

Store tal frå trondheimsk tale¹

Metodologisk drøfting av ei nettbasert dialektgranskning

Av Stian Hårstad

A linguist cannot help but be impressed by the Internet. It is an extraordinarily diverse medium, holding a mirror up to many sides of our linguistic nature.

(David Crystal 2005)

1 Innleiing

Midt i 1990-åra, medan Internett enno var eit temmeleg ukjent fenomen for allmenta, kom marknadsføringsforskaren Pete Comley (1996:335) med følgjande spådom: "The Internet will become an established method of data collection of the future". Godt og vel eit tiår seinare kan vi slå fast at Comley fekk rett. Bruken av Internett i vitenskapleg samanheng har auka kolossalt (jf. t.d. Pötschke & Simonson 2001), og *nettundersøkingismetodologi* har etter kvart vakse fram som eit eige fagfelt (jf. Batinic 2001).²

Også lingvistar har på ulike vis gjort bruk av Internett for å skaffe til veges forskingsdata (jf. Atteslander 2005:1072ff.). Dette gjeld særleg korpuslingvistikk og leksikografiske studiar (jf. t.d.

-
- 1 Denne artikkelen er ein omarbeidd versjon av innlegget eg heldt på SONE-konferansen *Urbanitet, etnisitet og språk* i Kristiansand 15.–17. mars 2007. Fleire av tilhøyrarane der skal ha takk for nyttige kommentarar. Ein ekstra takk går til Randi Solheim som har kommentert også ei skriftleg utgåve av innlegget.
 - 2 Sjå til dømes nettstaden *Web Survey Methodology* (<http://www.websm.org>), som mellom anna inneheld ein rikhaldig bibliografi om nettbasert datainnsamling.

Grefenstette 2002), men òg innanfor dialektometri og dialekt-geografi har ein freista å utnytte verdsveven som innsamlingskanal.

I denne framstillinga skal eg gjere greie for ei slik utrøyning av Internett som medium for talemålsgransking. Inspirert av "Bergensartesten"³ som Gunnstein Akselberg (Universitetet i Bergen) fekk i stand saman med *Bergens Tidende* i 2004, utarbeidde eg ein dialekttest berekna på trondheimarar, som *Adresseavisen* gjekk med på å leggje ut i nettutgåva si (*adresa.no*). I løpet av fem månader i 2006 svarte nær 28 000 avislesarar på 28 spørsmål om trondheimsmålet, og bidrog på den måten til å samle inn store mengder talespråklege data. Nedanfor skal eg gå inn på ein del av dei metodiske spørsmåla eg støyte på i arbeidet med å utforme denne testen. Eg vil òg kommentere utfordringar som dukka opp då talmaterialet låg føre, og deretter freiste å seie noko om mogleg framtidig utnytting av denne innsamlingsmåten. Dei konkrete resultatata (dvs. dei språklege dataa) vil eg derimot halde utanfor i denne samanhengen, slik at den metodologiske diskusjonen får kome i framgrunnen.

2 Dialektgransking på nettet

Innsamling av talemålsdata gjennom Internett kan sjåast som ei forlenging av *spørjeliste-tradisjonen* slik vi kjenner han i klassisk dialektologi (jf. t.d. Chambers & Trudgill 1980:24f., Eichhoff 1982, Hoff 1968, Knoop, Putschke & Wiegand 1982:41f.). Spørjelistene inneber ei (hovudsakleg) *indirekte utspørjing* som tek sikte på å få medlemmer i eit visst språksamfunn til å rapportere språkbruken sin på utvalde punkt gjennom å fylle ut eit skriftleg skjema.⁴ Det er kanskje særleg det leksikalske nivået som har vorte utforska på denne måten (jf. Eichhoff 1982), men òg ulike grammatiske og fonetiske trekk har vore representerte på spørjelistene.

