

Helge Sandøy

Talebanken ved LLE, UiB.

Ei orientering

15.3.2014

<http://www.folk.uib.no/hnohs/Talebanken.html>

Formål og premissar

Talebanken er eit digitalt arkiv (korpus) for transkriberte lydfiler som skal brukast i talemålsforskning. Her blir lyd og transkripsjon kopla saman. Arkivet er etablert med tanke på forskningsformål, og derfor er tilgangen avgrensa til forskarar med løyve. Denne avgrensinga gjer at ein kan sleppe å sortere vekk så mykje frå samtaleopptaka som vi elles burde gjort viss vi skulle la bruken vere friare. Dette gjeld først og fremst gamle opptak, for ved nye opptak gjeld ei generell retningslinje om at ein skal unngå tema som kan skape problem med tanke på personvernet. Talebanken følgjer retningslinjene for personvern og er bygd opp i samråd med Norsk samfunnsvitskapleg datateneste (NSD).

Talebanken er utvikla teknisk med tanke på å effektivisere analysar og å gjere analysane sikrare. Materialet kjem inn i Talebanken ved at det går inn i konkrete sosiolingvistiske talemålsstudiar. Det eit mål å ta vare på alt materialet inklusiv kodingar for at det skal kunne gjenbrukast i andre og seinare prosjekt.

For å effektivisere og betre analysane er grunntanken at transkripsjonen skal vere i ein standardortografi, og vi har valt nynorsk rettskriving. Med å følge ein slik standard kan ein vere sikrare på å finne att og sortere relevant materiale når ein skal søkje, kode og analysere elektronisk. Den faktiske talen, som sjølvsgatt er viktig i talemålsprosjekt, blir derimot fokusert ved kodinga i variablar. Då tilfredsstiller ein den premissen for nøyaktigheit som seier at noko er nøyaktig berre i forhold til ein definisjon eller ei problemstilling. (Ut frå det siste kan lydskrift aldri bli nøyaktig, aller minst med tanke på at materialet skal kunne brukast til analysar der verken problemstillingar eller variablar enno ikkje er fastlagt.)

Historikk

Utviklinga av Talebanken gjekk inn som del av *Dialektendringsprosessar*, eit NFR-prosjekt som tok til i 2008 med Helge Sandøy som prosjektleiar. Det tekniske utviklingsarbeidet er gjennomført av Paul Meurer og Øystein Reigem ved UniDigital i tett samarbeid med Helge Sandøy og medarbeidarar ved *Dialektendringsprosessar* i tida 2008-2011. Programmeringa av Talebanken byggjer på det grunnleggjande programmeringsverktøyet Corpuscle, og Talebanken har liknande struktur som andre korpus utvikla siste åra av Paul Meurer, slik at programutvikling i eine korpuset kan lett overførast til eit anna korpus (jf. Paul Meurer: «Corpuscle – A new corpus management platform for annotated corpora» i Gisle Andersen: *Exploring Newspaper Language*, 31–49, Amsterdam: Bejnamins).

Det er ønskjeleg å få mest mogleg forskingsmateriale inn i Talebanken – både til fordel for enkeltprosjekta og for den samla oppbygde «felleskapitalen» som banken kan vere

for framtidig forskning. Derfor legg både *Dialektendringsprosessar* og *Industristadprosjektet* inn sine data der, og dessutan utnyttar også andre mindre talemålsprosjekt ved LLE, UiB, *Talebanken*.

Kort om arbeidsmåten

Når materialet er innlagt i *Talebanken*, kan ein søkje i det frå ein meny der ein gjer ymse val for kva type trefford ein ønskjer seg ifrå kva materiale:

Her kan du lage et søk ved å bruke menyer.

Velg et **delkorpus**:

subdialekt = "Hå, Ogna"
interviewer-p = "Nei"
legg til:
attributt: - <input type="text"/>

Velg **posisjonelle betingelser**. | Ignorer strukturelle posisjoner

<input type="checkbox"/> <i>target</i>
word = <input type="text" value="eg"/> <input type="checkbox"/> %c
gjentagelse: <input type="text" value="1"/> <input type="text"/>
legg til:
attributt: - <input type="text"/>
struct: - <input type="text"/>

Her søker ein etter ordet *eg* i materialet frå Ogna i Hå kommune. Når ein har trykt *Søk*, får ein opp alle treffa (belegga, førekomstane) i ein kontekst på denne måten i ein kwic-konkordans:

Enkelt søk | Avansert søk

Søk: "eg" :: dialect-sub = "Hå, Oгна" & interviewer-p = "Nei" | Søk | Lagre søket som | Hent lagret: - | Slett søket

Done. Real time: 0.1731 sec. (0.037 CPU sec.)

