

For ansatte

Det humanistiske fakultet

Søk

Forsiden

Aktuelt

Ansettelsesforhold

Arbeidsstøtte

Drift og servicefunksjoner

Kompetanseutvikling

Organisasjon

Aktuelt

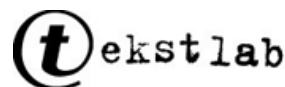
Aktuelle saker

- 2010
- 2011
- 2012
- 2013
- 2014
- 2015

Midler til ny språkteknologi

Tekstlabs prosjekt Language Infrastructure made Accessible (LIA) får 27.3 millioner av Norges forskningsråd. Prosjektet samler opptak av norsk og samisk språk i en søkbar database.

Tildelingen er del av en større [investering i moderne forskningsinfrastruktur fra Forskningsrådet](#). Totalt en halv milliard kroner deles ut til 16 prosjekter for å finansiere laboratorier, databaser og utstyr.



Utvikling av språkteknologi

Gamle lydopptak på norsk og samisk skal samles og tilgjengeliggjøres i en felles database. Formålet med å samle opptakene, transkribere og tagge dem i et korpus, er at forskerne skal kunne søke på ord og tekstbiter og lytte på tilhørende lydklipp.

Materialet vil være et verdifullt tilskudd i videreutviklingen av programvare for taleteknologi, språkgjenkjenning, maskinoversettelse og dialogsystemer.

– Materialet legges inn i korpussøkesystemet Glossa, som er utviklet av Tekstlab, og som brukes av andre korpus i dag. Det blir en verdifull empirisk kilde for lingvister og etnografer. Dessuten blir materialet gjort tilgjengelig for separat teknologiutvikling der det er behov for mer trenings- og testmateriale. For eksempel ved utvikling av taleteknologi. Siden materialet er gammelt, er det juridisk friere fra et rettighetsperspektiv, forteller Bondi Johannessen.

Tekstlaboratoriet skal transkribere opptakene fra muntlig til skriftlig form, lydrett og ortografisk.

– Vi har utviklet et digitalt spesialverktøy, en dialekt-transliteratør, som oversetter fra lydskrift til rettskriving.

Prosjektleder for LIA, professor Janne Bondi Johannessen.

Gjøre talespråket tilgjengelig

– Materialet ligger i tilgjengelige formater, som gamle lydbånd eller voksruller. Disse må digitaliseres for at forskerne skal kunne høre på opptakene, forklarer Bondi Johannessen.

Nasjonalbiblioteket er i gang med å digitalisere voksruller med opptak av norsk i Amerika, som ble gjort av språkforskerne Seip og Selmer på 1930-tallet. Disse vil bli gjort tilgjengelige sammen med gamle opptak gjort av språkforskeren Einar Haugen.

Med prosjektet blir også samisk talespråk samlet i en felles database for første gang.

– Jeg er veldig glad for tildelingen. Å samle gamle og nye data åpner for historiske studier av språkene, som ikke har vært mulig tidligere.

Om Language Infrastructure made Accessible (LIA):

LIA er et samarbeid mellom Tekstlaboratoriet ved ILN, Universitetet i Bergen, Universitetet i Tromsø, NTNU, Norsk Ordbok 2014, Nasjonalbiblioteket og UNINETT Sigma. I tillegg er Humboldt-universitetet (Tyskland), University of Wisconsin, Pennsylvania State University (USA) og universitetene i Odense, Uppsala og Gøteborg tilknyttet prosjektet.

Prosjektet skal vare i fire år.

Prosjektleder er er Janne Bondi Johannessen.

Av Vibeke Enstrøm Nøking, informasjonskonsulent, ILN
Publisert 6. nov. 2013 13:31 - Sist endret 27. okt. 2014 12:00

 Del på e-post  Del på Facebook  Del på Twitter

UiO > Det humanistiske fakultet > For ansatte > Aktuelt

Aktuelle saker	Leder	Signert	2 minutter med	Arkiv HF-aktuelt	Arrangementer
2010	2010	2010	2010	2012	
2011	2011	2011	2011	2011	
2012	2012	2012	2012	2010	
2013	2013		2013	2009	
2014	2014		2014	2008	
Mer...	Mer...		Mer...	Mer...	

Forskningsrådet	Fra UH-sektoren	HFSU	Leserinlegg	Min kollega	Publikasjoner
2012		2010	2010	2013	
2013		2011		2014	

[Undervisningsrådet](#)

2012
2013

UiO : Det humanistiske fakultet



Besøksadresse
Det humanistiske fakultet
(kart)
Niels Henrik Abels vei 36
0313 OSLO

Postadresse
Postboks 1079
Blindern
0316 OSLO

Telefon, faks, e-post
228 56293
228 54550
postmottak@hf.uio.no

Ansvarlig for denne siden:
[Nettredaksjonen på HF](#)
[Logg inn](#) | ▾

Les om [hvordan UiO behandler personopplysninger](#), hvorfor vi bruker informasjonskapsler, og hvordan du kan reservere deg 