

# Nordisk dialektkorpus

Janne Bondi Johannessen

*Universitetet i Göteborg 5. nov. 2008*

	<p><b>NoTa-Oslo</b> Norsk talespråkskorpus-Oslo-delen</p>	<p><a href="#">Kort info</a></p>
	<p><b>TAUS</b> Talemålsundersøkelsen i Oslo</p>	<p><a href="#">Kort info</a></p>
	<p><b>BigBrother</b></p>	<p><a href="#">Kort info</a></p>
	<p><b>Nordisk dialektkorpus - Scandinavian Dialect Corpus</b></p>	<p><a href="#">Kort info</a></p>
	<p><b>UPUS</b> Utviklingsprosesser i urbane språkmiljøer</p>	<p><a href="#">Kort info</a></p>

# Nordisk dialektkorpus - Scandinavian Dialect Corpus

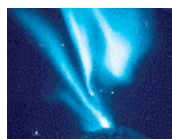
Nordisk dialektkorpus - Scandinavian Dialect Corpus

-2007 - 2008: Tekstlaboratoriet utvikler første del av et korpus og en talespråksdatabase for nordiske dialekter. Vi samler også inn materiale til et norsk dialektkorpus fra ca 50 målepunkter (dvs ca 200 informanter) rundt omkring i Norge. Innsamlingen skjer i samarbeid med våre partnere ved UiT og NTNU.

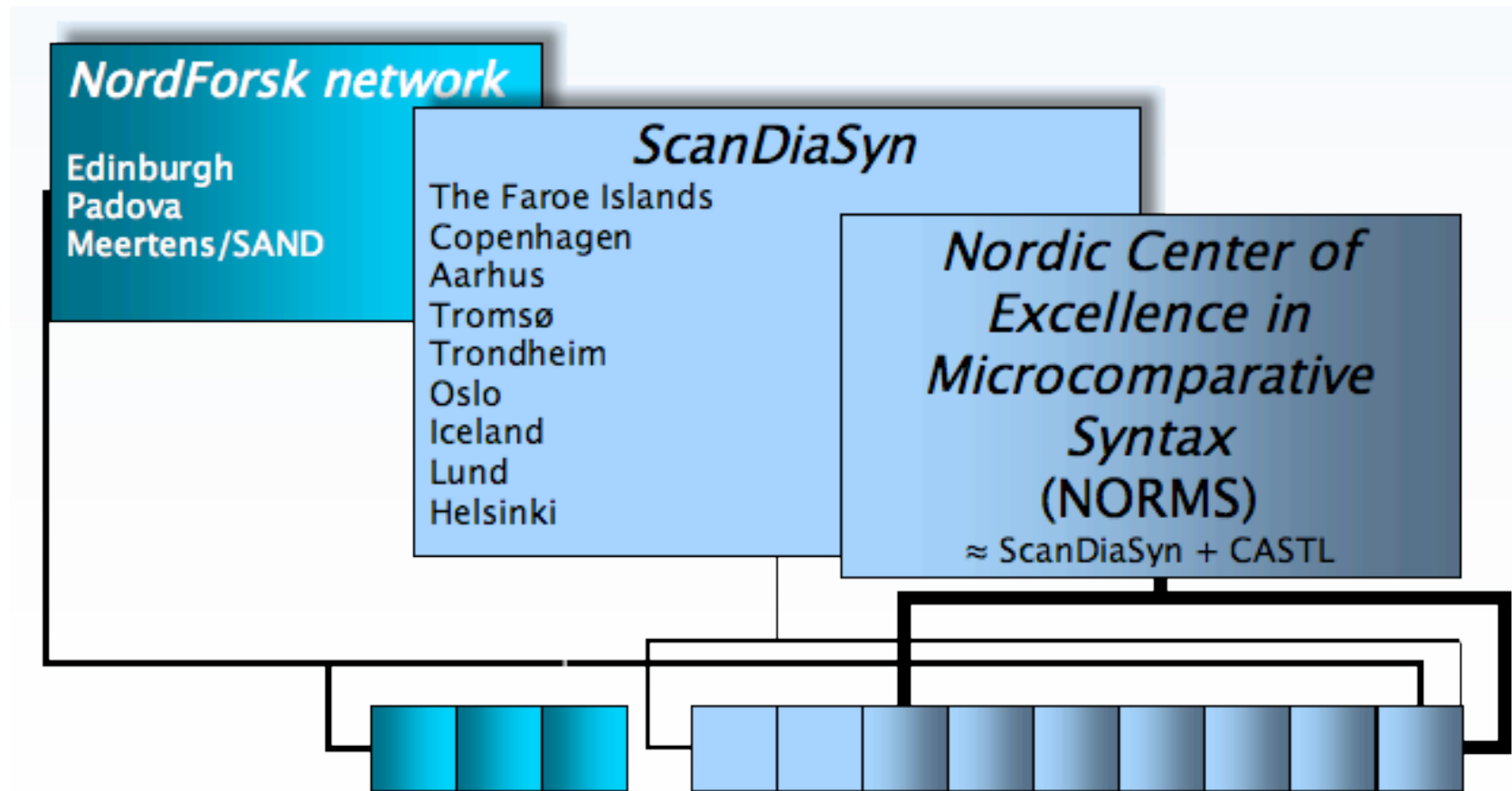
2008 - 2009: Den norske delen av dialektkorpuset utvides med 25 målepunkter.

Det norske dialektkorpuset vil bestå av intervjuer og samtaler som skal transkriberes og kobles sammen med tilsvarende lyd- og videoopptak. Det vil også bli brukt spørreskjemaer som skal kartlegge ulike syntaktiske forhold i informantenes talemål. Korpuset vil dessuten bestå av eldre transkripsjoner og opptak, for eksempel fra Målførearkivet ved ILN, dels i samarbeid med NO2014. De øvrige nordiske dataene vil leveres av våre samarbeidspartnere i ScanDiaSyn.

