

SPRÅKA VÅRE STÅR I FARE: EI UTFORDRING FOR SPRÅKTEKNOLOGIEN

Vi er vitne til ein digital revolusjon som påverkar kommunikasjonen og samfunnet dramatisk. Den seinaste utviklinga i digital informasjons- og kommunikasjonsteknologi blir nokre gonger samanlikna med Gutenbergs oppfinning av trykkpressa. Kva kan denne analogien fortelje oss om framtida for det europeiske informasjonssamfunnet generelt og for stillinga til språka spesielt?

Vi er vitne til ein digital revolusjon som kan samanliknast med Gutenbergs oppfinning av trykkpressa.

I kjølvatnet av Gutenbergs oppfinning skjedde fleire store gjennombrøt i kommunikasjon og kunnskapsutveksling, som til dømes Luthers omsetjing av Bibelen til eige morsmål. Sidan Gutenbergs tid har ein utvikla fleire teknikkar for betre handsaming av språkbehandling og kunnskapsutveksling:

- standardisering av rettskriving og grammatikk for dei vanlegaste språka har gjeve ei hurtigare spreiding av nye vitskaplege og intellektuelle idear;
- utviklinga av offisielle språk har gjort det lettare for innbyggjarane å kommunisere innanfor visse (som oftast politiske) grenser;
- undervising og omsetjing mellom språk har bidrege til utveksling på tvers av språk;

- etablering av redaksjonelle og bibliografiske retningslinjer har sikra kvaliteten og tilgangen på trykt materiale;
- etablering av ulike medium som aviser, radio, fjernsyn, bøker og andre medium har dekt ei rekkje kommunikasjonsbehov.

Dei siste tjue åra har informasjonsteknologi bidrege til å automatisere og forenkle mange av desse prosessane:

- publiserings- og teksthandsamingsprogram har erstatta skrivemaskin og dokumentproduksjon;
- Microsoft PowerPoint har erstatta overheadtransparantar;
- e-post gjer det mogleg å sende og ta mot dokument raskare enn med ei faksmaskin;
- Skype tilbyr billeg telefonsamtaler via Internett og legg til rette for videokonferansar;
- ulike format for lagring av lydar og video gjer det enkelt å utveksle multimedia-innhald;
- søkemotorar gjer det enkelt å søkje i nettsider;
- nettbaserte tenester som Google Translate produserer raske, omtrentlege omsetjingar;
- sosiale medium som Facebook, Twitter og Google+ forenkler hurtig kommunikasjon, samarbeid og informasjonsdeling.

Sjølv om slike verktøy og program er nyttige, er dei enno ikkje i stand til fullt ut å fylle rolla som ein berebjelke

for innbyggjarane i eit fleirspråkleg europeisk samfunn, med fri flyt av informasjon og varer.

2.1 SPRÅKGRENSE HINDRAR UTVIKLINGA AV EIT EUROPEISK INFORMASJONSSAMFUNN

Vi kan ikkje føreseie nøyaktig korleis informasjonssamfunnet vil sjå ut i framtida. Men det er svært sannsynleg at den viktigaste revolusjonen i moderne kommunikasjonsteknologi vil liggje i nye måtar å samle folk som snakkar ulike språk. Dette legg press på den einskilde, som må lære nye språk, og på programutviklarar, som må lage nye applikasjonar som kan sikre gjensidig forståing og tilgjenge til deleleg kunnskap. I ein økonomi og eit informasjonssamfunn som blir stadig meir globalisert, vil nye medium føre til enklare interaksjon på tvers av språk, språkbrukarar og ulike typar innhald. Sosiale medium som Wikipedia, Facebook, Twitter, YouTube, og nyleg Google+ har vorte stadig meir utbreidde, men dette er berre toppen av isfjellet.

Ein stadig meir globalisert økonomi og informasjonssamfunn konfronter oss med fleire språk, ulike språkbrukarar og ulike typar innhald.

Ifølgje ein fersk rapport frå Europakommisjonen kjøper 57% av Internettbrukarane i Europa varar og tenester på språk som ikkje er deira eige morsmål (engelsk er det vanlegaste framandspråket, følgt av fransk, tysk og spansk). 55% av brukarane kan lese innhald på eit framandspråk, medan berre 35% bruker eit anna språk til å skrive e-post eller poste kommentarar på nettet [6]. For nokre år sidan var kanskje engelsk Internettet sitt *lingua franca*, men no er situasjonen dramatisk forandra. Mengda av nettbasert innhald på andre europeiske språk (og dessutan asiatiske og språk frå Midtausten) har eksplodert.

