

Kronikk/tirsdag

LES MER PÅ db.no/meninger

KRONIKK: Arne Dvergsdal
Skriv til: kronikk@dagbladet.no
Kronikk lengden er fra 6000 til 7000 tegn
 inklusive ordmellomrom.
Kronikken på nett: www.dagbladet.no/kronikk

Nye utfordringer i språkforskningen

Moderne språkforskning forutsetter store dataressurser og søkbare elektroniske tekstsamlinger. Slike korpus er kostnadskrevenne å bygge opp, skriver kronikkforfatterne som etterlyser en langt mer aktiv holdning fra Forskningsrådet på dette feltet.



KRONIKK

Janne Bondi Johannessen

Universitetet i Oslo



Øystein Alexander Vangsnes

Universitetet i Tromsø



Tor Anders Åfarli

Norges teknisk-naturvitenskapelige universitet

FORSKNING I HUMANIORA generelt og norsk språk spesielt er i ferd med å bli en stor taper i kampen om forskningsmidlene. Skal fremtidig språkforskning bare dreie seg om engelsk? Skal vi holde oss til foreldede forskningsmetoder og bli liggende i en bakejevne både hva gjelder metodologi og kunnskap? Dette kan bli resultatet dersom ikke forskningsmidler også kanaliseres inn mot våre virksomheter.

Moderne informasjonsteknologi har åpnet opp uante muligheter for nye forskningsmetoder innenfor et vidt spekter av vitenskaper. Elektroniske arkiv og samlinger, databaser og søkeverktøy gjør at man nå etter noen minutter eller timer foran datamaskinen kan finne fram til informasjon det ville ta år og dag med reising og leting i fysiske arkiver å framskaffe – hvis det i det hele tatt hadde latt seg gjøre å finne. Det kalles e-infrastruktur og man har begynt å snakke om e-vitenskap og e-forskning som egne begreper.

MEN DET KOSTER å bygge opp e-infrastruktur. Det synes Norges forskningsråd til en viss grad å ha innsett, og i disse dager lanseres et eget forskningsprogram hos NFR kalt eVITA, som ifølge programplanen er «et forsknings- og infrastrukturprogram innrettet mot beregnings- og dataintensive utfordringer i naturvitenskap, teknologi og medisin.» Det er liten tvil om at

denne satsingen på e-vitenskap er en oppfølging av den siste forskningsmeldingens fokus på realfaglig forskning. Men e-vitenskap er i alle høyeste grad også relevant og nødvendig for humaniora. Dessverre har NFR så langt vist liten forståelse for at også humanistisk forskning bør bringes inn i den nye e-vitenskapelige æra.

Vi som arbeider med oppbygging av språkressurser for forskning, ser nemlig igjen og igjen at behovene våre oversees, og våre sonderinger mot Forskningsrådet tyder på at eVITA i første omgang i praksis vil være lukket for søknader fra annet hold enn medisinske og matematisk-naturvitenskapelige fag. De ordinære midlene under Forskningsrådets humaniorapott FRIHUM samt de smalere forskningsprogrammene er alle stengt for søknader om ressursoppbygging fordi dette ikke ansees som forskning. Forskningsrådets dører synes altså stengt. Samtidig har det vist seg umulig å få penger gjennom de ordinære utstyrsbevilgningene hos universitetene og Forskningsrådet.

DEN ANBEFALTE årlige rammen for eVITA er ifølge programplanen 100 millioner, mens Forskningsrådet kun bevilget 23 millioner NOK til nye forskningsprosjekt innenfor humaniora i 2006. Vi tre stod bak en søknad til sistnevnte pott, om et prosjekt som kombinerer ressursoppbygging og teoretisk lingvistisk forskning, men prosjektet fikk ikke bevilgning til tross for at det av tre uavhengige eksperter fikk karakterene sju, sju og seks på en skala hvor sju var høyest, til tross for at prosjektet inngår som en del av et større nordisk prosjekt der i alt ni forskergrupper samarbeider, til tross for at det nordiske nettverket nylig ble utnevnt til nordisk senter for fremragende forskning, og til tross for at det i sin tur er en del av et større europeisk nettverk. Hovedbegrunnelsen for avslaget var at prosjektet var for omfattende!