Nordisk dialektkorpus og database utvikles i samarbeid med våre partnere i det nordiske nettverket ScanDiaSyn og det nordiske senteret for fremragende forskning, NORMS. Korpuset og databasen vil bli allment tilgjengelig for forskning, og vil være et viktig verktøy for språkforskere nasjonalt og internasjonalt.



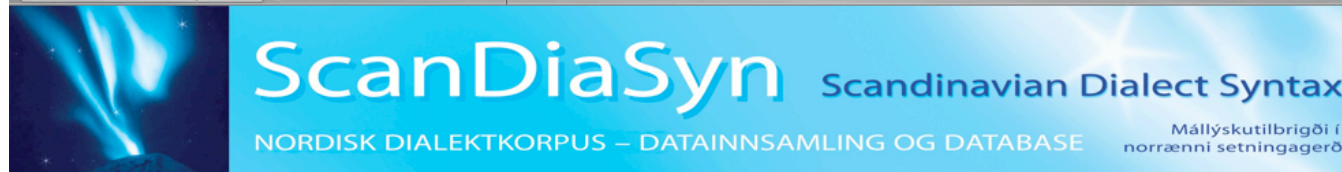
# Partners



# The ScanDiaSyn-project

Two goals:

- Investigate
  - systematically map and study the syntactic variation across the Scandinavian dialect continuum
- Document
  - create a database and a corpus of transcribed and tagged speech material linked with audio and video. Available and easily accessible for a variety of research types, not just syntax, through a user friendly interface on the internet.



Hovedside
Informantbrosjyre
Opptakssteder
Transkripsjon
Ansatte
Fra prosjektsøknaden
Aktiviteter
I pressen
Kontakt
Søk i korpuset
Dokumentkiste for ScanDiaSyn
ScanDiaSyn
NORMS
Tekstlaboratoriet
Institutt for lingvistiske og nordiske studier

## Nordisk dialektkorpus - Scandinavian Dialect Corpus

2007 - 2008: Tekstlaboratoriet utvikler første del av et korpus og en talespråksdatabase for nordiske dialekter. Vi samler også inn materiale til et norsk dialektkorpus fra ca 50 målepunkter (dvs ca 200 informanter) rundt omkring i Norge. Innsamlingen skjer i samarbeid med våre partnere ved UiT og NTNU.

2008 - 2009: Den norske delen av dialektkorpuset utvides med 25 målepunkter.

Det norske dialektkorpuset vil bestå av intervjuer og samtaler som skal transkriberes og kobles sammen med tilsvarende lyd- og videoopptak. Det vil også bli brukt spørreskjemaer som skal kartlegge ulike syntaktiske forhold i informantenes talemål. Korpuset vil dessuten bestå av eldre transkripsjoner og opptak, for eksempel fra Målførearkivet ved ILN, dels i samarbeid med **NO2014**. De øvrige nordiske dataene vil leveres av våre samarbeidspartnere i ScanDiaSyn.

Nordisk dialektkorpus og database utvikles i samarbeid med våre partnere i det nordiske nettverket **ScanDiaSyn** og det nordiske senteret for fremragende forskning, **NORMS**. Korpuset og databasen vil bli allment tilgjengelig for forskning, og vil være et viktig verktøy for språkforskere nasjonalt og internasjonalt. Den norske delen av dialektkorpuset vil kunne utgjøre en del av en framtidig norsk språkbank.

ScanDiaSyn og Nordisk dialektkorpus har fått støtte fra:



[Summary in English](#)



interval:  +  
 criteria»  1 min  criteria»  -  
 start of word  max

Regular expressions:    Randomize   
 Search within:   Skip total frequency

informant   


---

 country    [**>**] [**<**]  
 Norway    


---

 agegroup  sex

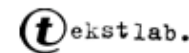
Display:  Search

- word »
- occurrences »
- num »
- degr »
- case »
- sex »
- nlex »
- pers »
- temp »
- def »
- desc »
- pos »
- type »
- phonetic
- (spm)
- adj
- adv
- adv:interj
- adv:sbu
- cbl
- det
- inf-merke
- interj
- konj
- konj:adv
- konj:prep:adv
- konj:sbu
- konj:sbu:adv
- konj:sbu:det
- pause
- pause2
- prep
- prep:sbu
- pron
- pron/det
- sbu
- sbu:adv
- sbu:prep
- subst
- subst:adj
- sänn
- ukjent
- verb
- verb:subst

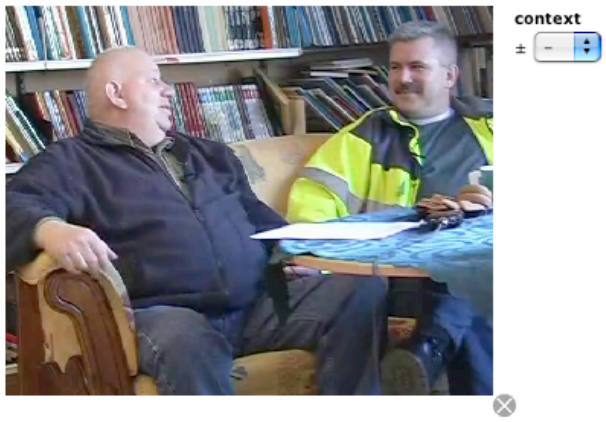
Recording locations  
Transcriptions

[Velg subkorpus](#)

Denne Glossaversjonen er under utvikling.  
 This version of Glossa is under development.