³ Sjå omtale i nettutgåva av *Bergens Tidende* 18. januar 2005: <http://www2.bt.no/tettpa/article331552>.

⁴ Det finst òg fleire døme på at spørjelistar har vore grunnlag for *direkte utspørjing* (jf. t.d. Chambers & Trudgill 1980:24f.), men i denne samanhengen er det særleg listeutfylling *utan* forskarar til stades som er mest relevant å sjå på som førebilete.

Den største føremonen med å overføre denne utspørjingsmåten til Internett er at ein dermed kan nå svært mange språkbrukarar raskt og til låge kostnader. Det er monaleg mindre ressurskrevjande å distribuere elektroniske skjema enn papirbaserte, og den vidare handsaminga av datatilfanget er òg mykje enklare i og med at materialet frå første stund er digitalisert.

Hittil er det truleg amerikanske dialektologar som har gått lengst i å utnytte Internett på denne måten.⁵ "Pop versus soda"-målinga⁶, som fekk inn over 250 000 svar på kva ord amerikanarar og kanadiarar brukar for "brus", er berre eitt av ei lang rekkje døme på meir eller mindre vitenskaplege "dialect surveys" som har vorte gjennomførte i USA ved hjelp av verdsveven. Ein kan òg nemne det britiske prosjektet "BBC Voices"⁷, som er eit samarbeid mellom BBC og fleire universitet, og som mellom anna har ført til at meir enn 62 000 personar har sendt inn leksikalsk materiale gjennom nettsidene til BBC. I det tyske språkområdet har Internett mellom anna vorte teke i bruk i innsamlingsarbeidet til det ordgeografiske prosjektet *Atlas zur deutschen Alltagssprache* (jf. Elspaß & Möller 2006)⁸. Dei same to forskarane har tidlegare prøvd ut mediet i to mindre, leksikografiske studiar (jf. Möller 2003 og Elspaß 2005), og begge konkluderer med at metoden kan ha mykje for seg i talemålsforskninga.

Så vidt eg kjenner til, er det gjort få nettbaserte dialektstudiar av noko større omfang i det skandinaviske språkområdet. Den måten Gunnstein Akselberg nytta ei nettavis til å samle inn talemålsdata på, var såleis eit pionereksperiment som det har vore spennande å følgje opp. Akselbergs undersøking høyrer til den typen av nettgranskingar som er *sjølvrekrutterande*. Det vil seie at spørjelista i

5 Sjå til dømes nettstaden <http://cfprod01.imt.uwm.edu/Dept/FLL/linguistics/dialect/index.html> for breiare omtale av (og resultat frå) ein av dei større nettbaserte dialektstudiane i USA dei siste åra.

6 Sjå nettstaden <http://www.popvsoda.com>.

7 Sjå nettstaden <http://www.bbc.co.uk/voices>.

8 Sjå òg nettstaden <http://www.philhist.uniaugsburg.de/de/lehrstuehle/germanistik/sprachwissenschaft/ada/>.

"Bergensartesten" ikkje vart sendt direkte til utvalde informantar, men i staden lagt ut på ein open nettstad der alle besøkjande kunne respondere. "Bergensartesten" hadde i tillegg eit meir *sosiodialektologisk* tilsnitt enn dei nettstudiane eg har nemnt ovanfor. Ved sida av dei reint lingvistiske sjekkpunkta vart det lagt inn spørsmål om sosiale makrofaktorar som bustad, kjønn og alder. Då eg utarbeidde den trondheimske versjonen var det naturleg å halde fast på Akselbergs opplegg, og grunnrissa til dei to testane er såleis nært innpå identiske.