00:00 00:00

Treff 1 - 30 av 770 | siste 15 treff | Gå til: | Last ned (Excel-modus) | Type: kwic | Vis annotasjon | Vis linjefilter | Columns ... | Structures ... | Linjer per side

Choose variable: | Position: | Shortcuts:

count	cpos	treff
1	1631620	r til dagen spelar du, tru? To timar kanskje? Det er ikkje så mykje. Nei. Du då? Ikkje så mykje, eg pleier vera heime ein del. Ja. Elles går på rulleski. Å. Ja. Har du mange plassar, gode plassar å gå her? Ikkje akk der er. Ja. Er der det? Det var mange. Ja. Og det er ein til to? Som... Ja. I skal gå på Øksnevad. Mm. Far din, min arbeider der. Er ikkje det, er ikkje ikkje gjort så mykje meir. Eg har vore hjå slektingar og overnatta. Ålgård skal bli bonde. Ja. Vera på gard? Ja. Lyst å overta? Ja. Så då, så du kjem kjem på. Nei. Nei. Ingenting negativt eg kjem på, så, Mm. Ja. Skal me sjå det var det var snø der oppe. Då gjekk me på ski, og så var der renn. Å. var med på, der inne. Matningsdal. Og så er... litt heime. Og så, Bygd for kjem på, unntake, det var... Det var når eg var liten, så vart eg redd, for me var hjå nokre slektingar, og så heldt me oppi ei sånn, plastikkting. Og så var eg redd for at eg skulle drukna, så eg stod og så ropte, hjelp, hjelp, hjelp. M

I denne konkordansvisinga kan ein så kode (annotere) alle tilfella av treffordet. Det gjer ein med å lytte på ordet i konteksten for så å bestemme kva variant forma av *eg* tilhøyrer og så sette inn med hurtigtast den aktuelle koden:

00:00 00:00

30 av 4097 | neste 30 treff | Gå til: | Last ned (Excel-modus) | Type: kwic | Vis annotasjon | Vis linjefilter | Columns ... | Structures ... | Linjer per side: |

variable: | Position: | Shortcuts:

cpos	treff	annotation
1440033	ig var med då, for, kor mange år sidan blir det? Og synest det var artig å bli spurt på nytt, så	eg tenkte det var greit å følge det opp. Hm. Eg heiter @namn. Fødd og
1440615	a. Jo eldre ein blir, så, så legg ein og kanskje meir og meir merke til det. Meir og meir. Ja, men	eg L02:e-sv trur det, i motsetnad til Oslo så trur eg det er, ikkje så lett å på ein m
1440625	nskje meir og meir merke til det. Meir og meir. Ja, men eg trur det, i motsetnad til Oslo så trur	eg L02:e-sv det er, ikkje så lett å på ein måte kjenne dei, få dei linkane. Nei. Det e
1440851	altså det var på ein måte, dette var veldig fredeleg. Mm. I motsetnad til kva det er i dag. Kor	eg L02:e-sv opplever at det er meir skummelt å vere ute ein laurdagskveld. I tol-
1440874	ver at det er meir skummelt å vere ute ein laurdagskveld. I tol- eitt-tida i sentrum, det gjer	eg L02:eg-sv helst ikkje. Hm. Når vi voks opp, så var det, stort sett trygt og fredel
1440993	ikka, Stokka var jo ein sånn ein... Fredeleg greie. Trur ikkje det har skjedd... Nei, å ja, det er	eg L02:e-sv heilt einig i. Var ikkje ein gjeng. Nei da. Folk var jo altså, sånn. Mødre
1441301	gdommen sjøl... Eg veit jo at vi ikkje skal gå der, så det er jo heilt trygt for meg. Hm. Ja. Så	eg L02:eg-sv trur dei føler det er trygt. Du blir liksom kjent med ein, med rammene
1441357	, der du, der du veks opp. Sånn at du er kjent og du veit smothola, kor du skal stikke av. Så	eg L02:eg-sv trur dei fleste synest det er trygt. Og vi syntest det var trygt, den gr
1441381	e av. Så eg trur dei fleste synest det er trygt. Og vi syntest det var trygt, den gongen. Men	eg L02:eg-sv likar jo ikkje den dagen @namn skal begynne å gå på byen. Nei. Nei. S
1441477	ita-ruta.] Ja. Sant? Det er ikkje gitt du får. Ikkje gitt du får loypene til... Men for henne så er	eg L02:eg-sv sikker på ho synest det er kjempetrygt. Ja. Heilt greit, går saman me
1441559	lag enn det det var når de voks opp? Nokre forskjellar, store eller små? Sosiokulturelt tenkjer	eg L02:eg-sv at den gongen vi voks opp, så var dette eit nybyggjarområde. Det va
1441735	fu får eit anna sjikt av befolkninga enn det som det blei bygd til den gongen. Mm. Ja. Tenkjer	eg L02:eg-sv er ein forandring. Same... Pluss at bondegardane her i all hovudsak ei
1441917	t var eit godt og... Så, de bur på Stokka begge to no òg? Eg bur på Stokka. Ja, så godt som,	eg L02:eg-sv bur på Kampen, eller tilstoytande til, mm. Ja, ok. Byhaugen, ja. Kva, i
1442149	anger. Han høyrtes sånn ut, sånn reint dialektisk. Jo jo, men så hadde han no anar i Lund, så	eg L02:e-sv visste ikkje kva tid han kom hit. Nei då, eg er tredjegenarasjons
1442252	ag for eksempel, har de eit, var det eit sterkt inntrykk på eit eller anna vis, eller? Eg trur ikkje	eg L02:eg-sv hugsar min første skuledag, bortsett frå det eg er trygg på at, at,
1442261	trykk på eit eller anna vis, eller? Eg trur ikkje eg hugsar min første skuledag, bortsett frå det	eg L02:eg-sv er trygg på at, at, at mor må ha vore med meg, og at eg sikkert mått
1442283	nin første skuledag, bortsett frå det eg er trygg på at, at, at mor må ha vore med meg, og at	eg L02:eg-sv sikkert måtte klippe negler og vaske dei grundig, at det skulle vere pe
1442310	klippe negler og vaske dei grundig, at det skulle vere pent og skikkeleg. Ein blå ransel, meiner	eg L02:eg-sv, eg hugsar han ikkje. Eg hugsar frøken @namn som ein urgammal gre
1442312	eg negler og vaske dei grundig, at det skulle vere pent og skikkeleg. Ein blå ransel, meiner	eg L02:eg-sv hugsar han ikkje. Eg hugsar frøken @namn som ein urgammal gre
1442418	it femten års, eller tjuve års-jubileum, då såg ho likedan ut. Mm. Nei, ho var staseleg. Ja. Men	eg L02:eg-sv lurer på om ikkje søndagsskulen sluttar i dei tider no. Eg gjekk på son
1442441	Men eg lurer på om ikkje søndagsskulen sluttar i dei tider no. Eg gjekk på søndagsskulen. Ja,	eg L02:nøyt gjorde ikkje. Nei, ok. Det er difor det har, både, det er der danninga, i
1442507	... Ja. Det forklarar ein heil masse. For du har lurt litt på det? Kva som gjorde at... Ja. Ja, nei,	eg L02:eg-st har rett og slett meir utdanning enn deg. Ja, ja, du har det. Det var n
1442538	eg har rett og slett meir utdanning enn deg. Ja, ja, du har det. Det var mange søndagar. Ja,	eg L02:eg-sv har nokre stjerner og nokre hesteko. Det var, ja, det er litt deilig me
1442600	dig om, om at dei har ein heil samling med Jesus-vers og, og, ja. Ja, det er jo ein høydare. Og	eg L02:nøyt gjekk på skule på laurdag. Mm. Gjorde du? Ja, eg lurer på om ikkje eg
1442620	ers og, og, ja. Ja, det er jo ein høydare. Og eg gjekk på skule på laurdag. Mm. Gjorde du? Ja,	eg L02:e-sv lurer på om ikkje eg gjorde det ei stund. Så vidt, ok. Ja. Men eg bar ik