Denne digitale 'klasseskilnaden' mellom språka har overraskande nok ikkje fått mykje offentleg merksemd, trass i at det gjennomsyrrar heile samfunnet. Men det aktualiserer eit viktig spørsmål: Kva for europeiske språk vil overleve i eit nettverksbasert informasjons- og kunnskapssamfunn, og kva for språk er dømde til å forsvinne?

2.2 SPRÅKA VÅRE STÅR I FARE

Medan trykkjeteknologien bidrog til å auke informasjonsspreiinga i Europa, førde han òg til språkdød. Regionale og minoritetsspråk vart sjeldan trykte, slik at språk som kornisk og dalmatisk vart avgrensa til munnleg overføring, noko som i sin tur avgrensa bruksområda. Vil Internett har same verknad på språka våre?

Det språklege mangfaldet i Europa er ein av dei viktigaste delane av kulturarven vår.

Dei om lag 80 språka i Europa utgjer ein av dei viktigaste delane av kulturarven vår, og ein sentral del av den europeiske samfunnsmodellen [7]. Medan språk som engelsk og spansk sannsynlegvis vil overleve på den nye digitale marknaden, risikerer mange europeiske språk å bli irrelevante i eit nettverksbasert samfunn. Dette ville kunne svekkje Europas posisjon på verdsbasis og svekkje målet om likeverdig deltaking for alle europeiske borgarar, uavhengig av språk. Ifølgje ein UNESCO-rapport om fleirspråklegheit er språk eit viktig middel for å nyte godt av grunnleggjande rettar, som politisk ytringsfriheit, utdanning og samfunnsdeltaking [8].

2.3 SPRÅKTEKNOLOGI KAN LEGGJE TIL RETTE FOR SPRÅKBRUK

Tidlegare vart det først og fremst investert i språkopplæring og omsetjing. Utrekningar viser at den europe-

iske marknaden for omsetjing, tolking, programvarelokalisering og nettstadsglobalisering utgjorde 8,4 milliardar euro i 2008, og dette talet blir forventet å vekse med 10% årleg [9]. Men denne investeringa dekkjer berre ein liten del av det noverande og framtidige behovet for kommunikasjon mellom språk. Eit viktig tiltak for å sikre breidda og mangfaldet av språkbruk i morgondagens Europa er å bruke riktig teknologi, akkurat som vi bruker teknologi til å løyse utfordringar innan transport, energi og universell utforming.

Digital språkteknologi (retta mot alle former for tekst og munnleg tale) kan hjelpe menneske til å samarbeide, drive handel, dele kunnskap og delta i sosiale og politiske debattar på tvers av språkbarrierar og datakunnskapar. Språkteknologi er ofte innebygd i komplekse system som hjelper oss med å:

- finne informasjon med ein Internett-søkemotor;
- sjekke staving og grammatikk i eit teksthandsamingsprogram;
- vise produkttilrådingane i nettbutikkar;
- høyre taleinstruksjonar frå eit bilnavigasjonssystem;
- omsetje nettsider via nettbaserte tenester.

Språkteknologi består av ei rekkje kjerneapplikasjonar som legg til rette for ulike prosessar innanfor eit større applikasjonsrammeverk. Føremålet med META-NETs språkrapportar er å undersøkje i kva grad og kor godt desse kjerneteknologiane er utvikla for dei europeiske språka.

Vi treng robust og rimeleg språkteknologi for alle dei europeiske språka.

For å oppretthalde ein leiande posisjon i global innovasjon treng Europa ein språkteknologi som er tilpassa alle europeiske språk og som er robust, rimeleg og tett integrert i relevant programvare. Utan språkteknologi vil vi

ikkje kunne skape ei effektiv, interaktiv, multimedial og fleirspråkleg brukaroppleving i den nære framtida.

2.4 SPRÅKTEKNOLOGI GJEV MOGLEGHEITER

I ei verd basert på trykkjeteknologi var det viktige teknologiske gjennombrøtet rask kopiering av ei tekstsider ved hjelp av ei trykkpresse. Det omstendelege arbeidet med å slå opp, lese, omsetje og oppsummere kunnskap måtte framleis utførast av menneske. Ikkje før Edison kunne ein lagre tale, og då berre som analoge kopiar.