3 Førebuingar og utforming

Tanken om å bruke *adressa.no* til å gjennomføre ein dialekttest vann temmelig raskt gehør då eg lanserte han for redaksjonsleiinga i *Adresseavisen*.⁹ At konseptet allereie var utprøvd med hell av *Bergens Tidende*, hadde truleg ein del å seie for velviljen eg vart møtt med, men det var òg tydeleg at emnet i seg sjølv engasjerte avisfolket. Det er nok heller ingen tvil om at utsiktene til auka trafikk på nettsidene var med på å motivere *adressa.no*-leiarane til å gå inn i prosjektet. For å generere flest mogleg besøk vart dialekttesten presentert for lesarane som eit "test-deg-sjølv"-konsept der deltakarane skulle få vurdert om dei snakka "fint" eller "breitt" trondheimsmål. Det må seiast at eg i utgangspunktet ikkje var så begeistra for denne sterkt forteikna og temmeleg tabloide vinklinga, men eg såg òg at det var ein effektiv og gagnleg måte å nå fram til lesarane på. I den vidare utforminga av testen gjekk eg derfor med på å leggje opp til ein "dobbel agenda": På den eine sida skulle testen munne ut i eit "pseudoresultat" – ein poengskår som plasserte respondenten i éin av fire kategoriar (sjå figur 1). På den andre sida skulle testformatet samle svara frå deltakarane i ein "skuggerekneskap" som utgjorde det eigentlege (vitskaplege) resultatet.

⁹ Ein takk går til dei impliserte i *Adresseavisen* – i fremste rekkje Anita Gustavsen, John Sund og Kai Vaade.

Score 27–35	Score 36–48	Score 49–62	Score 63–71
"Fintrøndersk"	"Moderat fintrøndersk" / "Dannet trondheimsmål"	"Moderat breittrøndersk" / "Folkelig bysmål"	"Breittrøndersk"
"Ditt talemål ligger tett opp til det som tradisjonelt blir kalt 'fintrøndersk'"	"Ditt talemål er en 'dannet' trondhjemsdialekt, men du har også innslag av 'breie' dialektformer"	"Ditt talemål er for en stor del 'breittrøndersk', men du har også innslag av 'penere' dialektformer"	"Ditt talemål ligger tett opp til det som tradisjonelt blir kalt 'breittrøndersk'"

Figur 1: Dei fire kategoriane som testdeltakarane vart sorterte i

Det som gav skin av å vere ei slags *diagnostisering* av talemålet til den einskilde (dvs. den noko diskutabile båssettjinga), hadde som einaste funksjon å vere eit insentiv til å gjennomføre testen. Redaksjonen og eg gjekk altså ut frå at lysta til å finne ut om ein er "fintrønder" eller "breittrønder" (eller noko midt imellom), var drivkraft nok til at lesaren ville spandere nokre minutt på skjemautfylling. Ei stund vurderte vi å bruke ei eller anna form for premiering som lokkemat, sidan dette har vist seg å høgne gjennomføringsprosenten i liknande testar (jf. t.d. Frick et al. 1999), men vi gjekk bort frå denne ideen, ikkje minst for å sleppe dei ekstra formalitetane som lagring av personopplysingar fører med seg. Det vart dessutan snart klart at rekrutteringa gjekk lett, sjølv utan freistingar om loddtrekning.

Den datatekniske designen overlét eg til mannskapet hjå *Adresseavisen*, og eg kjem derfor ikkje til å kommentere denne her. Eg vil i staden ta for meg arbeidet med å konsipere *innhaldet* i "Trondheimstesten" og utdjupe nokre av dei vanskanen eg møtte i denne prosessen.