Når alle aktuelle treffa i materialet er koda, ønskjer ein å lage statistikk over variasjonen i denne språklege variabelen og krysse dei tala mot ymse bakgrunnsopplysningar ein har lagt inn om informantane. Det gjer ein med funksjonen *Distribusjon*, der ein kan f.eks. legge inn at ein ønskjer å sjå korleis variantane av pronomenet *eg* er spreidd på aldersgrupper og kjønn:

Enkelt søk | Avansert søk

Søk: "eg" :: dialect = "Stavanger" & interviewer-p = "Nei" | Lagre søket som | Hent lagret: |

Done. Real time: 0.1911 sec. (0.069 CPU sec.). Match size: 4097

Fractions sum up to 1.0 in each row. Fractions in blue are unweighted means of group fractions. Fractions in green are distributions of total numbers.

av Δ: filter: relative to Δ: filter: grupper etter Δ:

range30+2010 og Δ: filter: type: | counts only | include structures

Side 1/1 av 1x1. | Last ned

(sum) - L02:e-
sv L02:eg-
sv L02:eg-
st L02:ei-
sv L02:je-
sv L02:jei-
sv L02:je-
st L02:e-
st L02:ei-
st

(sum)	3819 (100,0)	498 (13,0)	433 (11,3)	2174 (56,9)	361 (9,5)	70 (1,8)	245 (6,4)	18 (0,5)	14 (0,4)	5 (0,1)	1 (0,0)
-	20 (100,0)	20 (100,0)									
Kvinne	20 (100,0)	20 (100,0)									
1890-1919	382 (100,0)	58 (15,0)	21 (5,8)	131 (27,1)	4 (1,0)		141 (44,0)	15 (3,4)	11 (3,6)	1 (0,2)	
Kvinne	134 (100,0)	19 (14,2)	9 (6,7)	4 (3,0)	1 (0,7)		91 (67,9)	2 (1,5)	8 (6,0)		
Mann	248 (100,0)	39 (15,7)	12 (4,8)	127 (51,2)	3 (1,2)		50 (20,2)	13 (5,2)	3 (1,2)	1 (0,4)	
1920-1949	1129 (100,0)	193 (16,8)	106 (9,3)	593 (53,0)	124 (11,1)	1 (0,1)	104 (8,9)	3 (0,3)	3 (0,3)	2 (0,2)	
Kvinne	584 (100,0)	146 (25,0)	62 (10,6)	224 (38,4)	40 (6,8)	1 (0,2)	104 (17,8)	3 (0,5)	3 (0,5)	1 (0,2)	
Mann	545 (100,0)	47 (8,6)	44 (8,1)	369 (67,7)	84 (15,4)					1 (0,2)	
1950-1979	1520 (100,0)	168 (10,4)	167 (11,6)	971 (63,0)	189 (13,0)	23 (2,0)				2 (0,1)	
Kvinne	972 (100,0)	124 (12,8)	92 (9,5)	644 (66,3)	107 (11,0)	3 (0,3)				2 (0,2)	
Mann	548 (100,0)	44 (8,0)	75 (13,7)	327 (59,7)	82 (15,0)	20 (3,6)					
1980-2009	768 (100,0)	59 (7,8)	139 (19,2)	479 (61,0)	44 (5,7)	46 (6,1)					1 (0,2)
Kvinne	447 (100,0)	31 (6,9)	56 (12,5)	309 (69,1)	27 (6,0)	24 (5,4)					
Mann	321 (100,0)	28 (8,7)	83 (25,9)	170 (53,0)	17 (5,3)	22 (6,9)					1 (0,3)

Denne tabellen med både talet på belegg og prosentutrekning per variant på kvar variabel i gruppa etter årsklasse og kjønn ein bruke til vidare analysearbeid. Tabellen kan lastast ned og bearbeidast vidare f.eks. i programmet Excel for å lage grafiske uttrykk for resultatet.

