskipprosjekt i forbindelse med IDIL 2022-2032 Tilgjengelig på https://innsyn.onacos.no/sametinget/mote/norsk/wfinnsyn.ashx?response=journalpost_detaljer&journalpostid=2021068184&

samiske språk, så trengs det store mengder med tekstsamlinger i korpus.»
heter det i saksframlegget.

Det pekes på at de store teknologiselskapene som Amazon, Google og Microsoft leverer språkteknologi hovedsakelig for de store språkene, mens små urfolksspråk blir utelatt. Et annet problem er at disse teknologiselskapene har lukket kildekode, noe som gjør at andre språk ikke får tilgang til deres verktøy eller plattform.

Norsk språkteknologi (nynorsk og bokmål)

Lenge var mangel på språkdata et problem i Norge slik at vi ikke hadde datagrunnlag nok for å lage talegjenkjenning. Det fantes ikke nok materiale å trene systemene på til at de kunne utvikle teknologien.

I 1999 ble Nordisk Språkteknologi stiftet og lagt til Voss. Selskapet skulle utvikle norsk, svensk og dansk taleteknologi og sørge for kanskje tusen arbeidsplasser i vestlandskommunen. Selskapet hadde etter oppstart 65 årsverk, men måtte redusere til 45 allerede i 2001.³⁵ Året etter var de nede i 30. Fjerde året var selskapet konkurs. Men i årene selskapet var aktivt, samlet de inn en rekke språkprøver fra hele landet. Dette ble overtatt av Universitetene i Bergen, Oslo og Trondheim sammen med IBM og Språkrådet som kjøpte opp serverne til Nordisk Språkteknologi etter konkursen. Dette førte til at språkdataene ble tilgjengelige for forskning og utvikling.³⁶

I 2010 ble Språkbanken etablert ved Nasjonalbiblioteket. Språkbankens oppgaver er å skaffe til veie grunnlagsressurser for utvikling av språkteknologiske tjenester for kommersiell utvikling og akademisk forskning. Alt materiale Språkbanken utvikler, er fritt og gratis tilgjengelig.³⁷ Alle data fra arbeidet til Nordisk Språkteknologi er overført til Språkbanken.

I 2019 ble Norge det første landet i Norden som fikk tale-til-tekst-teknologi da Microsoft gjennom språkdata fra Språkbanken, utviklet taletranskribering til bokmål i Word og Outlook.^{38, 39}

Talegjenkjenning på norsk var et tiltak i regjeringens handlingsplan for universell utforming 2015-2019 (IKT9).⁴⁰ Selskapet Max Manus AS (nå fusjonert med selskapet Conscriptor til

³⁵ Nordisk Språkteknologi (samlside), NRK Hordaland, tilgjengelig på <https://www.nrk.no/emne/nordisk-sprakteknologi-hordaland-1.189301>

³⁶ Røsjø, Bjarne (2009). Språkteknologien gjør fremskritt igjen. Forskning.no, tilgjengelig på <https://forskning.no/barn-og-ungdom-informasjonteknologi-internett/sprakteknologien-gjor-fremskritt-igjen/928518>

³⁷ <https://www.nb.no/sprakbanken/>

³⁸ Olsen, Stein Jarle (2019). Microsoft lanserer tale-til-tekst på norsk. Tek.no, tilgjengelig på <https://www.tek.no/nyheter/nyhet/i/xPyjnQ/microsoft-lanserer-tale-til-tekst-paa-norsk>

³⁹ Talegjenkjenning. (u.å.) *Wikipedia*, tilgjengelig på <https://no.wikipedia.org/wiki/Talegjenkjenning>, lest 6.10.2022

⁴⁰ Barne-, likestillings- og inkluderingsdepartementet (2016) Regjeringens handlingsplan for universell utforming 2015-2019, tilgjengelig på <https://www.regjeringen.no/no/dokumentarkiv/regjeringen-solberg/andre-dokumenter/bld/2016/regjeringens-handlingsplan-for-universell-utforming-2/id2473299/>

selskapet Omilon⁴¹) fikk midler over Arbeids- og sosialdepartementets budsjett i 2015 og 2016 for å levere en løsning for talegjenkjenningsteknologi på norsk. Løsningen forelå i juni 2017⁴², men hovedsakelig via godkjent søknad til NAV for personer med varige funksjonsnedsettelse. TUVAs løsning, som er ikke lenger tilgjengelig.⁴³

Rent teknologisk skjedde det en stor utvikling innen språkteknologien på 2000-tallet. Fram til da ble det innen språkteknologien benyttet regelbaserte metoder ved at man forsøkte å gi en uttømmende definisjon av betydningen og bruken av hvert ord. Dette er en arbeidskrevende metode, men for små språk som for eksempel samisk, kan dette være eneste løsning. Gjennom et semantisk nett av ord og begreper, koder man det enkelte ordet. For eksempel gjennom å kode ordet «barn» som del av begrepet «menneske», som del av begrepet «vesen» osv. og hvordan ordet barn kunne opptre i setninger. Utover 2000-tallet skjedde det en enorm utvikling i datakraft. Tilgang til data og stor datakraft endret språkteknologien og man startet å ta i bruk maskinlæring. Dette førte til en rask utvikling.

I dag finnes taletranskribering til bokmål i en rekke produkter. Man kan aktivere tale-til-tekst i de fleste Microsoftprodukter, Google, iOS- og Android-telefoner. Det finnes en rekke apper som både kan transkribere det som sies, oversette til et annet språk og få det lest opp med syntetisk tale. Nettsøk kan gjennomføres med stemmen i Google og man kan skrive tekstmeldinger på telefonen, diktere lange dokumenter i Word eller få Teams-møtene automatisk teksten i sanntid.