botnhamn\_06 \* ja  
 botnhamn\_03 **og ellers så har jeg ikke vært på Bærhaugan # for det at hvis man er så uheldig man tar ei moltebær på feil # side**  
 botnhamn\_06 så er det krig



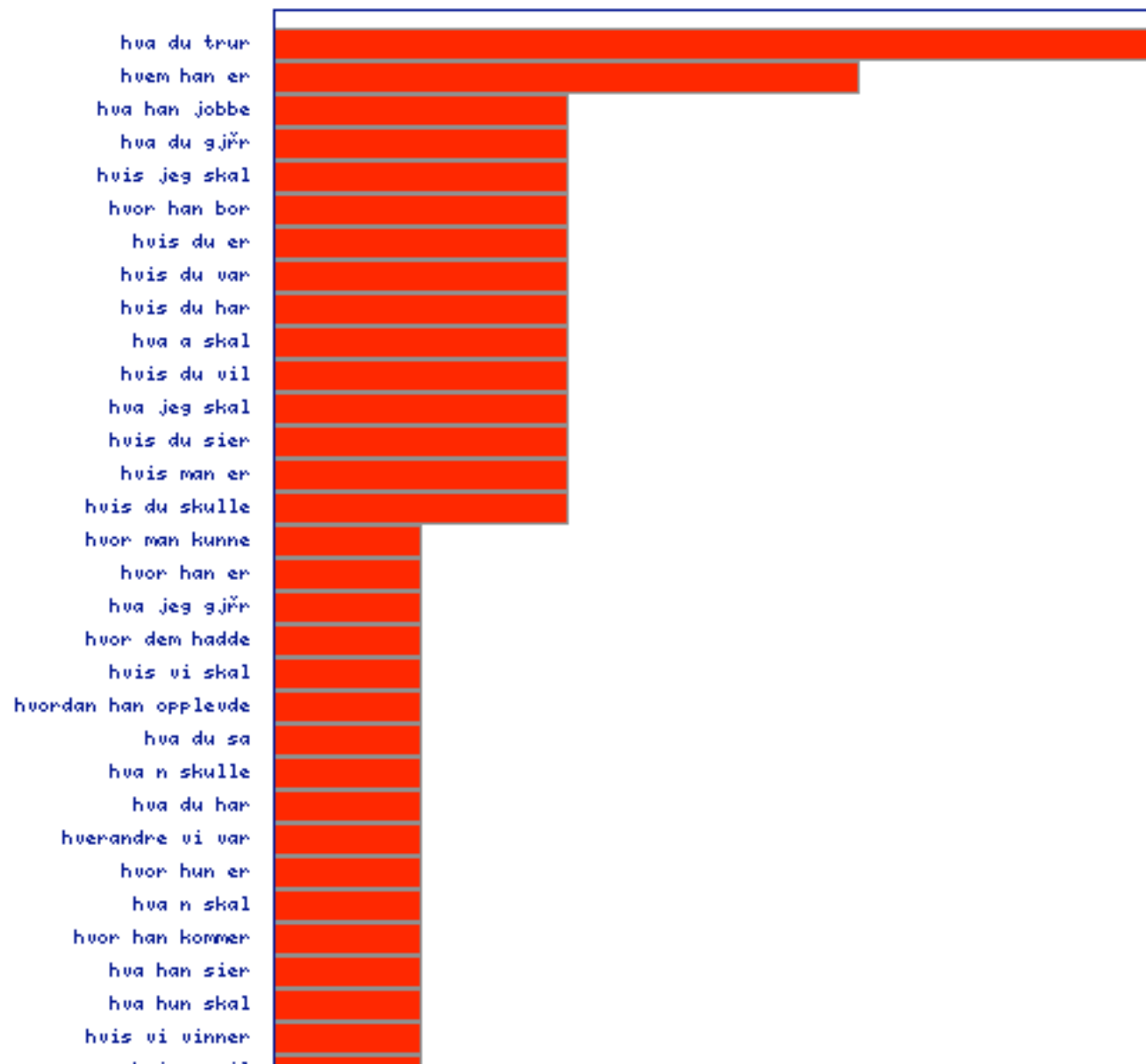
Informants: 22  
 scandiasyn:  
 CWB expression: "(((word="hv.\*" %c)))[(pos="pron")][((pos="verb"))];"  
 Action :   
 : 116  
 Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

- [andoeva\\_ma\\_01](#) **andoeva\_ma\_01** og jeg jeg ropte men jeg han hørte med ikke ## og så og så fortsatte jeg og og og gikk og da kom # kona til eldste b
- [botnhamn\\_06](#) **botnhamn\_06** ja har du hørt noe ko **hva han sier** om det ?
- [botnhamn\\_03](#) **botnhamn\_03** det svinger alle veier **hva du syns** om åpningen på sprinte
- [botnhamn\\_06](#) **botnhamn\_06** nei han hadde ikke og **hva dem er** fire stykker ?
- [botnhamn\\_06](#) **botnhamn\_06** så han er så burde ha **hva dem er** fire stykker ?
- [botnhamn\\_06](#) **botnhamn\_06** \* du **hvordan det gikk** med han Tor Arne e med han Hetland ?
- [botnhamn\\_03](#) **botnhamn\_03** og ellers så har jeg ikke vært på Bærhaugan # for det at **hvis man er** så uheldig man tar ei moltebær på feil # side
- [botnhamn\\_06](#) **botnhamn\_06** (latter) **hva du trur** hun e # hva du trur hun Snuppa # en kveld vi var lagt oss # og det var nå etter det var blitt fr
- [botnhamn\\_06](#) **botnhamn\_06** (latter) hva du trur hun e # **hva du trur** hun Snuppa # en kveld vi var lagt oss # og det var nå etter det var blitt fr
- [botnhamn\\_06](#) **botnhamn\_06** så nå i ettertid veit du **hva jeg gjør** nå ?