Tradisjonelt har det nok vært slik at humanistiske fag ikke er typisk utstyrskrevenne. Men situasjonen er en annen nå. Forskningsmetodene for moderne språkforskning og språkteknologiforskning forutsetter store dataressurser, elektronisk søkbare tekstsamlinger – dvs. elektroniske tekstsamlinger som er annotert på ulike

måter. Slike korpus er kostnadskrevenne å bygge opp, men på tross av det har det skjedd en viss oppbygging av korpus med skriftspråkstekster (både ved UiB og UiO). Dette skyldes at mange typer skriftspråkskorpus i stor grad tross alt kan bygges opp ved å benytte et ressursmateriale som er tilgjengelig allerede, slik som ferdige tekstfiler fra aviser og forlag. Slike ressurser er i utgangspunktet billige. Men all ekstra annotering for å gjøre korpusene mest mulig brukbare for språkforskning, samt utviklingen av gode brukergrensesnitt for søking, kommer selvsagt i tillegg, og gjør at også slike korpus er kostnadskrevenne å bygge opp.

SITUASJONEN er en helt annen for talespråk. Her finnes det ikke noe ferdig tilgjengelig materiale – det materialet som man kunne tenkt seg å overta fra radio og tv er i såpass stor grad manipulert underveis og etterpå, og talerne selv tatt opp i en så kunstig situasjon, at denne typen materiale ikke egner seg som eksempel på vanlig talespråk. For å bygge opp talespråkskorpus, må man begynne helt fra bunnen av. Man må finne informanter og gjøre opptak med dem, man må bearbeide opptakene digitalt etterpå. For at korpusene skal bli søkbare, må opptakene transkriberes med standardtranskripsjon (ortografi), slik at hvert eneste ord også gjøres tilgjengelig i

skrift. Uttalenær transkripsjon er i mange tilfeller dessuten også ønskelig. Både opptak og transkripsjon er svært kostnadskrevenne. I tillegg kommer selvsagt de samme kostnadene som for skriftspråkskorpus, knyttet til annotering og søkegrensesnitt.

At slike korpus er viktige, er hevet over tvil. Uten dem vil norske forskere ikke kunne være i forskningsfronten når det gjelder vårt eget språk. Både innen allmenn språkforskning, dialektforskning og språkteknologiforskning er det viktig at man har etterrettelige data, og i mange tilfeller også kvantitative data som gjør statistiske metoder mulige. Uten gode talespråksdata vil all forskning som gjøres om norsk språk i verste fall være om norsk skriftspråk, som jo er en annen og mye mindre spontan variant enn den vi uttrykker oss på gjennom de fleste av livets situasjoner, eller om engelsk, hvor store datamengder finnes. Fra et kulturperspektiv vil det være et enestående tap dersom kunnskap om og forskning på dialekter og talespråk skal falle bort. Samtidig er det faktisk en kjensgjerning at det allerede i dag finnes store hull i kunnskapen om både talespråk og dialekter, hull som aldri har blitt fylt, og som først nå kunne blitt det fordi det i dag finnes et godt begreps- og teoriapparat i den teoretiske språkvitenskapen samt gode systemer for behandling av store datamengder i form av databaseteknologi.

SLIKE DATABASES kan ikke kjøpes fiks ferdige! De må bygges opp. Ingen andre land tar ansvar for oppbyggingen av norske språkdatabaser. Det må gjøres av oss ved de norske forskningsinstitusjonene. Derfor er det mer enn bare litt foruroligende når alle deler av Forskningsrådet fraskriver seg ansvaret for dette. Vi etterlyser en langt mer aktiv holdning fra Forskningsrådet på dette området. Det er av avgjørende betydning både for norsk forskning og for norsk kultur, og selv om vårt spesifikk anliggende gjelder språkvitenskapelig e-forskning, bør bekymringene deles også av forskere innenfor andre human- og samfunnsvitenskapelige fag.

KRONIKK I MORGEN

Professor Svein Sjøberg spør om norsk skole skal kopiere Singapore.

«Slike databaser kan ikke kjøpes fiks ferdige!»



NY TEKNOLOGI: Norsk språkforskning krever moderne informasjonsteknologi. Foto: Lars Eivind Bones

Dagbladet TIPSTELEFONEN 22 20 00 00

Akersgata 49, Oslo.
 Postboks 1184 Sentrum, 0107 Oslo.