Når det gjelder nynorsk er ikke situasjonen tilsvarende. Her har vi kun funnet ett kommersielt tilgjengelig produkt som leverer diktering på nynorsk, nemlig Dragon Professionally Speaking Anywhere. Dictus som leverer talegjenkjenning til Stortinget både på nynorsk og bokmål, har ikke kommersielt tilgjengelige produkter på noen av målformene.

Samisk språkteknologi

I 2001 ble Giellatekno, Senter for samisk språkteknologi ved Universitetet i Tromsø⁴⁴, startet som et prosjekt for samisk grammatisk analyse. I 2005 startet Sametinget opp Divvun⁴⁵, et prosjekt og en arbeidsgruppe for å lage samiske retteprogram. Giellatekno og Divvun har siden den gang jobbet sammen om kildekode for språkteknologi for de samiske språkene. Divvun ble en del av Universitetet i Tromsø i 2011. Til nå er det blitt utviklet språkressurser for tekst-til-tale for nordsamisk, maskinoversettelse fra nordsamisk til bokmål, og oversettelse mellom samiske språk.⁴⁶

Språkteknologimiljøet ved Universitetet i Tromsø (UiT) har i samarbeid med Sametinget samlet inn et samisk tekstkorpus. Verktøyene som er utviklet ved Sametinget og UiT, er

⁴¹ Omilon (19.1.2021) Max Manus og Conscriptor blir Omilon, *Pressemelding*, tilgjengelig på <https://omilon.com/no/article/vi-blir-omilon>,

⁴² Talegjenkjenning. *Wikipedia*, tilgjengelig på, <https://no.wikipedia.org/wiki/Talegjenkjenning> (lest 6.10.2022)

⁴³ NAV (udat.) Tuva - talegjenkjenning på norsk, tilgjengelig via <https://www.hjelpemiddeldatabasen.no/r11x.asp?linkinfo=52935>, lest 6.10.2022

⁴⁴ <https://giellatekno.uit.no/>

⁴⁵ <https://divvun.no/>

⁴⁶ Ressurser for nordsamisk tilgjengelig på: <https://giellatekno.uit.no/cgi/index.sme.nob.html>

grammatikkbaserte og altså ikke bygget på maskinlæringsbaserte metoder. Det er en av hovedgrunnene til at det er mulig å lage verktøyene, selv om det finnes få eller ingen annoterte språkdata tilgjengelig.⁴⁷ Å produsere nok tekst som kan brukes av KI er langt mer arbeidskrevende.

Samisk talegjenkjenning er i dag kun på eksperimentstadiet. En utfordring er at samisk ikke er definert datateknisk. «Konsekvensane av at eit språk ikkje er definert, er fleire. Brukarane vil berre sjå ein språkkode i staden for det riktige namnet på språket. Det gjer at dei ikkje kjenner att språket sitt, og verktøya blir såleis utilgjengelege. På Windows blir det uråd å nytta korrekturprogram, og det blir umogleg å nytta operativsystemtenester for språket, sjølv om ein installerer dei komponentane som trengst.»⁴⁸

Det er utviklet syntetisk tale for nordsamisk (tekst-til-tale). Det er i bruk i noen sammenhenger som på nettsidene til Troms og Finnmark fylkeskommune og flere kommuner innenfor det samiske forvaltningsområdet.⁴⁹

I dag finnes det ikke noen form for taletranskripsjon eller stemmestyring for noen av de samiske språkene.

Tegn-til-tale og tale-til-tegn-teknologi

Prosessering av tegnspråk enten som tegn-til-tale eller tale-til-tegn er et stort og komplekst område. Akkurat som med talte språk finnes det nasjonale tegnspråk og lokale varianter i form av tegnspråksdialekter. Det finnes også et internasjonalt tegnspråk⁵⁰, som verdens døveforbund ønsker skal omtales som «internasjonale tegn».

På europeisk nivå ble det i 2021 opprettet to store konsortier som arbeider med et utvalg av europeiske språk når det gjelder å digitalt oversette fra tegnspråk til tekst og omvendt.^{51,52}

Appen «Hand Talk Translator» oversetter automatisk tekst og tale til amerikansk og brasiliansk tegnspråk. Google m.fl. har utviklet tegn-til-tekst-konvertering for engelsk og kinesisk.⁵³ Det finnes også en rekke andre selskaper som arbeider med tegnspråksprosessering internasjonalt. I 2018 ble det rapportert om et nederlandsk firma som hadde utviklet en app kalt GnoSys basert på kunstig intelligens som skulle kunne

⁴⁷ Nørstebø Moshagen, S. (2022). Samisk språkteknologi i 2021. *Sprog I Norden*, 52(1), 93–102. Hentet fra <https://tidsskrift.dk/sin/article/view/132416>

⁴⁸ Nørstebø Moshagen, S. (2022). Samisk språkteknologi i 2021. *Sprog I Norden*, 52(1), 93–102. Hentet fra <https://tidsskrift.dk/sin/article/view/132416>. s. 96

⁴⁹ Antonsen, Lene og Trond Trosterud (2020) Med et tastetrykk. Bruk av digitale ressurser for samiske språk. Samiske tall forteller. Tilgjengelig på <https://samilogutmuitalit.no/nb/2020/med-et-tastetrykk-bruk-av-digitale-ressurser-samiske-sprak-2020>

⁵⁰ World Federation of the Deaf (2019) FAQ on International Sign. Tilgjengelig på <http://wfdeaf.org/news/resources/faq-international-sign/>

⁵¹ <https://signon-project.eu/>

⁵² <https://www.project-easier.eu/>

⁵³ <https://omnibridge.ai/>

transkribere tegnspråk til tekst.⁵⁴ Vi har ikke funnet noe om denne appen i senere kilder, eller App Store. Dette skyldes etter all sannsynlighet at det var et lanseringsoppslag og ikke er blitt videreført.