lemma: jeg  
 phon: æ  
 pos: pron  
 num: ent  
 type: pers:hum  
 pers: 1.  
 case: nom  
 descr: me  
 nlex: 6



## Lexical Statistics



# OPPTAKSSTEDER

## 50 + 25 målepunkter finansiert hittil

### HEDMARK

Kirkenær  
Stange  
Drevsjø  
Ljørdalen, Trysil  
Aldal

### OPPLAND

Jevnaker  
Skreia  
Kvam  
Gausdal  
Vang  
Ringebu  
Lom

### TELEMARK

Tinn  
Nissedal  
Langesund  
Vinje  
Hjartdal  
Nome

### AGDER

Vennesla  
Valle  
Evje  
Vegårshei  
Kr.sand  
Lyngdal  
Åseral  
Sirdal  
Landvik

### SOGN OG FJORDANE

Gaupne  
Stryn  
Kalvåg  
Jølster  
Hyllestad  
Bremanger

### TRØNDELAG INKL. MØRE

Namdalen  
Inderøya  
Meråker  
Oppdal  
Røros-distriktet  
Gauldal  
Fosen  
Hitra/Frøya  
Trondheim  
Aure  
Todalen  
Selbu

### NORDLAND

Evenes

### TROMS

Medby  
Botnhamn  
Mefjordvær  
Stonglandseidet  
Tromsø  
Lavangen  
Kirkesdalen (Målselv)  
Mannalen  
Karlsøy  
Kvæfjord

# Forts.

- **ROGALAND**

Suldal

Sokndal

Gjesdal

Karmøy

Hjelmeland

Time

+ Stavanger

- **BUSKERUD**

Hallingdal

Syilling, Lier

Rollag, Numedal

Ål, N.Hallingdal

Øvre Eiker

- **FINNMARK**

Ikke klart ennå

## HORDALAND

Fusa

Voss

Eidfjord ,Y.Hardanger

Kvinnherad,I.Hardanger

Bergen

LindØY

Bømlo

## NORDLAND

Øksnes (Vesterålen)

Vestvågøy(Lofoten)

Ballangen (Ofoten)

Steigen

Beiarn

Rana

Hattfjelldal

Alstahaug

Sømna

# Transkripsjon og tagging av ScanDiaSyn-opptakene

- **To transkripsjoner:**

  - Lydnær**

    - Verktøy: Transcriber som automatisk kopler sammen transkripsjonen med lyd og bilde.

  - Ortografisk**

    - Verktøy: halvautomatisk dialektoversetter utviklet ved Tekstlaboratoriet

- **Tagging:** NoTa-taggeren, en statistisk talemålstagger som er trent på materialet fra NoTa-Oslo.

# Transkripsjon

- Talemålsnær transkripsjon som i hovedsak følger Papazian og Helleland: *Norsk talemål*
- Unntak: Vi bruker bare norske tegn, jf TAUS, f.eks:
  - L for tjukk l
  - å** for /o/, **o** for /u/ og **u** for /ū/
  - sj** for /ʃ/, **kj** for /ç/
- Se transkripsjonsveiledningen på hjemmesiden <http://www.tekstlab.uio.no/nota/scandiasyn/>

# Øversettelse

- Alle norske transkripsjoner i Nordisk dialektkorpus blir øversatt til standard bokmål (2005) vha av en halvautomatisk transliterator utviklet av Lars Nygård
- Hvorfor bokmål?
  - Kan benytte NoTas talemålstagger
  - NoTa-Oslo er transkribert til bokmål
  - Kan senere øversettes halvautomatisk til nynorsk vha NyNo

# Eksempel på lydnær og ortografisk transkripsjon

003 + 006

\* 1: ja da # de de va m de de toks kje vare på nåkken ting # æ husske æ hadde ett æ hadde # em ee fannt enn sånn em jærnlæsst de veit du ka de e førr slag de?

2: \* nei

003 + 006

\* 1: jærnlæsst de va sånn dæmm brukkte dæmm laggede jo sko sjøll +[pron=uklart] # å m å denn tænnge æ denn må æ ta vare på førr de e jo enn ee # unik tinn

2: \* ja åkei

\* 1: å æ ta tok hann # å ann pappa heiv denn # å æ hennta ann å jømmte ann å hann klarte å finne ann å heiv ann

2: \* oi

\* 1: så sa æ "koffer hiv du [pron=uklart-] detta her [-pron=uklart] ?" # "vess du veit ko mykkje æ banntes [leende] på denn dær jærnlæsst'n nå vi hållte +[pron=uklart] på # sku lage sko # så har du å ville hevve ann"

botnhamn\_03 ja da # det det var m det det toks ikke vare på noen ting # jeg husker jeg hadde et jeg hadde # em e fant en sånn em jernlest det veit du hva det er for slag det?

botnhamn\_06 \* nei

botnhamn\_03 jernlest det var sånn dem brukte dem lagde jo sko sjøl # og m og den tenkte jeg den må jeg ta vare på for det er jo en e # unik ting

botnhamn\_06 \* ja ok

botnhamn\_03 og jeg ta tok han # og han pappa heiv den # og jeg henta han og gjømte han og han klarte og finne han og heiv han

botnhamn\_06 \* oi

botnhamn\_03 så sa jeg " hvorfor hiver du dette her ? " # " hvis du veit hvor mye jeg bantes (leende) på den der jernlesten når vi holdt på # skulle lage sko # så har du òg ville hivd han "