I ein del arbeid har ein òg nytte av å studere ordforrådet og ordfrekvensar i eit materiale. Det får ein ut på denne måten for f.eks. materialet frå Ognå i Hå:

Enkelt søk | Avansert søk

Søk: ".*" :: dialect-sub = "Hå, Ogna" & interviewer-p = "Nei" Søk | Lagre søket som | He

Done. Real time: 14.8821 sec. (15.3597 CPU sec.)

Antall treff: 13435, unike ord/fraser: 493. Attributt: | ignorer storskriving | sorter: | Last ned include c

Side 1 av 2. Previous Next

3260 (24,17%) .	22 (0,16%) Eller	6 (0,04%) delar	4 (0,03%) denne	2 (0,01%) "Yngres"	2 (0,01%) bruka
2892 (21,44%) ,	20 (0,15%) alltid	6 (0,04%) delte	4 (0,03%) di	2 (0,01%) /ka/	2 (0,01%) bruker
1213 (8,99%) det	20 (0,15%) altså	6 (0,04%) dette	4 (0,03%) din	2 (0,01%) /mor/	2 (0,01%) brukt
770 (5,71%) eg	20 (0,15%) dag	6 (0,04%) dialektar	4 (0,03%) drama	2 (0,01%) /vi/	2 (0,01%) by
468 (3,47%) dei	19 (0,14%) aldri	5 (0,04%) Alle	4 (0,03%) eine	2 (0,01%) Aftenblad	2 (0,01%) byassar
410 (3,04%) der	18 (0,13%) Brusand	5 (0,04%) Av	4 (0,03%) einen	2 (0,01%) Aftenbladet	2 (0,01%) byggjer
354 (2,62%) Det	16 (0,12%) At	5 (0,04%) Bergen	3 (0,02%) /stivla/	2 (0,01%) Alltid	2 (0,01%) bøker
285 (2,11%) du	15 (0,11%) alt	5 (0,04%) Berre	3 (0,02%) Altså	2 (0,01%) Andre	2 (0,01%) campingvogn
240 (1,78%) Eg	14 (0,10%) :	5 (0,04%) Danmark	3 (0,02%) Bella	2 (0,01%) Anten	2 (0,01%) da
208 (1,54%) ?	14 (0,10%) begynner	5 (0,04%) Eigentleg	3 (0,02%) Dalane	2 (0,01%) Austlandsk	2 (0,01%) dataingenior
192 (1,42%) då	13 (0,10%) annan	5 (0,04%) arbeida	3 (0,02%) Ei	2 (0,01%) Bibbok	2 (0,01%) delt
186 (1,38%) at	11 (0,08%) Bjerkreim	5 (0,04%) bedehuset	3 (0,02%) Eldre	2 (0,01%) Boksekasting	2 (0,01%) derfor
182 (1,35%) ein	11 (0,08%) annleis	5 (0,04%) begynna	3 (0,02%) alder	2 (0,01%) Bonden	2 (0,01%) derifrå
162 (1,20%) au	11 (0,08%) begge	5 (0,04%) bli	3 (0,02%) attande	2 (0,01%) Bru	2 (0,01%) disko
120 (0,89%) alle	11 (0,08%) budde	5 (0,04%) boka	3 (0,02%) august	2 (0,01%) Bygde	2 (0,01%) diskutert
111 (0,82%) av	11 (0,08%) både	5 (0,04%) byen	3 (0,02%) bad	2 (0,01%) Eigersundsk	2 (0,01%) dokker
104 (0,77%) ...	11 (0,08%) deg	5 (0,04%) desse	3 (0,02%) bensinstasjonen	2 (0,01%) aktivitetar	2 (0,01%) drukna
102 (0,76%) eller	11 (0,08%) eitt	5 (0,04%) eigen	3 (0,02%) bokstavar	2 (0,01%) aktivitetsskild	2 (0,01%) du/
85 (0,63%) berre	10 (0,07%) Ein	4 (0,03%) /boen/	3 (0,02%) brann	2 (0,01%) all	2 (0,01%) døra
75 (0,56%) den	10 (0,07%) barnehage	4 (0,03%) Alt	3 (0,02%) brei	2 (0,01%) altfor	2 (0,01%) døydd
73 (0,54%) Dei	10 (0,07%) eksempel	4 (0,03%) Austlandet	3 (0,02%) breiare	2 (0,01%) an	2 (0,01%) døyr