Når det gjelder norsk tegnspråk fikk et studentinitiativ midler fra Norges forskningsråd for å utvikle norsk tegnspråkteknologi i 2021-2022.⁵⁵ DeepSigns ambisjon var å bygge bro over kommunikasjonsgapet mellom brukere av tegnspråk og det hørende samfunnet. Formålet var å bygge teknologi som kunne bistå til å utvide repertoaret av tegnspråklige individers muligheter innen utdanning og sosial kontakt med ikke-tegnspråklige mennesker og omvendt. Konkret ønsket DeepSign å utvikle en programpakke som skulle oversette verbalt språk fra enhver form for lyd-, tekst- eller bildefil til tegnspråk. Teknologien skulle baseres på fremskritt innen kunstig intelligens, maskinlæring, språkprosessering og animasjonsteknologi til å gjennomføre oversettelser i sanntid. Prosjektet ble avsluttet tidlig og selskapet ble lagt ned uten å ha ferdigstilt sin teknologi eller lansert et produkt på det norske markedet.

Prosjektets ambisjonsnivå oversteg betydelig hva som var mulig å frembringe av realistiske resultater innenfor rammene av bevilgningen og kompetansen innad i kjernegruppen. Daglig leder for selskapet flyttet innovasjonen og selskapet til Nederland, da det her allerede finnes en etablert forskningsgruppe for tegnspråkteknologi ved Universitetet i Amsterdam. Det nederlandske døvemiljøet er godt integrert i utviklingen.⁵⁶

En av utfordringer de møtte var tilgang til transkribert tegnspråk. En kilde til dette er NRKs programmer med tegnspråk. DeepSign fikk benytte dette materialet til forskning og utviklingsformål, men ikke til kommersialisering.

Generelt for utvikling av språkteknologi i Norge, enten det er tegnspråk, norsk eller samisk, er at NRKs tekstete og tegnspråktolkete program kunne være en god kilde til å lage datasett som kunne vært brukt for å trene et system. Dessverre er det meste belagt med klausuler og kan derfor ikke brukes.

Det finnes i dag ingen tegn-til-tale-teknologi for norsk tegnspråk.

Språkdød

Språk utvikler seg over tid både på grunn av politiske forhold, folkevandring og generell globalisering. Alt dette påvirker språket man snakker i et geografisk område. I norsk historie kan vi for eksempel peke på fornorskingen av samene som førte til undertrykkelse av samisk språk og kultur og påvirkning av dansk gjennom unionstiden. Kolonialisering har vært kanskje den viktigste faktoren for endrete språk i store deler av verden. På det amerikanske kontinentet snakkes nå hovedsakelig spansk, engelsk og portugisisk, mens de lokale

⁵⁴ Financial Express (2018) New AI-powered app translates sign language into speech in real time: How it works. Tilgjengelig på: <https://www.financialexpress.com/life/technology-new-ai-powered-app-translates-sign-language-into-speech-in-real-time-how-it-works-1360933/>

⁵⁵ Norges forskningsråd – prosjektbanken (udat) STUD: DeepSign AS. Tilgjengelig på <https://prosjektbanken.forskningsradet.no/project/FORISS/327456?Kilde=FORISS&distribution=Ar&chart=bar&calcType=funding&Sprak=no&sortBy=date&sortOrder=desc&resultCount=30&offset=0&Prosjektleder=Jonas+Gahr+Sturtzel+Lunde>

⁵⁶ DeepSign AS (2022) - Avsluttende rapport til Norges forskningsråd (upublisert)

språkene enten har forsvunnet eller er i ferd med å forsvinne. På verdensbasis regner man med at totalt 571 språk er utdødd, det vil si at det ikke er i bruk, finnes noen som snakker det eller skriftlige kilder.⁵⁷ Innen utgangen av dette århundret regner man med at 1500 av språkene det snakkes i verden kan være utdødd.⁵⁸ Globalisering fører også til at mange behersker engelsk som ofte blir et felles *lingua franca* for store deler av den vestlige verden. Dette fører også til press på små språk, spesielt hvis de ligger tett opptil et svært utbredt språk. Dette fører sannsynligvis også til at små språk forsvinner.

Norge har hatt stor variasjon i det muntlige språket, det vil si et stort mangfold av dialekter. På samme måte som globaliseringen påvirker internasjonal språkbruk, påvirker regionalisering nasjonal språkbruk. På Østlandet snakker nå svært mange unge tilnærmet Oslo-dialekt, mens den eldre garde har holdt på stedets opprinnelige dialekt.⁵⁹

Mange minoritetsspråk står i fare for å dø ut i den digitale verden. Utvikling av taleteknologi gjøres først der hvor gevinsten økonomisk sett og tilgangen å datamateriale størst. De utrydningstruede språkene står i fare for å dø digitalt da det ikke finnes ressurser verken økonomisk eller nok data for å trene språkmodeller til å utvikle god talegjenkjenningsteknologi eller annen form for språkteknologi.

Samisk språkteknologi vanskeliggjøres ytterligere ved at det brukes tre ulike samiske språk. De samiske språkene står overfor de samme utfordringene nynorsk – for lite treningsdata, men samisk i langt større grad. Som følge av fornorskingspolitikken er det i dag neppe noen med samisk bakgrunn som ikke snakker norsk og svært mange behersker også engelsk som følge av undervisning i skolen. Dette betyr at mange med samisk bakgrunn behersker andre mer utbredte språk.