# Innsamlingsmetoder – dagens innsamling

**Intervju** Samtaler Spørreskjema Oversettelse Andre data

- Opptaksassistent intervjuer
- Varighet 5 til 10 minutter
- Enkle spørsmål om oppvekst og bosted
- Tas opp på video og transkriberes med talemålsnær transkripsjon



# Innsamlingsmetoder – dagens innsamling

Intervju **Samtaler** Spørreskjema Oversettelse Andre data

- To informanter snakker fritt sammen i ca 20 minutter
- Forsøker å skape uformell stemning med lett servering og duk på bordet
- NB! Ikke bringe sensitive personopplysninger inn i samtalen
- Tas opp på video og transkriberes



# Innsamlingsmetoder – dagens innsamling

Intervju Samtaler **Spørreskjema** Oversettelse Andre data

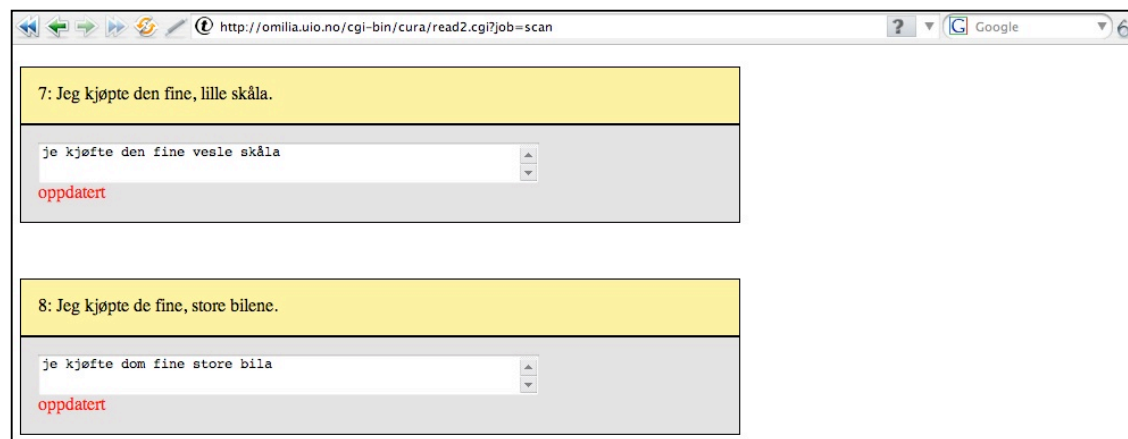
- Informanten bedømmer 130 setningers akseptabilitet
- Graderes på en skala fra 1 (dårligst) til 5 (best)
- Spilles av på lokal dialekt fra medbrakt pc (ev. leses opp)

	A	B	C	D	E	F	G
1	VIKTIG: Kolonne A (unike setningsnumre) må aldri endres. Nye setninger og nye numre kan kun legges til på slutten av skjemaet. Se veiledningen i fanen nederst på arket.						
2							
3	Unike		Opptakssted -->	Opptakssted			
4	nr.		Informant nr. -->	1	2	3	4
269	Ja/nei-spørsmål med om.						
270				Test 127 kun hvis positivt svar på 126.			
271	1245	126	Om du har vært i Tromsø? (bet: Har du vært i Tromsø?)				
272	1246	127	Om du har aldri vært i Tromsø? (bet: Har du aldri vært i Tromsø?)				

# Innsamlingsmetoder – dagens innsamling

Intervju Samtaler Spørreskjema **Oversettelse** Andre data

- Kartlegging av syntaktiske og morfologiske forhold
- Informanten vurderer 55 konstruksjoner, og oversetter de som vurderes som akseptable til lokal dialekt



# Innsamlingsmetoder – dagens innsamling

Intervju Samtaler Spørreskjema Oversettelse **Andre data**

- Inkludere data fra Målførearkivet og fra de andre nordiske landene
  - Ulike mengder
  - Ikke nødvendigvis alle typer av data fra alle steder
  - Dataene tilpasses og presenteres i samme format

# Opptak fra Målførearkivet

- Utfordringer:

Finne samtaler/intervjuer med fri dialog

- Ikke opplesing av ordlister
- Ikke for stort fokus på stedsnavn
- Ikke for mange personopplysninger/sensitive opplysninger
- Ikke teiper som inneholder for mange småopptak
- Ikke teiper der intervjueren snakker mest

Finne opptak uten for mye støy

- Dårlig opptaksutstyr
- Utydelig tale

Finne opptakssteder som er i nærheten av  
ScanDiaSyns målepunkter

# Opptak fra Målførearkivet

- **Hedmark:** Drevsjø ɹ, Ljørdalen, Rendalen, Åsnes ɹ
- **Oppland:** Brandbu ɹ, Nordre land, Skjåk ɹ, Valdres, Vågå
- **Telemark:** Bø, Hjartdal ɹ, Rauland, Seljord, Valle ɹ
- **Agder:** Eiken ɹ, Lista, Søgne, Søgneled, Tonstad, Vegårdshei, Øvre Sirdal ɹ, Åseral ɹ
- **Sogn og Fjordane:** Aurland ɹ, Eid (Nordfjord), Hyen, Gloppen, Årdal
- **Troms:** Andøya ɹ, Bardu ɹ, Holt (Målselv) ɹ, Kirkesdalen (Målselv) ɹ, Lyngen ɹ, Sørreisa ɹ, Torsken ɹ, Tranøy ɹ, Tromøysund ɹ

ɹ transkribert

→ DEMO

## Fra lydfil til websøkeside:

andoeya\_ma\_01

- ja de di hørrte tell i Trommsø # ee lokalbåta

fm

- husker du årstallene for det der for de for de forlisene der?