Digital språkteknologi kan no automatisere sjølve omsetjingsprosessen, innhaldsproduksjon og kunnskapshandsaming for alle europeiske språk. Språkteknologi kan òg bidra til intuitive talestyrte grensesnitt for hushaldningsmaskiner, bilar, datamaskiner og robotar. Vi er enno på eit tidleg stadium av utviklinga av å bruke kommersielle og industrielle applikasjonar, men FoU har skapt mange nye høve. Til dømes er maskinomsetjing alt vorten rimeleg nøyaktig innanfor visse område, og eksperimentelle applikasjonar mogleggjer fleirspråkleg informasjons- og kunnskapsstyring og dessutan innhaldsproduksjon for mange europeiske språk.

Som med dei fleste teknologiane vart den første bruken innan bl.a. talebaserte brukargrensesnitt og dialogsystem utvikla for svært spesialiserte domene, og dei hadde ofte ei nokså avgrensa yting. Men det ligg store marknadspotensial innanfor utdanningssektoren og underhaldningsindustrien ved å integrere språkteknologi i spel, kulturminnestader, skule og anna opplæring, bibliotek, osv. Mobile informasjonstenester, datastøtta språklæring, eLæringsmiljø, eigenvurderingsverktøy og plagiattkontrollprogram er berre nokre av bruksområda der språkteknologi kan spele ei viktig rolle. Populariteten til sosiale medium som Twitter og Facebook illustrerer behovet for avanserte språkteknologiar som kan overvake innlegg, oppsummere diskusjonar, analy-

sere meiningstrendar, oppdage kjenslereaksjonar, identifisere brot på lover og reglar eller spore misbruk.

Språkteknologi kan bidra til å bryte ned språkbarrierane som det språklege mangfaldet skaper.

Språkteknologi representerer eit enormt potensial for EU. Han kan bidra til å handsame fleirspråklegheit i Europa – det faktumet at ulike språk lever i naturleg sameksistens i europeiske føretak, organisasjonar og skular. Men innbyggjarane treng å kommunisere på tvers av desse språkgrensene og på kryss og tvers av den felles europeiske marknaden. Språkteknologi kan bidra til å overvinne denne siste barrieren samstundes som han støttar fri og open bruk av det einskilde språket. Ser ein lenger framover, vil ein nyskapande og fleirspråkleg europeisk språkteknologi gje ein målestokk for dei globale partnerane våre når dei utviklar sine eigne fleirspråklege samfunn. Språkteknologi er ei form for ‘hjelpemiddel’-teknologi som hjelper oss å bryte ned språklege barrierar og gjere språksamfunn meir tilgjengelege for kvarandre. Eit anna viktig og aktivt forskingsfelt er nytta av språkteknologi i redningsoperasjonar i katastrofeområde, der teknologiying kan bli eit spørsmål om liv og død: Framtida sine intelligente robotar med tverrspråklege funksjonar kan redde liv.

2.5 UTFORDRINGAR FOR SPRÅKTEKNOLOGI

Sjølv om språkteknologien har gjort betydelege framsteg dei siste åra, skjer den noverande teknologiske utviklinga og produktinnovasjonen for sakte. Vanlege verktøy som stave- og grammatikkontroll i tekstbehandling er vanlegvis einspråklege og berre tilgjengelege for ei handfull språk. Nettbaserte maskinomsetjingstenester er nyttige for å få eit rask oversyn over innhaldet i do-

kumentet, men gjev store problem når svært nøyaktige og fullstendige omsetjingar trengst. På grunn av kompleksiteten i menneskeleg språk er det å modellere naturleg språkbruk i programvare for deretter å teste det ut i den verkelege verda ein tidkrevjande og kostbar operasjon som krev ei stabil finansiering. Dei europeiske landa må difor vere aktive i møte med dei teknologiske utfordringane eit fleirspråkleg samfunn står overfor gjennom aktivt å utvikle nye metodar for å skunde på utviklinga. Dette kan vere både uttrekningsorienterte framsteg og teknikkar som ‘crowdsourcing’.

Den teknologiske utviklinga går for langsamt.