Digital språkdød er et fenomen beskrevet av András Kornai⁶⁰ hvor han peker på faktorer som påvirker at språk ikke greier å komme over det digitale skillet, og mener at 95 % av verdens språk ikke vil greie overgangen til den digitale verden. Han spår at nynorsk ikke vil overleve bokmål digitalt sett da hans undersøkelser viser at til nynorskbrukere bruker bokmålswikipedia og nettopp dette fenomenet påvirker at nynorsk ikke blir benyttet digitalt på samme måte som et større språk med mer prestisje. Dette kan bli tilsvarende for mange små språk ikke bare når det gjelder taleteknologi, men generelt. Digital språkdød oppstår ikke bare på grunn av manglede teknologi, men på grunn av tap av digitale brukere, tap av prestisje, tap av språket som identitet, mente Kornai.

Teknologirådet peker på farene ved manglede språkteknologi i Norge i sin rådgiving om teknologi til Stortinget i desember 2022. «Fortsetter tjenester å fungere dårlig på norsk kan

⁵⁷ International Language Services (2019) Understanding Extinct Languages: When and Why They Die Off. Tilgjengelig på <https://www.ilstranslations.com/blog/understanding-extinct-languages-when-and-why-they-die-off/>

⁵⁸ World Economic Forum (2022) 1,500 endangered languages could disappear by the end of the century. Tilgjengelig på: <https://www.weforum.org/agenda/2022/01/languages-endangered-diversity-loss-spoken/>

⁵⁹ Abusland, Å., Engan, O., Tørdal, R. M., Federl, M. (2020, 8. juni). Norsk dialektmangfold i fare?. NDLA. Tilgjengelig på <https://ndla.no/article/24524>

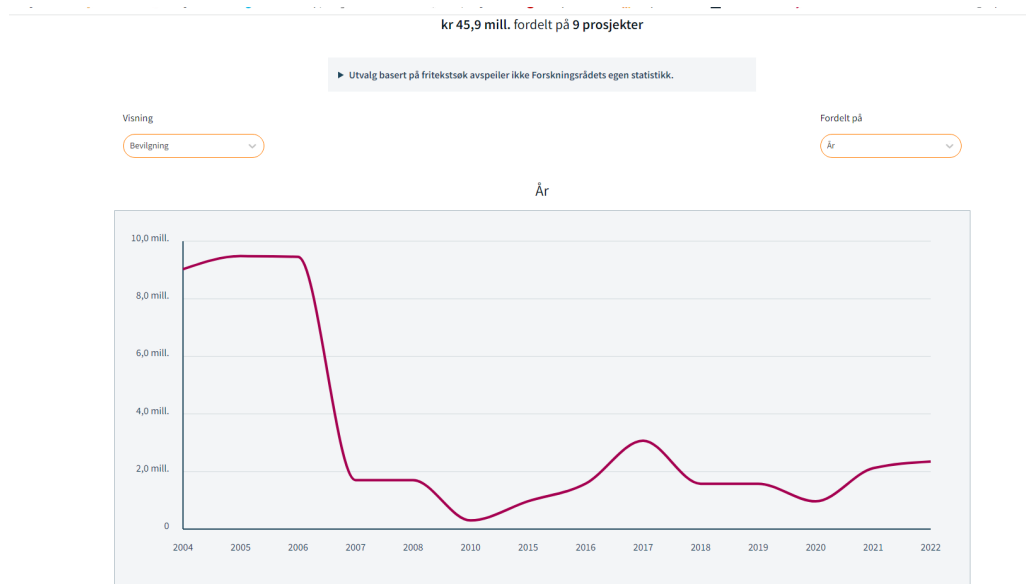
⁶⁰ Kornai A (2013) Digital Language Death. PLoS ONE 8(10): e77056. <https://doi.org/10.1371/journal.pone.0077056>

norsktalende bli ekskludert eller bytte til engelsk. Trenden kan gjøre det krevende for små språksamfunn å beholde språkbrukere, og det kan også true talemangfoldet dersom teknologien kun lærer seg bokmål.»⁶¹

⁶¹ Teknologirådet (2022) Taleteknologi med kunstig intelligens. Tilgjengelig på <https://teknologiradet.no/publication/taleteknologi-og-kunstig-intelligens/>

5. Noen prosjekter innen talegjenkjenningsteknologi

Et søk i Norges forskningsråds prosjektbank på termen «taleteknologi» viser at fra 2004 til 2022 ble det bevilget 45,9 millioner fordelt på ni prosjekter og at det særlig ble satset 2004-2007. Dette er kun prosjekter med det norske ordet «taleteknologi» i tittelen.



Figur 1 Utvikling av bevilgninger til «taleteknologi» 2004-2022. Skjerm bilde fra nettsiden

De ni prosjektene er ⁶²:

1. FONEMA - Metodeutvikling for naturtro norsk talesyntese. Tildelt: kr 8,0 mill. 2003 – 2007
2. Et norsk talespråkkorpus (NoTa) Tildelt: kr 1,9 mill. 2004-2006
3. Digitale utfordringer: deltakelse og funksjonshemming i informasjonssamfunnet. Tildelt: kr 5,0 mill. 2006 – 2009
4. BRAGE - Brukergrensesnitt med naturlig tale. Tildelt: kr 25,7 mill. 2002 – 2007. Prosjektet skulle blant annet forbedre systemenes evne til å forstå spontan tale.
5. Stemmestyrte Interaksjon. Tildelt: kr 0,25 mill. Prosjektperiode: 2006 – 2006
6. Lydstyring. Tildelt: kr 0,30 mill. 2010
7. Atomiske enheter for universell representasjon av tale. Tildelt: kr 8,4 mill. 201-20220
8. En pilotstudie for å vurdere bruk av taleteknologi til å forbedre skriveferdigheter blant ungdommer med lav lese- og skriveprestasjon. 2019 – 2022
9. SCRIBE: Maskintranskripsjon av norsk samtale. Tildelt: kr 12,0 mill. 2021-2025

SCRIBES målsetting er å utvikle et norsk tale-til-tekst-transkripsjonssystem for flerpartssamtaler under realistiske opptaksforhold⁶³. Prosjektet skal utvikle modeller som håndterer spontan samtale, som kan håndtere turtaking og utnytte konteksten i dialogen.