andoeya\_ma\_01

- næi # æ tørr ikkje å sei de dærran # ee åssjtallan førr # ee denn føssjte Sænnjen då då arbeidde vi på Annes ## dær på moloen
- å hann ee å hann E3 # hann va sånn matrialførrvalltar dær å # hann sa de hann hadde ført de +[pron=uklart] inn i # i sjurnal'n ## datoen då når den føsstte Sænnjen # føliste
- å denn anner Sænnjen ## de # kan æ ikkje sei # akkorat daton # førr då # då då ee hadde vi fått hann M2 hann såmm nu e # såmm æ bor hoss her i i Ossjlo
- å +[pron=uklart] hann va smågut # å hann kåmm ee ner te mæ dær æ arbeidd nere hoss hann (uforståelig) å hållt på me nånn ee vånngnjul å arbeidde dær
- å å vi hørrte denna tutingga a denna båt'n ## så # hann kunne være enn # enn par år gammel # trur æ hann M2
- førr hann kåmm åleina nerråver dær tell mæ ### å ee ## æ kann ikkje s- nu sei sei åssjtalle # akkorat

andoeya\_ma\_01-02-fm\_b\_amg\_lh.tr  
andoeya\_ma\_01-02\_fm\_b\_amg.wav

	fm + andoeya_ma_01	andoe	fm + ...	andoeya_ma_01	fm	a
ver dær tell mæ ### å ee ## æ kann ikkje s- nu sei åssjtalle # akkorat	nei # det er så # husker du noe om av ...	de...	å ja # den ...	ja ja då bægge to # de va snetykke # svært drev	gikk de...	j
	* [pron=me-] ja de de +[pron=uklart] de var ee	..=me]	* ja ja		...karene?	

8:40                      8:45                      8:50                      8:55                      9:00                      9:05

Cursor : 08:58.172



# Fra lydfil til websøkeside:

## Transcriber-fil vist som tekst

```
</Turn>
<Turn speaker="spk2" startTime="437.621" endTime="442.204">
<Sync time="437.621"/>
ja de di hørrte tell i Tromsø # ee lokalbåta
</Turn>
<Turn speaker="spk1" startTime="442.204" endTime="445.758">
<Sync time="442.204"/>
husker du årstallene for det der for de for de forlisene der?
</Turn>
<Turn speaker="spk2" startTime="445.758" endTime="523.950">
<Sync time="445.758"/>
næi # æ tørr ikkje å sei de dærran # ee åssjtallan førr # ee denn føssjte Sænnjen då då arrbedde vi på Annes ## dær på moloen
<Sync time="458.64"/>
å hann ee å hann E3 # hann va sånn matrialførrvalhtar dær å # hann sa de hann hadde ført de
<Event desc="uklart" type="pronounce" extent="previous"/>
inn i # i sjurnal'n ## datoen då nårr den føste Sænnjen # føliste
<Sync time="472.452"/>
å denn anner Sænnjen ## de # kan æ ikkje sei # akkurat daton # førr då # då då ee hadde vi fått hann M2 hann såmm nu e # såmm æ bor hoss her i i Ossjlo
<Sync time="488.811"/>
å
<Event desc="uklart" type="pronounce" extent="previous"/>
hann va smågut # å hann kânn ee ner te mæ dær æ arrbeidd nere hoss hann
<Comment desc="uforståelig"/>
å hållt på me nånn ee vånngnjul å arbeidde dær
<Sync time="498.612"/>
å å vi hørrte denna tutingga a denna båt'n ## så # hann kunne være enn # enn par år gammel # trur æ hann M2
<Sync time="510.896"/>
førr hann kânn åleina nerråver dær tell mæ ## å ee ## æ kann ikkje s- nu sei sei åssjtalle # akkurat
</Turn>
<Turn speaker="spk1 spk2" startTime="523.950" endTime="532.366">
<Sync time="523.950"/>
<Who nb="1"/>
nei # det er så # husker du noe om av omstendighetene ved de forlisene der # for eksempel den første som forliste?
<Who nb="2"/>
*
<Event desc="me" type="pronounce" extent="begin"/>
ja de de
<Event desc="uklart" type="pronounce" extent="previous"/>
de var ee
</Turn>
<Turn speaker="spk2" startTime="532.366" endTime="533.706">
<Sync time="532.366"/>
de var snetykk
<Event desc="uklart" type="pronounce" extent="previous"/>

<Event desc="me" type="pronounce" extent="end"/>

</Turn>
```

