

SPRÅKENE VÅRE STÅR I FARE: EN UTFORDRING FOR SPRÅKTEKNOLOGIEN

Vi er vitner til en digital revolusjon som påvirker kommunikasjon og samfunnet dramatisk. Den seneste utviklingen i digital informasjons- og kommunikasjonsteknologi blir noen ganger sammenlignet med Gutenberg's oppfinnelse av trykkpressen. Hva kan denne analogien fortelle oss om fremtiden for den europeiske informasjonssamfunnet generelt og for språkernes stil-ling spesielt?

Vi opplever en digital revolusjon som kan sammenlignes med Gutenberg's oppfinnelse av trykkpressen.

I kjølvannet av Gutenberg's oppfinnelse skjedde flere store gjennombrudd i kommunikasjon og kunnskapsutveksling, som for eksempel Luthers oversettelse av Bibelen til eget morsmål. Siden Gutenberg's tid har man utviklet flere teknikker for bedre håndtering av språkbehandling og kunnskapsutveksling:

- standardisering av rettskriving og grammatikk for de vanligste språkene har gitt en hurtigere spredning av nye vitenskapelige og intellektuelle ideer;
- utviklingen av offisielle språk har gjort det lettere for innbyggerne å kommunisere innenfor visse (of- test politiske) grenser;
- undervisning og oversettelse mellom språk har bi- dratt til utveksling på tvers av språk;

- etablering av redaksjonelle og bibliografiske ret- ningslinjer har sikret kvaliteten og tilgjengeligheten av trykt materiale;
- etablering av ulike medier som aviser, radio, fjernsyn, bøker og andre medier har dekket en rekke kommu- nikasjonsbehov.

De siste tjue årene har informasjonsteknologi bidratt til å automatisere og forenkle mange av disse prosessene:

- publiserings- og tekstbehandlingsprogrammer har erstattet skrivemaskin og dokumentproduksjon;
- Microsoft PowerPoint har erstattet overheadtrans- parenter;
- e-post gjør det mulig å sende og motta dokumenter raskere enn med en faksmaskin;
- Skype tilbyr billige telefonsamtaler via Internett og legger til rette for videokonferanser;
- ulike formater for lagring av lyd og video gjør det en- kelt å utveksle multimedial-innhold;
- søkemotorer gjør det enkelt å søke i nettsider;
- nettbaserte tjenester som Google Translate produse- rer raske, omtrentlige oversettelser;
- sosiale medier som Facebook, Twitter og Google+ forenkler hurtig kommunikasjon, samarbeid og in- formasjonsdeling.

Selv om slike verktøy og programmer er nyttige, er de ennå ikke i stand til fullt ut å fylle rollen som en bære-

bjelke for innbyggerne i et flerspråklig europeisk samfunn, med fri flyt av informasjon og varer.

2.1 SPRÅKGRENSE HINDRER UTVIKLINGEN AV ET EUROPEISK INFORMASJONSSAMFUNN

Vi kan ikke forutsi nøyaktig hvordan fremtidens informasjonssamfunn vil se ut. Men det er svært sannsynlig at den viktigste revolusjonen i moderne kommunikasjonsteknologi vil ligge i nye måter å samle folk som snakker forskjellige språk. Dette legger press på den enkelte, som må lære nye språk, og på programutviklere, som må lage nye applikasjoner som kan sikre gjensidig forståelse og tilgang til felles kunnskap. I en økonomi og et informasjonssamfunn som blir stadig mer globalisert vil nye medier føre til enklere interaksjon på tvers av språk, språkbrukere og ulike typer innhold. Sosiale medier som Wikipedia, Facebook, Twitter, YouTube, og nylig Google+ har blitt stadig mer utbredt, men dette er bare toppen av isfjellet.

En stadig mer globalisert økonomi og informasjonssamfunn konfronterer oss med flere språk, ulike språkbrukere og ulike typer innhold.

Ifølge en fersk rapport fra Europakommisjonen kjøper 57% av Internettbrukerne i Europa varer og tjenester på språk som ikke er deres eget morsmål (engelsk er det vanligste fremmedspråket, fulgt av fransk, tysk og spansk). 55% av brukerne kan lese innhold på et fremmedspråk, mens bare 35% bruker et annet språk til å skrive e-post eller poste kommentarer på nettet [6]. For noen år siden var engelsk kanskje Internettets *lingua franca*, men nå er situasjonen dramatisk forandret. Mengden av nettbasert innhold på andre europeiske språk (samt asiatiske og språk fra Midtøsten) har eksplodert.