⁶² prosjektbanken.forskningsradet.no

⁶³ SCRIBES nettsider - <https://scribe-project.github.io/>

Modellene skal også støtte bruk av ulike dialekter og ulike ortografier (bokmål, nynorsk eller dialektspesifikke). De arbeider også med å forbedre målemetodene for kvaliteten på transkripsjonen og innholdet, ikke bare opptelling av feil ord. Som for eksempel hvis en person sier «ikkje» i en setning, og det blir transkribert som «ikke» blir det tradisjonelt målt som feil transkribering, men meningsinnholdet er det samme. Derfor er dette en annen type feil om det hadde blitt transkribert til «i kjelen». Prosjektet skal derfor søke å utvikle en annen type evalueringsteknologi enn *word error rate*. SCRIBE skal også bidra til teoretisk og metodisk utvikling av maskinlæring med sparsomme data. Deltakere i SCRIBE-prosjektet er Telenor, NTNU, NRK, OpenAI og Språkbanken ved Nasjonalbiblioteket.

På tampen av året bevilget regjeringen 3 millioner ekstra til Divvun for utvikling av samisk språkteknologi i forbindelse med tiåret for urfolksspråk.⁶⁴

NorwAI (Norwegian Research Center for AI Innovation) er den største akademiske satsingen på KI-innovasjon i Norge. Senteret ligger ved NTNU i Trondheim og koordinerer forsknings- og innovasjonsaktiviteter mellom tre universiteter, to forskningsinstitutter og 11 bedrifter. Senteret ble opprettet i 2020. Et av forskningsfeltene til senteret er språk og språkprosessering og har som målsetting å bygge en stor skandinavisk språkmodell.⁶⁵

European Language Equality (ELE)⁶⁶ er et stort konsortium bestående av 52 partnere som dekker alle europeiske land, samt forskning og industri. ELE-prosjektet arbeider for å oppnå full digital språklig likestilling i Europa innen 2030. Språkrådet deltar på vegne av Norge. ELE leverer rapporter om 32 ulike språk på ensartet vis slik at det skal være mulig å sammenligne utviklingen og status. I rapporten om norsk som kom i 2022, *ELE-Report 26: D1.26 Report on the Norwegian Language*, gjennomgås hele språkteknologifeltet, ikke bare tale-til-tekst.⁶⁷

Her slås det fast at siden 2012 (da den forrige rapporten ble skrevet) er det tatt i bruk nye eller forbedrede utgaver av maskinoversettelse, taleteknologi, snakkeroboter (chatbots) og digitale assistenter på norsk, og at utviklingen innenfor maskinlæring har bidratt til dette. Likevel er språkteknologi for norsk langt fra det nivået som de store europeiske språkene som engelsk, tysk, fransk og spansk ligger på. Norsk språkteknologi havner omtrent midt på treet når man sammenligner med de nevnte språkene og andre europeiske språk ifølge forfatterne bak rapporten. Intensjonen bak rapporten var å undersøke hva som skal til for å få norsk språkteknologi opp på samme nivå som de store europeiske språkene; hvilke grunnlagsressurser må det satses på, hvilke verktøy mangler eller må bli bedre og hva skal til for at norsk også i fremtiden skal være et samfunnsbærende språk. For skal et språk fungere som samfunnsbærende språk i det 21. århundre, må det også fungere digitalt.

⁶⁴ Kommunal- og distriktsdepartementet (28.10. 2022) Ekstra løyving til samisk språkteknologi *Pressemelding*. Tilgjengelig på <https://www.regjeringen.no/no/aktuelt/ekstra-loyving-til-samisk-sprakteknologi/id2942028/>

⁶⁵NorwAI (udat.) Language and Personalization, <https://www.ntnu.edu/norwai/lap>

⁶⁶ <https://european-language-equality.eu/>

⁶⁷ Eide, Kristine, Kåsen, Andre og Dale, Ingerid Løyning (2022) *Report on the Norwegian Language*. ELE rapport 1.26. Tilgjengelig på https://european-language-equality.eu/wp-content/uploads/2022/03/ELE_Deliverable_D1_26_Language_Report_Norwegian_.pdf

I rapporten pekes det på flere utfordringer for norsk taleteknologi. For det første har ikke Norge et pågående forskningsprogram for språkteknologi. Det finnes riktig nok en god del mindre prosjekter og internasjonalt samarbeid, men ikke et stort forskningsprogram. Forskningsrådet har et program for kunstig intelligens hvor man også kan søke støtte til språkteknologiske prosjekter.

Andre utfordringer det pekes på i rapporten er at selv om det finnes godt med innsamlete språkdata på bokmål, er utvalget lite på nynorsk. Talegjenkjenning fungerer generelt best på standard østlandsk og det er store utfordringer knyttet til dialekter både når det gjelder å gjenkjenne og transkribere. Det pekes også på at vi mangler språkdata fra en rekke fagfelt, noe som gjør det vanskelig å utvikle språkteknologiske produkter for domenet og at mye arbeid gjenstår med å utvikle nok grunnlagsressurser for gode språkteknologiske produkter på både bokmål og nynorsk. Videre trekkes det frem at utenlandske leverandører må bli mer oppmerksomme på likestillingen mellom bokmål og nynorsk, og det understrekes at det er viktig at det offentlige bruker sin innkjøpsmakt ved offentlige anskaffelser og peker på reglene i språkloven.