# Fra lydfil til websøkeside: Transcriber-fil med originaltranskripsjon og oversettelse

```
<Turn speaker="spk2" startTime="437.621" endTime="442.204">
  <Sync time="437.621"/>
  ja det de hørte til i Tromsø # e lokalbåter
  @ja de di hørte tell i Tromsø # ee lokalbåta
  </Turn>
  <Turn speaker="spk1" startTime="442.204" endTime="445.758">
    <Sync time="442.204"/>
    husker du årstallene for det der for de for de forlisene der?
    @husker du årstallene for det der for de for de forlisene der?
  </Turn>
  <Turn speaker="spk2" startTime="445.758" endTime="523.950">
    <Sync time="445.758"/>
    nei # jeg tør ikke å si det derre # e årstallene for # e den første Senjen da da arbeidet vi på Andenes ## der på moloen
    @nei # æ tørr ikkje å sei de dærran # ee åssjtallan førr # ee denn føssjte Sænnjen då då arrbedde vi på Annes ## dær på moloen
    <Sync time="458.64"/>
    og han e og han E3 # han var sånn materialforvalter der og # han sa det han hadde ført det
    @å hann ee å hann E3 # hann va sånn matrialførrvalltar dær å # hann sa de hann hadde ført de
    <Event desc="uklart" type="pronounce" extent="previous"/>
    inn i # i journalen ## datoen da når den første Senjen # forliste
    @ inn i # i sjurnal'n ## datoen då når den føssste Sænnjen # føliste
    <Sync time="472.452"/>
    og den andre Senjen ## det # kan jeg ikke si # akkurat datoen # for da # da da e hadde vi fått han M2 han som nu er # som jeg bor hos her i i Oslo
    @å denn anner Sænnjen ## de # kan æ ikkje sei # akkurat daton # førr då # då då ee hadde vi fått hann M2 hann såmm nu e # såmm æ bor hoss her i i Ossjlo
    <Sync time="488.811"/>
    og
    @å
    <Event desc="uklart" type="pronounce" extent="previous"/>
    han var smågutt # og han kom e ned til meg der jeg arbeidet nede hos han
    @ hann va smågut # å hann kamm ee ner te mæ dær æ arrbeidd nere hoss hann
    <Comment desc="uforståelig"/>
    og holdt på med noen e vognhjul og arbeidet der
    @ å hållt på me nånn ee vånngnjul å arbeidde dær
    <Sync time="498.612"/>
    og og vi hørte denne tutinga av denne båten ## så # han kunne være en # en par år gammel # trur jeg han M2
    @å å vi hørte denna tutinga a denna båt'n ## så # hann kunne være enn # enn par år gammel # trur æ hann M2
    <Sync time="510.896"/>
    for han kom aleine nedover der til meg ### og e ## jeg kan ikke s- nu sier sier årstallet # akkurat
    @førr hann kamm åleina nerråver dær tell mæ ### å ee ## æ kann ikkje s- nu sei sei åssjtalle # akkurat
  </Turn>
  <Turn speaker="spk1 spk2" startTime="523.950" endTime="532.366">
    <Sync time="523.950"/>
    <Who nb="1"/>
    nei # det er så # husker du noe om av omstendighetene ved de forlisene der # for eksempel den første som forliste?
    @nei # det er så # husker du noe om av omstendighetene ved de forlisene der # for eksempel den første som forliste?
    <Who nb="2"/>
  *

```

# Fra lydfil til websøkeside: Tagget og oversatt fil

```
Senjen_["Senjen" subst_appell_mask_be_ent]
</Turn>
<Turn speaker="spk1 spk2" startTime="433.733" endTime="437.621">
<Sync time="433.733"/>
<Who nb="1"/>
var_["være" verb_pret] det_["det" det_dem_nøyt_ent] Troms_["Troms" subst_appell_mask_ub_ent] var_["være" verb_pret] det_["det" det_dem_n
<Comment desc="uforståelig"/>

<Who nb="2"/>
*["*" pause2] ja_["ja" interj] ja_["ja" interj] ja_["ja" interj]
</Turn>
<Turn speaker="spk2" startTime="437.621" endTime="442.204">
<Sync time="437.621"/>
ja_["ja" interj] det_["det" pron/det] de_["de" pron_fl_pers_3_nom_obj] hørte_["høre" verb_pret] til_["til" prep] i_["i" prep] Tromsø_["T
</Turn>
<Turn speaker="spk1" startTime="442.204" endTime="445.758">
<Sync time="442.204"/>
husker_["huske" verb_pres] du_["du" pron_ent_pers_hum_nom_2] årstallene_["årstall" subst_appell_nøyt_ub_ent] for_["for" prep] det_["det"
</Turn>
<Turn speaker="spk2" startTime="445.758" endTime="523.950">
<Sync time="445.758"/>
nei_["nei" interj] #_["#" pause] jeg_["jeg" pron_ent_pers_hum_nom_1] tør_["tørre" verb_pres] ikke_["ikke" adv] å_["å" inf-merke] si_["si
<Sync time="458.64"/>
og_["og" konj] han_["han" pron_mask_ent_pers_hum_3_nom] e_["e" interj_fyll] og_["og" konj] han_["han" pron_mask_ent_pers_hum_3_nom] E3_["
<Event desc="uklart" type="pronounce" extent="previous"/>
inn_["inn" prep] i_["i" prep] #_["#" pause] i_["i" prep] journalen_["journal" subst_appell_mask_be_ent] ##_["##" pause] datoen_["dato" s
<Sync time="472.452"/>
og_["og" konj] den_["den" det_dem_mask_ent] andre_["annen" adj_be_ent_pos] Senjen_["Senjen" subst_appell_mask_be_ent] ##_["##" pause] de
<Sync time="488.811"/>
og_["og" konj]
<Event desc="uklart" type="pronounce" extent="previous"/>
han_["han" pron_mask_ent_pers_hum_3_nom] var_["være" verb_pret] smågutt_["smågutt" subst_appell_mask_ub_ent] #_["#" pause] og_["og" konj]
<Comment desc="uforståelig"/>
og_["og" konj] holdt_["holde" verb_perf-part] på_["på" prep] med_["med" prep] noen_["noen" det_mask_ent_kvant] e_["e" interj_fyll] vognh
<Sync time="498.612"/>
og_["og" konj] og_["og" konj] vi_["vi" pron_fl_pers_hum_nom_1] hørte_["høre" verb_pret] denne_["denne" det_dem_fem_ent] tutinga_["tuting
<Sync time="510.896"/>
for_["for" konj] han_["han" pron_mask_ent_pers_hum_3_nom] kom_["komme" verb_pret] aleine_["aleine" adv] nedover_["nedover" prep] der_["d
</Turn>
<Turn speaker="spk1 spk2" startTime="523.950" endTime="532.366">
<Sync time="523.950"/>
<Who nb="1"/>
nei_["nei" interj] #_["#" pause] det_["det" pron/det] er_["være" verb_pres] så_["så" konj/sbu/adv] #_["#" pause] husker_["huske" verb_pr
<Who nb="2"/>
*["*" pause2]
<Event desc="me" type="pronounce" extent="begin"/>
ja_["ja" interj] det_["det" pron_nøyt_ent_pers_3] det_["det" pron_nøyt_ent_pers_3]
<Event desc="uklart" type="pronounce" extent="previous"/>
det_["det" pron/det] var_["være" verb_pret] på_["på" prep] med_["med" prep] noen_["noen" det_mask_ent_kvant] e_["e" interj_fyll] vognh
```

