

Unifob AKSIS lager nettbaserte ordbøker for innvandrere. LEXIN-prosjektet sikret dem Rosings språkpris.

Det interaktive bildet på dataskjermen viser et supermarked omgitt av ulike spiselige produkter. Et raskt klikk på den avlange matvaren med stekeskorpe forteller deg at det du omtaler som *rooti* på ditt somaliske morsmål, heter *brød* på norsk.

– Mange ordbøker blir gjort tilgjengelig på nett uten at man klarer å utnytte de mulighetene som digital distribusjon gir. I trykte bøker må man ta plasshensyn, men piksler er gratis. På nettet kan man dermed bruke bilder og illustrasjoner på en helt annen måte. Ikke minst kan man bruke hyperlenking til å koble bildene til ordbokoppslag, sier Sindre Sørensen, som er utvikler hos AKSIS.

Unifobs Avdeling for kultur, språk og informasjonsteknologi har i flere år produsert lavterskel-ordbøker for innvandrere på oppdrag for Utdanningsdirektoratet. Hittil foreligger det nettbaserte LEXIN-ordbøker på arabisk, engelsk, kurmanji, sorani, persisk, somali, tamil, tigrinja og tyrkisk med både bokmål og nynorsk som kildepråk. Fire av ordbøkene finnes også i trykte utgaver.

– Det praktiseres trolig rundt 150 språk i Norge, men mange av disse har så få brukere at kommersielle aktører ikke har mulighet til å satse på dem. Dermed må staten ta ansvar, sier Tove Bjørneset, som er prosjektleder for LEXIN-prosjektet. Ettersom vi bruker et statisk kildepråk, kan vi også relativt raskt og billig produsere nye ordbøker med andre språk.

I skrivende stund er ordbøker på thai og urdu i ferd med å ferdigstilles, mens russisk og polsk er i startfasen. I tillegg jobbes det blant annet med å utvikle et søkegrensesnitt for mobiltelefoner. Enkelhet og brukervennlighet er hele veien sentrale stikkord, og i 2008 ble LEXIN belønnet med Rosings språkpris for nettopp dette. Juryen for prisen, som deles ut av Den norske dataforening og Språkrådet, berømmet LEXIN for en ujalte, rask og praktisk tjeneste av stor verdi for brukerne.

– Vi er i kontinuerlig dialog med brukerne og lytter til deres innspill. At vi i fjor lanserte nynorsk som kildepråk for samtlige ordbøker, var ikke et resultat av statlige føringer, men derimot nærmest et krav fra innvandrere bosatt i nynorskrområder. Dette viser hvor stor verdi språk har både for vår



Prosjektleder Tove Bjørneset og utvikler Sindre Sørensen.

identitet og som en kulturell fellesnevner, sier Tove Bjørneset.

Egne verktøy

Ideen om et statlig ordbokprosjekt for innvandrere er opprinnelig svensk og mer enn 30 år gammel. Den svenske løsningen ble så populær at både Norge, Danmark og Island takket ja til tilbudet om lån av materialet for å utvikle tilsvarende språkressurser. Etter et vellykket pilotprosjekt med import av svensk kildepråksmateriale 1996–1999 og deretter gjenbruk av utvalgte oversettelser, ble de første nettordbøkene lansert i 2002. De to énspråklige norske ordbøkene og den norsk–engelske nettordboken består av 36 000 oppslagsord, sammensetninger og avledninger, mens de andre ordbøkene inneholder mellom 7 600 og 11 000.

– LEXIN er en ressurs som hjelper brukerne opp på et språklig nivå som gjør dem bedre i stand til både å forstå og å bli forstått. På sikt vil de kunne bruke andre og større

ordbøker, noe som er helt nødvendig for å få et godt norsk språk, sier Tove Bjørneset. Unifob AKSIS har selv utviklet redigeringsverktøyet, som lar oversetterne redigere ordboksmaterialet via en egen nettside. Oversetterne kan dermed sitte hvor som helst i verden og jobbe på LEXIN. Også Utenriksdepartementet bruker verktøy fra AKSIS, blant annet i forbindelse med sin terminologidatabase EØS-EU-basen.