Dette digitale 'klasses skillet' mellom språkene har overraskende nok ikke fått mye offentlig oppmerksomhet, på tross av at det gjennomsyrrer hele samfunnet. Men det aktualiserer et viktig spørsmål: Hvilke europeiske språk vil overleve i et nettverksbasert informasjons- og kunnskapssamfunn, og hvilke er dømt til å forsvinne?

2.2 SPRÅKENE VÅRE STÅR I FARE

Mens trykkteknologien bidro til å øke informasjons-spredning i Europa, førte den også til språkdød. Regionale språk og minoritetsspråk ble sjelden trykt, slik at språk som kornisk og dalmatisk forble begrenset til muntlig overføring, noe som i sin tur begrenset bruksområdene. Vil Internett ha samme virkning på språkene våre?

Det språklige mangfoldet i Europa er en av de viktigste delene av vår kulturarv.

Europas omtrent 80 språk utgjør en av de viktigste delene av vår kulturarv og en sentral del av den europeiske samfunnsmodellen [7]. Mens språk som engelsk og spansk sannsynligvis vil overleve på det nye digitale markedet, risikerer mange europeiske språk å bli irrelevante i et nettverksbasert samfunn. Dette vil kunne svekke Europas posisjon på verdensbasis, og svekke målet om likeverdig deltakelse for alle europeiske borgere, uavhengig av språk. Ifølge en UNESCO-rapport om flerspråklig-het er språk et viktig middel for å nyte godt av grunnleggende rettigheter, som politisk ytringsfrihet, utdanning og samfunnsdeltakelse [8].

2.3 SPRÅKTEKNOLOGI KAN TILRETTELEGGE FOR SPRÅKBRUK

Tidligere ble det først og fremst investert i språkopplæring og oversettelse. Beregninger viser at det europeiske

markedet for oversettelse, tolkning, programvarelokalisering og nettstedsglobalisering utgjorde 8,4 milliarder euro i 2008, og dette tallet forventes å vokse med 10% årlig [9]. Men denne investeringen dekker bare en liten del av det nåværende og fremtidige behovet for kommunikasjon mellom språk. Et viktig tiltak for å sikre bredden og mangfoldet av språkbruk i morgendagens Europa er å bruke riktig teknologi, akkurat som vi bruker teknologi til å løse utfordringer innen transport, energi og universell utforming.

Digital språkteknologi (rettet mot alle former for skrevet tekst og muntlig tale) kan hjelpe mennesker til å samarbeide, drive handel, dele kunnskap og delta i sosiale og politiske debatter på tvers av språkbarrierer og datakunnskap. Språkteknologi er ofte innebygget i komplekse systemer som hjelper oss med å:

- finne informasjon med Internett-søkemotorer;
- sjekke staving og grammatikk i tekstbehandlingsprogram;
- vise produktanbefalingene i nettbutikker;
- høre taleinstruksjoner fra bilnavigasjonssystemer;
- oversette nettsider via nettbaserte tjenester.

Språkteknologi består av en rekke kjerneapplikasjoner som legger til rette for ulike prosesser innenfor et større applikasjonsrammeverk. Formålet med META-NETs språkrapporter er å undersøke hvorvidt og hvor godt disse kjerneteknologiene er utviklet for de europeiske språkene.

Vi trenger robust og rimelig språkteknologi for alle de europeiske språkene.

For å opprettholde en ledende posisjon innen global innovasjon trenger Europa en språkteknologi som er tilpasset alle europeiske språk og som er robust, rimelig og tett integrert i relevant programvare. Uten språkteknologi vil vi ikke kunne skape en effektiv, interaktiv,

multimedial og flerspråklig brukeropplevelse i overskuelig fremtid.

2.4 MULIGHETER FOR SPRÅKTEKNOLOGI

I trykketeknologiens dager besto det viktige teknologiske gjennombruddet av rask kopiering av en teksts side ved hjelp av en trykkpresse. Det omstendelige arbeidet med å slå opp, lese, oversette og oppsummere kunnskap måtte fremdeles utføres av mennesker. Ikke før Edison kunne man lagre tale, og da kun som analoge kopier.