70 (0,52%) @namn	9 (0,07%) barnehagen	4 (0,03%) Elles	3 (0,02%) brus	2 (0,01%) aner	1 (0,01%) ""fjell
65 (0,48%) "	9 (0,07%) begynte	4 (0,03%) allikevel	3 (0,02%) bussen	2 (0,01%) apokope	1 (0,01%) "Eple"
65 (0,48%) Der	9 (0,07%) eldre	4 (0,03%) anten	3 (0,02%) butikkar	2 (0,01%) april	1 (0,01%) "Har"
54 (0,40%) andre	8 (0,06%) /åsse/	4 (0,03%) arbeid	3 (0,02%) butt	2 (0,01%) arbeidsveka	1 (0,01%) "Sauda"
49 (0,36%) del	8 (0,06%) arbeider	4 (0,03%) arbeidd	3 (0,02%) byggja	2 (0,01%) att	1 (0,01%) "Sokndal"
47 (0,35%) ei	8 (0,06%) bergensk	4 (0,03%) arbeidde	3 (0,02%) bygt	2 (0,01%) austlending	1 (0,01%) "austlending"
44 (0,33%) blir	8 (0,06%) blei	4 (0,03%) atten	3 (0,02%) dagar	2 (0,01%) bak	1 (0,01%) "bonden"
38 (0,28%) Egersund	8 (0,06%) byass	4 (0,03%) barneskulen	3 (0,02%) dansk	2 (0,01%) bebyggelse	1 (0,01%) "brennmanet"
36 (0,27%) eigentleg	8 (0,06%) dialekten	4 (0,03%) bedehus	3 (0,02%) de	2 (0,01%) begynnelsen	1 (0,01%) "eg"
36 (0,27%) eit	8 (0,06%) einannan	4 (0,03%) bok=	3 (0,02%) dela	2 (0,01%) begynt	1 (0,01%) "eh"
35 (0,26%) !	7 (0,05%) Den	4 (0,03%) bort	3 (0,02%) dett	2 (0,01%) behov	1 (0,01%) "genser"
31 (0,23%) Då	7 (0,05%) Eigersund	4 (0,03%) bra	3 (0,02%) dialekta	2 (0,01%) besøk	1 (0,01%) "kyrme"
31 (0,23%) akkurat	7 (0,05%) austlandsk	4 (0,03%) bu	3 (0,02%) dumt	2 (0,01%) betydde	1 (0,01%) "mor"
30 (0,22%) Bryne	7 (0,05%) blanding	4 (0,03%) budd	3 (0,02%) eigersundsk	2 (0,01%) biten	1 (0,01%) "regnkled"
30 (0,22%) Du	7 (0,05%) bygd	4 (0,03%) buss	3 (0,02%) eig	2 (0,01%) blanda	1 (0,01%) "regntøy"
30 (0,22%) bur	7 (0,05%) einaste	4 (0,03%) byas	3 (0,02%) eige	2 (0,01%) boksekasting	1 (0,01%) "svie"
29 (0,22%) dialekt	6 (0,04%) betre	4 (0,03%) byqda	3 (0,02%) eiqentlig	2 (0,01%) boksen	1 (0,01%) "å"

På grunnlag av heile materialet i Talebanken i 2012 blei det laga ei frekvensordbok over talt materiale. Den er grunnlaget for ein del berekningsarbeid i sosiolingvistiske drøftingar.

Meir rettleing om arbeidsmåtar i Talebanken finn ein under ein fane Dokumentasjon og i ei eiga enkel rettleing på arbeidsområdet åt medarbeidarane.

Materialomfanget

Per 15. mars 2014 omfattar Talebanken dette materialet:

<i>Kommune</i>	<i>Timar opptak</i>	<i>Tal på informantar</i>	<i>Eldste og yngste informant fødd</i>	<i>Talet på ord</i>
<i>Hå</i>				
<i>Stavanger</i>				
<i>Sauda</i>				
<i>Øygarden</i>				
<i>Bergen</i>				
<i>Odda & Tyssedal</i>				

<i>Årdal</i>				
<i>Vanylven</i>				
<i>Midsund</i>				
<i>Molde</i>				
<i>Andre kommunar utafor dei aktuelle prosjekt-områda</i>				
I alt				

Prosjekt og publikasjonar med grunnlag i Talebanken