3.1 Kva skal testast? Litt om variablane

Sjølv om eg kunne hatt god lyst til å bruke denne testen til å undersøkje trondheimsmålet på mange titals punkt, var det særleg to forhold som gjorde at eg måtte tøyne meg og skjere talet på variablar ned til 28: For det første er det sjølvsagt grenser for kor

mykje tid og tålmod ein avislesar har når det gjeld å fylle ut testskjema i ei nettavis (jf. Gräf 1999:157ff.). For det andre er det ein del språklege trekk som rett og slett eignar seg dårleg for eit slikt testformat fordi dei er vanskelege å framstille grafisk på ein lettfatteleg måte. Som eg kjem attende til nedanfor (3.3), er det heilt avgjerande for ei slik indirekte datainnsamling at respondenten utan for mykje møde forstår kva det blir spurt etter. Dersom eit sjekkpunkt synest uklart eller vanskeleg tilgjengeleg, er terskelen låg for å trekkje seg ut av heile testsituasjonen. Som Gräf (op.cit.:167) seier, er deltakarane berre interesserte i å henge med så lenge utspørjinga er *kognitivt lønsam* for dei. Testen må gje dei meir attende i form av til dømes underhaldning eller læring, enn det dei må leggje ned av innsats. I utveljinga av variablar vil ein måtte ta omsyn til dette, og då kan det hende at trekk må haldast utanfor, sjølv om dei utan tvil har stor lingvistisk interesse.

Eg støyte på dette problemet fleire gonger i mi eiga variabelselektering. Språkdrag som absolutt var aktuelle, måtte utelatast fordi eg ikkje fann nokon dugande måte å presentere dei på i det mediet eg ville bruke. Dette gjeld til dømes kvaliteten til konsonantsambandet /lt/ (som i *salt*), som høgst truleg har fleire variantar i dagens trondheimsmål. Men korleis eg skulle uttrykkje skilnaden mellom stemt og ustemt palatal lateral utan å ty til avanserte (og for dei fleste uforståelege) lydteikn, fann eg ikkje noko fullgodt svar på (sjå elles avsnitt 3.3 om lydskrift). Ei løysing kunne ha vore å leggje inn lydfiler med uttaledøme, men dette ville ha komplisert testformatet monaleg og dessutan ekskludert alle nettbrukarar som ikkje kan høyre lyd på datamaskinen. Det same problemet gjeld kvaliteten til lange dentalar (som utlyden i *snødd*, *vill* og *mann*). Dette trekket vart med som variabel i testen, men berre med variantane [+palatal] og [-palatal]. Eventuelle andre kvalitetar¹⁰ vart ikkje fanga opp – rett og slett fordi dei ikkje lét seg uttrykkje enkelt og eintydig i skrift.

¹⁰ Observasjonar frå eige feltarbeid tyder på at ein del trondheimarar tidvis uttaler desse konsonantane med noko i nærleiken av ein (apiko-)postalveolar eller retrofleks kvalitet.

Dei variablane eg vurderte som lettfattelege nok, vart omsette til 28 spørsmål med mellom to og fire svaralternativ (jf. vedlegg 1). Spørsmåla var med andre ord av typen "closed-ended" (jf. Best & Krueger 2004:52f.) i og med at dei moglege svara var gjevne på førehand. Eit slikt *lukka* format gjev eit materiale som er betrakteleg enklare å handsame, men det har òg uheldige sider. Ved å stengje for "fri" rapportering gjennom *opne* ("open-ended") spørsmål¹¹ risikerer forskarane å gå glipp av variantar som i utgangspunktet er ukjende for dei. I staden for at testdeltakarane blir stimulerte til å skrive inn svar på sjølvstendig grunnlag, blir dei leidde til å krysse av ved eit av dei oppgjevne alternativa, jamvel om det optimale svaret ikkje finst mellom desse. Sjølv om eg gjorde mykje for å kvalitetssikre dei ulike variablane, kan eg ikkje vere sikker på at alle aktuelle variantar var representerte i testen. Ei tilbakemelding frå ein av testdeltakarane tyder på at eg ikkje lykkast heilt: Då vedkommande skulle svare på kva ho sa for talordet *hundre*, fann ho nemleg ikkje "si" form mellom svaralternativa. Ho brukte etter seiande korkje *hundre*, *hunnjer* eller *honnjer*, men derimot *hundrede*. Og at denne forma inngår i trondheimsmålet, hadde eg ikkje teke høgde for.