– I tverrfaglige prosjektgrupper får blant annet vi filologer mulighet til å komme med innspill og ønsker til utviklere og programmerere, og vice versa. Dette kontinuerlige samspillet om utvikling og testing av skreddersydde verktøy og ressurser resulterer ofte i banebrytende løsninger. I tillegg er det en betydelig trivselsfaktor for oss alle i en hektisk arbeidshverdag. Erfaringsmessig er det mange utenfor vårt fagmiljø som har stort behov for vår tverrfaglige kompetanse uten å være klar over det, sier Bjørneset.



FAKTA

Unifob AKSIS driver forskning og utvikling på følgende områder:

Språkteknologi. Forskning på og utvikling av språkressurser, som verktøy for termekstraksjon, grammatikker, annotering og korpusverktøyer. Vi utvikler også nettbaserte ordbøker (LEXIN) og ulike løsninger for språkprosessering.

Språktesting. Utvikling av språktester i samsvar med internasjonale standarder for fremmedspråk på alle nivåer. Vi driver forskning på språktesting, testmetoder og standardisering.

Elektronisk publisering. Forskning på semantisk teknologi for å effektivisere søk i komplekse tekstsamlinger. Vi utvikler blant annet infrastrukturer for elektronisk publisering av store tekstsamlinger, som for eksempel Wittgensteinarkivet.

Digitale medier og elektronisk læring. Forskning på digitale medier og deres pedagogiske potensialer og begrensninger. Vi analyserer sosiale og kulturelle aspekter ved nye medier, samt utvikler teknologier og metoder for testing av disse.

Høydepunkter

- Unifob AKSIS bidro til publisering av flere berømte personers tekster: Henrik Ibsens dikt er en del av prosjektet Henrik Ibsens skrifter (utgivelse 2009). På eMunch.no ble de første sidene av Edward Munchs korrespondanse offentliggjort. 5000 sider av Ludwig Wittgensteins etterlatte skrifter blir gjort åpent tilgjengelig på wittgensteinsource.org.
- LFG-Parserbanker – et verktøy for oppbygging av databaser av grammatisk analyserte setninger ble ferdigstilt og lisensiert til flere firma og universiteter.
- LEXIN – ordbøker for minoritetsspråk – vant Rosings språkpris. Disse har vi utviklet siden 1996.
- Stortingsmeldinga *Mål og Mening* ble positivt mottatt av språkforskningsmiljøene. Særlig interessant for oss er prioriteringen av norsk terminologi og opprettelsen av en norsk språkbank.
- Eli Hagen tok over som ny forskningsdirektør 1. oktober. Hun kommer fra BMWs forsknings- og utviklingssenter i München, hvor hun jobbet med talespråk og var program manager for forskning og utvikling av interiørkomponenter.

Utvalgte prosjekter

Gjennom prosjektet **Semi-SemAntic Models for cross-sector portals (SeSAM4)** er avdelingen med på å gjøre det enklere for små og mellomstore bedrifter å nyttiggjøre seg mulighetene som ligger i semantisk web.

EU-prosjektet **Common LAnguage Resources and technology INfrastructure (CLARIN)** har som mål å etablere en elektronisk forskningsinfrastruktur for språkressurser til nytte for humaniora og samfunnsvitenskap.

I EU-prosjektet **Science Created by You (SCY)**, skal avdelingen forske på hvordan digitale læringsmiljøer kan bidra til å gjøre læring av naturvitenskap mer interessant.

Prosjektet **Termportalen** etablerer en nasjonal infrastruktur for norsk fagspråk og terminologi.

Samarbeidspartnere

Unifob AKSIS samarbeider med flere universiteter – særlig HF og SV fakultetet ved UiB. Andre samarbeidspartnere og oppdragsgivere er Forskningsrådet, EU, Språkrådet, næringslivet og det offentlige Norge.