Med språkteknologi kan man nå automatisere selve prosessene for oversettelse, innholdsproduksjon og kunnskapshåndtering for alle europeiske språk. Språkteknologi kan også bidra til intuitive talestyrte grensesnitt for husholdningsmaskiner, biler, datamaskiner og roboter. Vi er fortsatt på et tidlig stadium av utviklingen av anvendte kommersielle og industrielle applikasjoner, men FoU har skapt mange nye muligheter. For eksempel er maskinoversettelse allerede blitt rimelig nøyaktig innenfor bestemte områder, og eksperimentelle applikasjoner muliggjør flerspråklig informasjons- og kunnskapsstyring samt innholdsproduksjon for mange europeiske språk.

Som med de fleste teknologier ble de første anvendelsene innen bl.a. talebaserte brukergrensesnitt og dialogsystemer utviklet for svært spesialiserte domener, og de hadde ofte en nokså begrenset ytelse. Men det ligger store markedsmuligheter innenfor utdanningssektoren og underholdningsindustrien ved å integrere språkteknologi i spill, kulturminnesteder, skole og annen opplæring, biblioteker, osv. Mobile informasjonstjenester, datastøttet språklæring, eLæringsmiljøer, egenrederingsverktøy og plagiatkontrollprogrammer er bare noen av bruksområdene hvor språkteknologi kan spille en viktig rolle. Populariteten til sosiale medier som Twitter og Facebook illustrerer behovet for avanserte

språkteknologier som kan overvåke innlegg, oppsummere diskusjoner, analysere meningstrender, oppdage følelsesmessige reaksjoner, identifisere brudd på lover og regler eller spore misbruk.

Språkteknologi kan hjelpe oss til å bryte ned de språkbarrierer som språklig mangfold skaper

Språkteknologi representerer en enorm mulighet for EU. Den kan bidra til å håndtere flerspråkighet i Europa – det faktum at ulike språk lever i naturlig samskilt i europeiske bedrifter, organisasjoner og skoler. Men innbyggerne trenger å kommunisere på tvers av disse språkgrensene og på kryss og tvers av det felles europeiske markedet. Språkteknologi kan bidra til å bryte ned denne siste barrieren, samtidig som den støtter fri og åpen bruk av det enkelte språk. Ser man lenger framover, kan nyskapende og flerspråklig europeisk språkteknologi gi en målestokk for våre globale partnere når de utvikler sine egne flerspråklige samfunn. Språkteknologi er en form for 'hjelpemiddel'-teknologi som hjelper oss å bryte ned språklige barrierer og gjør språksamfunn mer tilgjengelig for hverandre. Et annet viktig og aktivt forskningsfelt er bruken av språkteknologi i redningsoperasjoner i katastrofeområder, hvor teknologiytelsen kan bli et spørsmål om liv og død: Fremtidens intelligente roboter med tverrspråklige funksjoner kan redde liv.

2.5 UTFORDRINGER FOR SPRÅKTEKNOLOGI

Selv om språkteknologien har gjort betydelige fremskritt de siste årene, skjer den nåværende teknologiske utviklingen og produktinnovasjonen for langsomt. Vanlige verktøy som stave- og grammatikkontroll i tekstbehandling er vanligvis enspråklige og bare tilgjengelig

for en håndfull språk. Nettbaserte maskinoversettelsjestjenester er nyttige for å få en rask oversikt over dokumentets innhold, men gir store problemer når svært nøyaktige og fullstendige oversettelser trengs. På grunn av kompleksiteten i menneskelig språk er modelleringen av naturlig språkbruk i programvare, som deretter skal testes ut i den virkelige verden, en tidkrevende og kostbar operasjon som krever en stabil finansiering. De europeiske landene må derfor være aktive i møte med de teknologiske utfordringene som et flerspråklig samfunn står overfor, gjennom aktivt å utvikle nye metoder for å fremskynde utviklingen. Dette kan være både beregningsorienterte fremskritt og teknikker som 'crowdsourcing'.

Den teknologiske utviklingen går for langsomt.