Eg skal ikkje presentere alle dei 28 utvalde sjekkpunkta i full breidd her, men samanfatningsvis kan eg seie at dei dekkjer dei fleste lingvistiske nivåa, og at dei stort sett dreier seg om variasjon som har vore påvist i fleire samanhengar tidligare (jf. t.d. Dalen 1990 og Fintoft & Mjåvatn 1980). Det er ei overvekt av leksikalske og fonologiske spørsmål, medan syntaktiske og suprasegmentale trekk er lite representerte. I spørsmålslista vart variantane stundom framstilte som einskildord, andre gonger inngjekk dei i større syntagme som gjorde det tydelegare kva former det var tale om (sjå døma i figur 2). Lista vart innleidd av ein kort instruksjonstekst som mellom anna presiserte at det var den forma respondenten brukte *oftast*, som skulle rapporterast. Ein kunne altså krysse av ved berre eitt alternativ i kvart spørsmål, og eventuell *intraindividuell* variasjon vart såleis ikkje ansa.

¹¹ Dette vil seie spørsmål av typen "Kva ord brukar du for ungen til geita?". *Lukka* ("closed-ended") variant vil vere "Kallar du ungen til geita 1) kje eller 2) killing?".

2. – Huset står på stranna – Huset står på strannja	15. – Vi har skårre opp kjøttet – Vi har skjært opp kjøttet – Vi har skåre opp kjøttet	19. – husa – husan – husene
---	--	--------------------------------------

Figur 2: Døme på spørsmålsoppsett

Etter dei lingvistiske spørsmåla følgde ein bolck med sosiale variablar. Her vart respondenten beden om å gje opplysningar om *kjønn*, *alder*, *bustad i dag* og *bustad i oppvekståra*, og den som ville, kunne til slutt krysse av for *partipolitisk tilhørslse*. Dei 24 svaralternativa i spørsmåla om bustad var baserte på bydelsinndelinga i *Trondheimsstatistikken* (2001). I tillegg kom kategorien "utanfor Trondheim", som vart lagt inn som eit mogleg val for å kunne luke ut "urettkomne" respondentar.¹²

3.2 Vekting og skårar

For å kome fram til det eg ovanfor omtalte som "pseudoresultatet" av testen, vart variantane innanfor kvar variabel *vekta* i høve til ein "fint-breitt"-dimensjon. Det vil seie at former som kan seiast å vere meir standardnære (og i mange tilfelle nyare), fekk låg skår, medan meir tradisjonelle og standardfjerne former skåra høgare. Vi kan bruke verbet "være" som eit døme (jf. figur 3): Den mest standardnære (og presumptivt "fine") forma (*være*) blir vekta med 1, den apokoperte forma *vær'* blir rangert som "nest finast" og derfor vekta med 2, den jamna (men ikkje fulljamna) forma *værra* er "nest breiast" og blir vekta med 3, og den mest tradisjonelle og standardfjerne jamningsforma *vårrå* blir vekta med 4.

18.	Vekting
– å være	1
– å vær'	2
– å værra	3
– å vårrå	4

Figur 3: Døme på vekting av variantar

¹² Sjølv om det vart presisert at testen var meint for språkbrukarar med bakgrunn frå Trondheim, var det av ein eller annan grunn eit stort tal "ikkje-trondheimarar" som gjennomførte han.