2.6 SPRÅKTILEIGNING HOS MENNESKE OG MASKINER

For å illustrere korleis datamaskiner handsamar naturleg språk, og kvifor det er vanskeleg å programmere dei til å prosessere ulike språk, skal vi kort sjå på korleis menneske tileignar seg første- og andrespråk, og deretter sjå korleis språkteknologiske system fungerer.

Menneske tileignar seg språkkunnskap på to ulike måtar. Babyar lærer eit språk ved å lytte til samspel mellom foreldre, sysken og andre familiemedlemmer. Frå toårsalderen produserer born dei første orda sine og korte setningar. Dette er berre mogleg fordi menneske har ein genetisk disposisjon til å imitere og rasjonalisere på grunnlag av det dei høyrer.

Å lære eit andrespråk på eit seinare stadium krev meir innsats, hovudsakleg fordi barnet ikkje er omgjeve av ein språkfelleskap, slik det er tilfelle for morsmålet. På skulen tileignar ein seg vanlegvis framandspråk gjennom å innarbeide grammatiske strukturar, ordtilfang og staving. Dette skjer ved hjelp av puggeøvingar som skildrar språklege kunnskapar gjennom abstrakte reglar, tabellar og døme.

Menneske tileignar seg språkkunnskap på to ulike måtar: Læring frå døme og læring frå underliggjande språkreglar.

Dei to hovudtypane av språkteknologiske system ‘tileignar’ seg språklege kunnskapar på ein liknande måte. Statistiske (eller ‘datadrivne’) tilnærmingar innhentar språkkunnskap frå store samlingar av konkrete eksempletekster. For å trene stavekontrollsystem er det tilstrekkeleg å bruke tekst frå eit enkelt språk, men skal ein trene opp eit maskinomsetjingssystem, treng ein eit sett av parallelle tekster for to (eller fleire) språk. På denne måten kan maskina ‘lære’ mønster for korleis ord, korte setningar og fullstendige setningar blir omsette.

Ei statistisk tilnærming kan krevje millionar av setningar, og kvaliteten aukar jo meir tekst som blir analysert. Dette er ein av grunnane til at søkemotorleverandørar vil samle inn så mykje tekst som mogleg. Tekstbehandlingsprogramma sine stavekontrollar, så vel som tenester som Google Search og Google Translate, er alle baserte på statistiske metodar. Den store fordelene med statistiske metodar er at maskina lærer raskt gjennom ein kontinuerleg serie av treningsrundar, men kvaliteten er varierande.

Den andre tilnærminga til språkteknologi, og særleg til maskinomsetjing, er å byggje regelbaserte system. Språkforskarar, datalingvistar og dataekspertar må først kode grammatiske analysar (omsetjingsreglar) og setje saman ordlister (leksikon). Dette er svært tid- og arbeidskrev-

jande. Nokre av dei viktigaste regelbaserte maskinomsetjingssystema har vore under kontinuerleg utvikling i meir enn tjue år. Den store fordelene med regelbaserte system er at ekspertane har ein betre kontroll over maskina si språkhandsaming. Dimeit kan ein systematisk rette opp feil i programvara og gje brukaren detaljerte tilbakemeldingar. Dette er spesielt nyttig når systema skal brukast til språklæring. Men på grunn av dei høge kostnadene har regelbasert språkteknologi så langt berre vorte utvikla for store språk.

Dei to hovudtypane av språkteknologiske system tileignar seg språk på ein liknande måte.

Sidan styrkane og veikskapane ved statistiske og regelbaserte system ofte utfyller kvarandre, fokuserer forskinga no på hybridtilnærmingsmåtar som kombinerer dei. Så langt har likevel nytta av desse metodane vore mindre vellukka i industrielle applikasjonar enn i forskingslaboratoria.

I dette kapitlet har vi sett at mange vanlege dataprogram er avhengige av språkteknologi. Dette gjeld særleg for Europa, i kraft av å vere eit felles økonomi- og informasjonsområde. Sjølv om kvaliteten på språkteknologi har vorte mykje betre dei siste åra, er det enno eit stort forbedringspotensial. Under vil vi skildre rolla norsk språk har i det europeiske informasjonssamfunnet og vurdere tilstanden for norsk språkteknologi.

The Norwegian Language in the Digital Age

Nynorskversjon

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

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

ISBN: 978-3-642-31432-2