Norges forskningsråds bevilgninger til kunstig intelligens har økt svært mye de siste årene og det er grunn til å anta at en del språkteknologiske prosjekter vil ha støtte gjennom denne satsningen. I tillegg kommer bevilgninger direkte fra regjeringen til konkrete FoU-prosjekter eller forskningssenter. Dette har vi ikke klart å skaffe fullstendig oversikt over innenfor prosjektet.

6. utfordringer

Basert på skriftlige kilder vi har gjennomgått i prosjektet, samtaler, møter, e-poster og annen kommunikasjon med fagfolk innen feltet, egne tester og brukertester med hørselshemmede, mener vi å ha identifisert en rekke utfordringer når det gjelder utvikling av tale-til-tekst, listet i tilfeldig rekkefølge.

Manglende språkdata

Dette gjelder både nynorsk, bokmål, samisk og tegnspråk, men særlig samisk og tegnspråk. Kilder som kunne ha vært benyttet, som transkriberte radioprogram eller tekstete TV-sendinger, er ofte belagt med opphavsrett eller andre begrensninger, og kan ikke bli brukt til formålet. Det finnes mest tilgjengelig materiale for bokmål, ganske lite på nynorsk og så godt som ingenting på samisk eller tegnspråk.

Manglende ressurser

Dette henger sammen med at det mangler annoterte data. Å samle inn språkprøver og få dette transkribert er svært ressurskrevende. Det antas at å transkribere én time med lyd tar 4-5 timer.

Kommersielle hensyn

Mye av utviklingen innen språkteknologi er drevet av de store internasjonale teknologiselskapene som Nuance, Facebook, Google og Amazon. Dette betyr at innsatsen legges der hvor det kommersielle potensialet er størst. Små språk kommer langt ut i køen med mindre myndighetene bevilger midler slik at grunnlagsressurser kan lages og teknologien utvikles. Siden det allerede er utviklet god taletranskripsjon for bokmål, er incentivene for å utvikle parallelle nynorskversjoner små. Det samme gjelder samisk, da som regel både nynorskbrukere og samer forstår og kan bruke bokmål.

Spranget mellom naturlig, spontan tale og transkribert formalspråk

Når mennesker snakker sammen brukes språket på en helt annen måte enn ved skriftlig kommunikasjon, selv om det for eksempel i private tekstmeldinger eller oppdateringer i sosiale medier brukes et språk med mer muntlig preg og dialekt da noen opplever seg selv som fremmedgjorte i sin egen kommunikasjon ved å skrive korrekt bokmål eller nynorsk. Likevel er muntlig kommunikasjon veldig langt unna en skriftlig tekst som en bok eller avisartikkel. Fyllord, tonefall, ikke-ord (meningsbærende lyder), gjentakelser og ufullstendige setninger er en helt naturlig måte å kommunisere på mellom mennesker. Hvis dette blir transkribert fonetisk, det vil si skrevet ned slik lydene framkommer (lydskrift), er det veldig langt unna slik en bokside ser ut. Skal automatisk transkripsjon brukes for eksempel til transkribering av naturlig, spontan tale, trengs det datasystemer som klarer det som den menneskelige hjerne klarer, nemlig å sitte igjen med essensen av innholdet - selv med mange omveier i det muntlige språket - og transkribere det til en fornuftig tekst.

Bakgrunnsstøy

Mange løsninger er i dag avhengig av god kvalitet på inndata. Det vil si at de er avhengig av god og klar lyd for at transkripsjonen skal bli bra. Dette fordrer både gode mikrofoner og lite bakgrunnsstøy. Tester vi gjorde i forbindelse med å lage guide for bruk av tale-til-tekst-

produkter viste at kvaliteten på inndata var svært viktig. Bakgrunnsstøy eller at flere snakker samtidig, gjør det vanskelig å skille ut og transkribere. Mennesker kan sile ut stemmen til en person som står og snakker til en og ikke høre dette, hvis vedkommende ser på et TV-program hen er svært interessert i. Foreløpig er ikke taletranskriberingen som er tilgjengelig i dag, i stand til den samme formen for selektiv hørsel, men hvis dette kan koples til biometriske data gjennom stemmegjenkjenning kan det være en måte å løse utfordringen på. Dette åpner igjen spørsmål knyttet til persondata og personvern.

Domenespesifikke treningsdata

Mennesker som snakker sammen forstår hverandre ofte ut fra hva som er temaet for samtalen, felles bakgrunn, aktuelle hendelser osv. Dess lenger unna man er faglig, kulturelt og geografisk, dess vanskeligere er det å forstå en del av det som blir sagt. To meteorologer som snakker sammen om nyere forskning innen værsystemer vil snakke om været annerledes enn to tenåringsjenter i Bergen. Slik er det også med maskiner som skal tolke og transkribere. Hvis ordtilfanget som systemet gjetter på blir sagt, hentes fra hva som er de vanligste ordene totalt sett, kan det bli feil. Forkortelser, egennavn og faguttrykk vil være vanskelig å få korrekt. Det som har vist seg nyttig, er da å trene systemet på data fra det området (domenet) det skal brukes på. Dette har for eksempel det danske selskapet Dictus gjort, som leverer talegjenkjenningsteknologi til Stortinget. Ved å trene systemet på lyd og transkripsjon fra innlegg og debatter i stortingssalen, blir systemet lært opp til ord og vendinger som brukes i Stortingssalen, men som ikke nødvendigvis er så vanlige ellers. Omilon som leverer taletranskribering på norsk blant annet til helsesektoren, har lagt til ordlister beregnet for nettopp denne sektoren slik at systemet har rett ordforråd for oppgaven. Bruk av domenespesifikke språkdata og ordlister rettet mot oppgaven, gjør at transkribering blir mer korrekt. Dette har også en økonomisk side, da det må være noen som er villige til å finansiere slike domenespesifikke treningsdata, eller at de som leverer talegjenkjenningsteknologi antar at det finnes et kommersielt marked for det domenespesifikke produktet slik som Nuance/Omilon har gjort med sitt taletranskripsjonsprodukt rettet mot medisinsk personell.