Det er sjølvsagt mykje å innvende mot ei slik konvertering av språklege element til diskrete talverdiar – det gjev til dømes lite meining å seie at *vårrå* er *eksakt fire gonger så "breitt"* som *være* – men denne framgangsmåten gjorde det iallfall mogleg å kvantifisere svara frå den einskilde respondenten og rekne dei om til ein totalskår. Denne skåren vart brukt til å rangere testdeltakarane i høve til ein skala med 27 og 71 som ytterpunkt¹³, og vidare grupperer dei i fire kategoriar (jf. figur 1). Slik fekk kvar deltakar eit konkret testresultat i form ein karakteristikk av talemålet sitt, og samstundes ein indikasjon på korleis medborgarane fordelte seg på aksen "fint"–"breitt".¹⁴ Eg tvilar ikkje på at ein del følte det som urimeleg når dei på slutten av testen fekk språkbruken sin stempla som "fintrøndersk" eller "breitrøndersk", men i ein slik samanheng vil det vere umogleg å finne nemningar som ikkje på ein eller annan måte er ladde med sterkt subjektive verdiar. Det seier seg sjølv at dei kategoriane og merkelappane som vart utarbeidde til denne delen av testen, ikkje er objektive storleikar, og eg skal vere den første til å uttrykkje skepsis overfor ei slik overflatisk og simplifiserande tilnærming til språket¹⁵. Det vart likevel naudsynt å operere med desse *ad hoc*-kategoriane fordi testkonseptet kravde at vi hadde eit "resultat" å lokke avislesarane til oss med.

Den *reelle* datainnsamlinga låg i akkumuleringa av responsar innanfor kvar einskild lingvistisk og sosial variabel. Dette materialet vart samla i ein database til bruk i vidare statistiske analysar, og her kunne ein sjå heilt bort frå sorteringa av respondentar i fire grupper.

-
- ¹³ Observante lesarar vil sikkert leggje merke til at det rimar heller dårleg med ein minimumsskår på 27 når ein veit at alle dei 28 spørsmåla gav minst eitt poeng i uttelling. Årsaka til at den lågast moglege skåren likevel var 27, er at eitt av sjekkpunkta (nemleg "kjøpe"/"sjøpe"-variabelen) vart halde utanfor denne poengrekninga, rett og slett fordi det var vanskeleg å sameine med "fint"–"breitt"-dimensjonen.
- ¹⁴ Styrkeforholdet mellom dei fire kategoriane vart framstilt i eit kontinuerleg oppdatert diagram på testsidene (jf. vedlegg 2).
- ¹⁵ Den utstrakte bruken av hermeteikn i dei ulike karakteristikkane (jf. figur 1) kan sjåast som eit uttrykk for dette.

For å få eit mest mogleg letthandterleg materiale vart det lagt inn ein "avkryssingstvang" i testformatet. Det vil seie at respondenten måtte svare på *alle* dei lingvistiske spørsmåla for å få testresultatet sitt. Å nøyde testdeltakaren på ein slik måte kan sjølvsagt ha sine negative sider, men eg valde denne varianten for å sleppe det ekstraarbeidet som ufullstendige responsar fører med seg.

Eit anna tiltak for å betre kvaliteten på responsen var å setje variantane innanfor kvar variabel i tilfeldig rekkjefølgd. Tanken bak dette var at vi unngjekk at testdeltakaren "gjennomskoda" testformatet og dermed sette kryss på same stad heile vegen for å oppnå ein ønskt skår. Ved å plassere "fine", "halvfine" og "breie" former om kvarandre vart risikoen mindre for at testresultatet skulle bli påverka av slik *rutinisering*.

3.3 Skriftfesting av tale

Ettersom indirekte talemålsgransking i all hovudsak må baserast på skriftlege medium, fører det med seg ei substansiell utfordring: Korleis skal talt språk framstillast ved hjelp av allminnelege skrifteikn? (jf. Eichhoff 1982:550) Som eg var inne på ovanfor, er det heilt avgjerande for utfallet av ei slik datainnsamling at testpersonane med minst mogleg umak klarer å *avkode* dei problemstillingane testen konfronterer dei med. Manglande forståing kan gje negativt utslag på to måtar: For det første er det lite som skal til før respondenten mister interessa og trekkjer seg når testsituasjonen er så lite forpliktande som han er i denne samanhengen. Dersom testkonseptet blir opplevd som utilgjengeleg eller uskjonleg, vil det vere lett å hoppe vidare til ei anna nettside. For det andre vil resultatata bli misvisande og ubrukande dersom testdeltakaren ikkje skjønner kva det blir spurt etter. I testsituasjonen er det ikkje mogleg med rettleiing eller forklaring undervegs, og det er ingen til stades som kan medverke til å sikre at rapporteringa frå respondenten faktisk reflekterer røynda (jf. Knapp & Heidingsfelder 1999).