TIPSTELEFONEN
22 20 00 00

SENTRALBORD: 22 31 06 00

Direkte linjer:

Sentralredaksjonen: 22 31 08 90

Vaktsjef: 22 31 07 81

Utenriksredaksjonen: 22 31 07 72

Sporten: 22 31 07 98

Deskfaktør: 22 31 07 22

Resepsjon: 22 31 08 88

Telegramadr.: Dagbladet Oslo

1000-tipset:

1000Tipset@dagbladet.no

Kort&Godt:

kortoggodt@dagbladet.no

På din side: pds@dagbladet.no

Redaksjonen: 22 42 95 48

Produksjonen (Desk): 22 31 06 05

Annonseavdelingen: 22 31 05 05

Administrasjonen: 22 31 05 60

Økonomiavdelingen: 22 31 06 06

Kundeservice: 22 31 09 25

Kulturavdelingen: 22 31 09 66

Lederavdelingen: 22 31 05 20

Kundeservice: 800 30 100

E-post: kundeservice@dagbladet.no

abonnement@dagbladet.no

Dagbladsentralen: 22 31 06 82

AS Dagblad-Trykk: 67 92 75 00

UTLANDET:

LONDON:

Lars Inge Staveland

las@dagbladet.no

13 Marlborough Road

Ealing W5 5NY, London

Telefon: +44 20 85 67 76 29

Faks: +44 20 88 40 68 13

JERUSALEM:

Yngvil Mortensen

ymo@dagbladet.no

Telefon: +972 26 28 42 61

Mobil: +972 5 42 08 69 52

BRUSSEL:

Einar Hagvåg, eha@dagbladet.no

Rue Haute 165, 1000 Brussel

Telefon: +32 (0) 25 13 52 46

Mobil: +32 (0) 4 73 97 03 86

NEW YORK:

Kirsten Margrethe Buzzi,

kbu@dagbladet.no

Telefon: 0047 91 32 25 73

ST. PETERSBURG:

Morten Strand, mstr@dagbladet.no

Telefon: +7 81 23 14 73 08

Mobil: +7 92 17 41 47 35

Faks: +7 81 23 14 73 08

LOKALREDAKSJONER:

TRONDHEIM:

Fjordgt. 82, 7010 Trondheim.

Telefon: 73 52 87 09.

Telefaks: 73 52 25 03

Kjell-Ivar Myhr, kim@dagbladet.no

Morten Stokkan, mst@dagbladet.no

BERGEN:

Leif Stang, lst@dagbladet.no

Farbotslia 8, 5124 Morvik

Telefon: 90 54 17 94

STAVANGER:

Nygt. 24, 4006 Stavanger.

Telefon: 51 89 61 11

Telefaks: 51 89 37 77.

Leiv Gunnar Lie, llie@dagbladet.no

Thomas Ergo, ter@dagbladet.no

Telefon: 91 60 04 34

SKIEN:

Bernt Blomquist, bb@dagbladet.no

Nygata 1, 3724 Skien.

Telefon: 91 60 04 14

KRISTIANSAND:

Eivind Pedersen, epe@dagbladet.no

Markens gt. 48, Postboks 659.

4666 Kristiansand.

Telefon: 38 02 11 55

Telefaks: 38 02 09 92

db.no

www.dagbladet.no

Dagbladets utgave på Internett.

Utgitt av DB Medialab AS

Redaktør/adm.dir.: Rune Røsten

run@dagbladet.no

Redaktør: Esten O. Sæther

eos@dagbladet.no

Kontakt Dagbladet.no:

Epost: nettred@dagbladet.no

Sentralbord: 22 31 06 00

Direkte telefon: 22 31 09 59

Dagbladet forbeholder seg retten til å distribuere innsendte innlegg i elektroniske formater.

PFU

• Dagbladet arbeider etter Vær Varsom-plakatens regler for god presseskikk. Den som mener seg rammet av urettmessig avisomtale, oppfordres til å ta kontakt med redaksjonen.
 • Pressens Faglige Utvalg (PFU) er et klageorgan som behandler klager mot pressen i presseetiske spørsmål.
 • Adr.: Rådhusgt. 17 3. etg., pb 46 Sentrum, 0101 Oslo. Tlf. 22 40 50 40, faks 22 40 50 55. E-post: pfu@presse.no