Stemmegjenkjenning

Skal man transkribere flere som snakker samtidig, kan det være nødvendig å knytte det til stemmegjenkjenning slik at det fremgår hvem som sier hva. I digitale møter med automatisk transkripsjon som for eksempel Microsoft Teams ser man hvem som snakker ut fra hvilken konto som er pålogget møtet. Hvis flere personer snakker på samme mikrofon (konto), vil det ikke fremkomme hvem som snakker. Dette kan løses ved å benytte stemmegjenkjenning. Stemmegjenkjenning er en biometrisk prosess og er like unik som fingeravtrykk. Gjennom å kople stemmegjenkjenning til taletranskribering kan dette løses. Samtidig gir man fra seg unik, biometrisk informasjon. Det samme gjør man når man bruker for eksempel Googlesøk med stemmen og du er pålogget med konto. Dette utløser spørsmål knyttet både til personvern og eiendomsrett til biometriske data.

Banale avatarer og utgangspunkt i hørende sine behov

Tegnspråk er et annerledes språk enn talespråk, og tegnspråk har heller ikke en direkte ekvivalent i tekst. Selv om muntlig språk formidles mye også med tonefall, gester og mimikk, så er det siste enda mer framtrædende i tegnspråk. Når man skal utvikle en tale-til-tegn-generator så må det nødvendigvis være en digital menneskefigur som snakker, eller en såkalt avatar. Tilbakemeldinger fra døve har vært at disse avatarene har vært veldig tegneserieaktige og at man som døv da framstår som barnslig og banal i sitt språk. Døve opplever også at teknologien utvikles først og fremst med tanke på hørendes behov for å kommunisere og formidle til døve.

Personvern og opphavsrett

Talegjenkjenningsteknologi har utfordringer knyttet til opphavsrett når det gjelder treningsdata, for eksempel gjennom at tekstete TV-program som kunne vært brukt som treningsdata er belagt med opphavsrett og ikke kan brukes. En annen utfordring knyttet til opphavsrett er knyttet til persondata fra taleprøver. Selv om taledata kan være anonyme, er stemmen et biometrisk kjennetegn som kan samles inn uten at man vet om det. En undersøkelse fra 2022 viste at videokonferanseløsninger som Zoom, Slack, Microsoft Teams/Skype, Google Meet og Cisco Webex fortsatte å samle inn taledata selv om møtedeltakere hadde satt mikrofonen på «mute». Når man har skrudd av mikrofonen regner man ikke med at selskapet som leverer tjenesten fortsatt har tilgang til mikrofonen og lyden i rommet du befinner deg i. «Forskerne ønsket å finne ut hvorvidt rådataene kan brukes til å avdekke hvilke aktiviteter som foregikk i bakgrunnen mens mikrofonen var dempet. Dette klarte de, riktignok ved hjelp av en spesialutviklet maskinlæringsalgoritme, med en presisjon på 82 prosent. Slike analyser kan ifølge forskerne brukes til «fingerprinting»-formål – altså profilering av brukere basert på ulike typer innhenting av informasjon om brukerne.»⁶⁸ Det har også vært flere tilfeller i USA hvor data fra smarte høyttalere som man styrer med stemmen, er forsøkt brukt som vitnebevis i kriminalsaker som drap.⁶⁹

Bias i treningsdata

Skjevhet (bias) i det materialet systemet er trent på, vil gi dårligere resultat for det som systemet ikke er trent på. Undersøkelser fra USA viser at fargede mennesker har større problemer med å bli transkribert korrekt enn hvite.⁷⁰ Hvis all trening av et talegjenkjenningssystem er gjort på menn fra Bodø i alderen 50-60 år, vil systemet bli veldig godt egnet til å gjenkjenne og transkribere det som sies av 50-60 år gamle menn fra Bodø. Dette er et reelt problem i taleteknologi. Systemene er ofte trent på voksne menn, ikke

⁶⁸ Gjessing, Marianne, Knudsen E. (19.4.2022) Nytt funn: Samtale-apper kan fortsette å samle lyddata selv om mikrofonen er dempet. Digi.no. Tilgjengelig på <https://www.digi.no/artikler/nytt-funn-samtale-apper-kan-fortsette-a-samle-lyddata-selv-om-mikrofonen-er-dempet/518816>

⁶⁹ Saumya, Rishabh (18.9.2021) Alexa: a catalyst in the evidence law? The Crimal Law Blog. Tilgjengelig på <https://criminallawstudiesnluj.wordpress.com/2021/09/18/alexa-a-catalyst-in-the-evidence-law/>

⁷⁰ Lopez, Claudia Lloreda (5.7.2020) Speech Recognition Tech Is Yet Another Example of Bias. Scientific American online. Tilgjengelig på <https://www.scientificamerican.com/article/speech-recognition-tech-is-yet-another-example-of-bias/>

kvinner, barn eller gamle. Dermed blir det oftere feil når personer som systemet ikke er trent på skal snakke og transkriberes. Personer med stamming, dysartri (problemer med å uttale ordene som følge for eksempel av hjerneslag, CP eller demens) eller andre med uklar uttale vil ha det vanskelig med å bruke produkter basert på generell talegjenkjenningsteknologi. Dette gjelder også døve og hørselshemmede som også kan ha dårlig diksjon. Det gjøres arbeid på dette feltet, også i Norge, men dette er på forskningsstadiet og ikke knyttet til teknologien som er gratis på markedet i dag.