Fra lydfil til websøkeside:

## Tagget fil og originaltranskripsjon satt sammen

- Resultat: fil som er satt sammen av originaltranskripsjonen, den oversatte transkripsjonen samt den morfologiske taggen:

og\_["å" "og" konj] han\_["hann" "han" pron\_mask\_ent\_pers\_hum\_3\_nom] e\_["ee" "e"  
interj\_fyll] og\_["å" "og" konj] han\_["hann" "han"  
pron\_mask\_ent\_pers\_hum\_3\_nom] E3\_["E3" "E3" subst\_prop] #\_["#" "#" pause]  
han\_["hann" "han" pron\_mask\_ent\_pers\_hum\_3\_akk] var\_["va" "være" verb\_pret]  
sånn\_["sånn" "sånn" sånn] matrialforvalter\_[" matrialførrvalltar " matrialforvalter"  
subst\_appell\_mask\_ub\_fl] der\_["dær" "der" prep]

Fra lydfil til websøkeside:  
**Innlegging av data**

- Transkripsjoner, tagging og informantopplysninger legges inn i informantdatabase og cqp

# 121 svenske informanter - fra Swedia 2000, ScanDiaSyn, NORMS

aasen_35
ankarsrum_om1
ankarsrum_om3
anundsjo_ow3
arjeplog_om1
arsunda_om1
asby_ym1
bara_om1
bengtsfors_ym2
blyberg_165
boda_ow3
bredsatra_om2
brunnsberg_57
evertsberg_188
faro_ym1
floby_om1
fole_om1
frillesas_ow2
indal_ow2
jamshog_ow1
klitten_144
kola_om1
leksand_ow4
loderup_ow2
nederlulea_ow1
orust_ow4
ossjo_ow2
overkalix_om1
segerstad_om1
skinnskatteberg_ym1
skolan_78
sorunda_ym1
sproge_om3
st_anna_ow3

# Informanter fra Dalarna

tid	sex	agegroup	country	region	area	place
aasen_35	F	B	Sweden	Svealand	Dalarna	Åsen
aasen_48	F	B	Sweden	Svealand	Dalarna	Åsen
blyberg_165	F	B	Sweden	Svealand	Dalarna	Blyberg
blyberg_168	M	B	Sweden	Svealand	Dalarna	Blyberg
brunnsberg_55	M	B	Sweden	Svealand	Dalarna	brunnsberg
brunnsberg_57	M	B	Sweden	Svealand	Dalarna	Brunnsberg
evertsberg_188	F	B	Sweden	Svealand	Dalarna	Evertsberg
klitten_141	M	B	Sweden	Svealand	Dalarna	Klitten
klitten_144	F	B	Sweden	Svealand	Dalarna	Klitten
leksand_om1	M	B	Sweden	Svealand	Dalarna	Leksand
leksand_om3	M	B	Sweden	Svealand	Dalarna	Leksand
leksand_ow3	F	B	Sweden	Svealand	Dalarna	Leksand
leksand_ow4	F	B	Sweden	Svealand	Dalarna	Leksand
skolan_69	M	A	Sweden	Svealand	Dalarna	Brunnsberg
skolan_77	F	A	Sweden	Svealand	Dalarna	Kåtilla
skolan_78	F	A	Sweden	Svealand	Dalarna	Älvdalen
skolan_79	M	A	Sweden	Svealand	Dalarna	Brunnsberg

# Noen verktøy og ressurser

- Integrert søkesystem for web: *Glossa* (UiO)
- Transliteratør: oversettelse fra dialekt til norm (UiO)
- Statistisk talemålstagger fra NoTa (UiO)
- Nettskjema: oversettelse til dialekt for informanter (UiO)
- Transkripsjonsveiledning for ScanDiaSyn (UiO)
- Transkripsjonsveiledning for NoTa (detaljer om turtaking o.a.) (UiO)
- Dokumentkiste: utveksling av data mellom forskergruppene (Eksternt + UiO)
- Blogg: rapport fra feltarbeid, utveksling av erfaringer (UiT)
- *Transcriber* til transkripsjon (Eksternt)



# Websider

- Glossa:  
<http://www.hf.uio.no/tekstlab/glossa.html>
- Skandinavisk dialektkorpus:  
<http://www.tekstlab.uio.no/nota/scandiasyn/>
- Transkripsjonsveiledning - fonetisk:  
<http://www.tekstlab.uio.no/nota/scandiasyn/transkripsjonsveiledning.html>
- Transkripsjonsveiledning ortografisk - med turer:  
<http://www.tekstlab.uio.no/nota/oslo/index.html> (se fane for transkripsjon)
- Norske talespråkskorpus v/ UiO:  
<http://www.tekstlab.uio.no/nota/>
- NoTa-Oslo m/transkripsjonsveiledning:  
<http://www.tekstlab.uio.no/nota/oslo/index.html>
- Dokumentkiste:  
<http://www.tekstlab.uio.no:8080/scandiasyn>
- Blog:  
<http://uit.no/scandiasyn/blog/>
- Transcriber:  
<http://trans.sourceforge.net/en/presentation.php>