2.6 SPRÅK TILEGNELSE HOS MENNESKER OG MASKINER

For å illustrere hvordan datamaskiner håndterer naturlig språk, og hvorfor det er vanskelig å programmere dem til å prosessere ulike språk, skal vi kort se på hvordan mennesker tilegner seg første- og andrespråk, og deretter se på hvordan språkteknologiske systemer fungerer.

Mennesker tilegner seg språkkunnskap på to forskjellige måter. Babyer lærer et språk ved å lytte til samhandling mellom foreldre, søsken og andre familiemedlemmer. Fra toårsalderen produserer barn sine første ord og korte setninger. Dette er bare mulig fordi mennesker har en genetisk disposisjon til å imitere og rasjonalisere på grunnlag av det de hører.

Å lære et andrespråk på et senere stadium krever mer innsats, hovedsakelig fordi barnet ikke er omgitt av et språkfelleskap, slik det er tilfelle for morsmålet. På skolen tilegnes fremmedspråk vanligvis gjennom innarbeiding av grammatiske strukturer, ordforråd og staving.

Dette skjer ved hjelp av puggeøvelser som beskriver språklige kunnskaper gjennom abstrakte regler, tabeller og eksempler.

Mennesker tilegner seg språkkunnskaper på to forskjellige måter: læring fra eksempler og læring fra underliggende språkregler.

De to hovedtypene av språkteknologiske systemer 'tilegner' seg språklige kunnskaper på en lignende måte. Statistiske (eller 'datadrevne') tilnærminger innhenter språkkunnskap fra store samlinger av konkrete eksempletekster. For å trene stavekontrollsystemer er det tilstrekkelig å bruke tekst fra et enkelt språk, men skal en trene opp et maskinoversettelsessystem trengs et sett av parallelle tekster for to (eller flere) språk. På denne måten kan maskinen 'lære' mønstre for hvordan ord, korte setninger og fullstendige setninger blir oversatt.

En statistisk tilnærming kan kreve millioner av setninger, og kvaliteten øker jo mer tekst som analyseres. Dette er en av grunnene til at søkemotorleverandører vil samle inn så mye tekst som mulig. Tekstbehandlingsprogrammenes stavekontroller, så vel som tjenester som Google Search og Google Translate, er alle basert på statistiske metoder. Den store fordelen med statistiske metoder er at maskinen lærer raskt gjennom en kontinuerlig serie av treningsrunder, men kvaliteten er varierende.

Den andre tilnærmingen til språkteknologi, og særlig til maskinoversettelse, er å bygge regelbaserte systemer. Språkforskere, datalingvister og dataeksperter må først kode grammatiske analyser (oversettelsesregler) og sette

sammen ordlister (leksika). Dette er svært tid- og arbeidskrevende. Noen av de viktigste regelbaserte maskinoversettelsessystemene har vært under kontinuerlig utvikling i mer enn tjue år. Den store fordelen med regelbaserte systemer er at ekspertene har en bedre kontroll over maskinens språkbehandling. Dermed kan man systematisk rette opp feil i programvaren og gi brukeren detaljerte tilbakemeldinger. Dette er spesielt nyttig når systemene brukes til språklæring. Men på grunn av de høye kostnadene har regelbasert språkteknologi så langt bare blitt utviklet for store språk.

De to hovedtypene av språkteknologiske systemer tilegner seg språk på en lignende måte.

Ettersom styrkene og svakhetene ved statistiske og regelbaserte systemer ofte utfyller hverandre, fokuserer forskningen nå på hybridtilnærminger som kombinerer dem. Så langt har imidlertid bruken av disse metodene vært mindre vellykket i industrielle applikasjoner enn i forskningslaboratoriene.

I dette kapitlet har vi sett at mange vanlige dataprogrammer er avhengige av språkteknologi. Dette gjelder særlig for Europa, i kraft av å være et felles økonomi- og informasjonsområde. Selv om kvaliteten på språkteknologi har blitt mye bedre de siste årene, er det fortsatt et stort forbedringspotensial. I det følgende vil vi beskrive rollen norsk språk har i det europeiske informasjons-samfunnet og vurdere tilstanden for norsk språkteknologi.

The Norwegian Language in the Digital Age

Bokmalsversjon

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 81 p. 24 illus. in color., Softcover

ISBN: 978-3-642-31388-2