Dialekt

Den store variasjonen i hvordan vi snakker i Norge gjør det vanskeligere å få god talegjenkjenning da det finnes lite språkdata som kan brukes i trening av språkmodeller. Å samle nok data er både tidkrevende og dyrt. Da norsk talegjenkjenning allerede er etablert og fungerer relativt godt for diktering på bokmål, kan en konsekvens bli at folk bruker standardisert bokmål som talemål ved bruk av systemer eller tjenester som benytter talegjenkjenning. Dette er ikke med på å fremme språklig likestilling eller personlig identitet knyttet til dialekten man har.

Valgfrihet og likestilling i skriftspråk

Når det gjelder norsk skriftspråk så er det stor valgfrihet i form og stil. Det kan hete både «en jente – jenten» og «ei jente-jenta/jenten» på bokmål, mens det på nynorsk bare kan hete «ei jente-jenta». Ordet «fremtid» kan også hete «framtid» på bokmål og være både hankjønn og hunkjønn og bøyes slik: en framtid- framtiden, en fremtid- fremtiden, ei/en framtid - ei/en fremtid. På nynorsk kan det hete både «vi» og «me» og man kan skrive både «å gjere» og «å gjera». Begge deler er like korrekt. Når man skriver, vil man ofte bestemme seg for stilnivå, om man skal bruke kløyvd infinitiv eller ei, eller tre kjønn på bokmål og være konsekvent med det. Denne språklige valgfriheten kan gjøre det vanskelig å få til god transkribering med mindre man har systemer hvor man kan bestemme stilnivå. Det er også lite kjennskap til de to norske målformene i utlandet eller til den språklige likestillingen mellom nynorsk og bokmål.

7. Oppsummering

Kort oppsummert kan vi si at taletranskribering i form av diktering med standard østnorsk talemål til bokmål fungerer greit og er tilgjengelig i flere apper og programmer i 2022. Egennavn, tall, forkortelser og faguttrykk er likevel vanskelig å få korrekt.

Har man et system som er knyttet til et spesifikt fagområde og trent spesielt på området, fungerer bedre enn hvis det benyttes en generell språkmodell.

Transkribering av naturlig, spontan samtale fungerer dårlig.

Diktering til nynorsk finnes, men har lavere kvalitet enn bokmål og er kun tilgjengelig i ett kommersielt produkt. Nynorsk finnes verken i Apple-, Microsoft- eller Google-produkter eller annen hverdagslig teknologi.

Dagens tale-til-tekst-teknologi er ikke på et nivå hvor det kan erstatte skrivetolker for personer med nedsatt hørsel.

Det er ikke utviklet noen teknologi for transkribering av tegnspråk eller noen av de samiske språkene.

Det er stort behov for mer taledata med ulike type stemmer og dialekter. Dette gjelder både nynorsk og bokmål.

Det er en rekke utfordringer når det gjelder utvikling av samisk taletranskribering og tegn-til-tekst-teknologi knyttet til ressurser, og dagens taleteknologi ofte er drevet fram kommersielle interesser og der hvor mest data er tilgjengelig.

For samisk må regelbasert metoder fortsatt brukes da samisk tekstmateriale ikke produseres på samme måte «av seg selv» i stort tempo slik som for eksempel engelsk tale og teksting frembringes på internett og sosiale medier daglig.

For videre arbeid er det derfor viktig å få utviklet parallellkorpus. For tegnspråk betyr det videomateriale med tale, tegn og tekst.

Selv om nye språkmodeller kan trenes på mindre data eller lignede språk, er man likevel avhengig av politisk vilje og tildeling av midler for å utvikle språkteknologi for både bokmål, nynorsk, samisk og tegnspråk. Hvis ikke kan vi oppleve at man blir tvunget over til bokmål eller engelsk i møte med taleteknologi.

I regjeringens handlingsplan for universell utforming 2021-2025, pekes det på at «språkteknologi er en forutsetning for at det kan lages universelt utformede tjenester, blant annet velferdstjenester. Alle tjenester må kunne fungere på nynorsk, bokmål og samiske språk, og digitale systemer må forstå norske dialekter.»⁷¹

Språkrådet peker også på at språkteknologi er en viktig del av universell utforming av digitale løsninger og at behovet for språkteknologi vil øke i tiden som kommer både fordi vi

⁷¹ Kulturdepartementet (2021) Bærekraft og like muligheter – et universelt utformet Norge (2021–2025), tilgjengelig på <https://www.regjeringen.no/no/dokumenter/barekraft-og-like-muligheter-et-universelt-utformet-norge/id2867676/>

kommuniserer skriftlig i større grad enn før, og det byr på utfordringer for brukere med lese- og skrivevansker eller svekket syn. Stadig flere arrangementer som lyd- og videoopptak uten teksting, ofte i sanntid. Dette kan skape problemer for brukere med hørselsvansker.⁷²

For at dette skal bli en realitet, må taleteknologien utvikles og forbedres.

Mange potensielle brukere av tale-til-tekst-teknologi er ikke oppmerksomme på de mulighetene som finnes.

Dessverre er ikke den lett tilgjengelige teknologien så presis at hørselhemmede er gruppen som vil ha mest glede av den. Blinde, dyslektikere eller personer som har problemer med å bruke tastatur kan ha stor nytte av teknologien slik den er i dag, så lenge man ønsker å **diktere til bokmål**.

Mange uten funksjonsnedsettelse vil også kunne ha glede av teknologien da den frigjør fingre, og det går mye raskere å diktere enn å bruke tastatur.

Tale-til-tekst-teknologi er et steg i retning av et mer universelt utformet samfunn, men teknologien er i dag ikke god nok til at den fungerer for alle.

⁷² Språkrådet (2021) Språkteknologi og universell utforming, tilgjengelig via <https://www.sprakradet.no/Sprakarbeid/Sprak-IKT/sprakteknologi-og-universell-utforming/>