For å styre unna slike problem må framstillinga av testspørsmåla skje på lekmannens premissar, og spesialisten må fire på krava sine om vitenskapleg grannsemnd og systematikk. I dei fleste høve vil dette seie at ein må halde seg til det vanlege alfabetet og elles følge det

som finst av "folkelege konvensjonar". Bruk av lydsymbol eller andre formaliserte teiknsett som berre er kjende for fagfolk, vil vere fåfengt og innebere ein risiko for misforståing og feiltolking.

I den endelege spørjelista i "Trondheimstesten" har ein slik "lekmannslogikk" fått gjennomslag fleire stader. For å framstille skilnaden i trykkplassering i ord som "problem" har eg til dømes brukt både apostrofar og kursivering på ein temmeleg ulingvistisk måte.¹⁶ Ustemt postalveolar frikativ blir notert *sj* eller *skj* i tråd med den offisielle stavemåten til det aktuelle ordet, og trykksvake e-ar (som i *være*) blir skrivne *e* sjølv om dei i realiteten nok er reduserte til ein schwa (ə). Det å skulle halde seg strengt til vanleg skrivepraksis viste seg likevel å vere vanskeleg og utilfredsstillande – ikkje minst når ein, som eg, har med *trønderske* talemål å gjere. I dette området finn vi ei rekkje språklydar (særleg konsonantvariantar) som ikkje blir omfemna av det ordinære norske alfabetet, og mange av desse inngår i variasjon som det utan tvil vil vere interessant å røkje etter i ein test av denne typen. Som eg allereie har vore inne på (3.1), var det fleire variablar eg måtte slå frå meg fordi dei medførte ei for komplisert skriftfesting. Men det vart òg til at eg valde å skiple det vanlege skriftbiletet på eit par punkt, slik at eg fekk lagt nokre av dei "trønderske lydane" under lupa. Palataliserte kvalitetar av lange dentalar (som i *snødd*, *vill* og *mann*) var eg svært huga til å inkludere, og eg måtte derfor finne ein måte å uttrykkje dei på som ikkje forvirra eller villeidde lesaren. Lars S. Vikør (1997, 2004) har gått nøye inn på korleis ikkje-lingvistar i ulike samanhengar har skriftfesta talespråk, og han slår fast at nettopp palatalisering av dentalar har vore ein vande for mange trønderske skribentar (2004:278). Innskot av *i* eller *j* føre konsonanten (*mainn* eller *majnn*) har vore dei vanlegaste noteringsmåtene, men ein kan òg finne *j*-en etterstilt (*mannj*) (Vikør 1997:342f. og 2004:278), og det er denne siste varianten eg bestemte meg for å nytte. Årsaka til at eg gjekk bort frå det som kan seiast å vere den mest utbreidde "lekmannskonvensjonen", er at eg var redd denne notasjonen ville "uroe" skriftbiletet for mykje. Vikør (ibid.) er òg inne på at bruken av

¹⁶ Førsteleddstrykk vart notert <'problem>, medan trykk på sisteleddet var framstilt slik: <pro'blem>.

innskoten *i* som palatalmerke kan tyngje teksten, særleg fordi det dermed oppstår mange nye diftongar. Også etterstilt *j* er sjølvsagt eit brot på skriftnormalen, men håpet mitt var at dette avviket var mindre påfallande og derfor mindre "sjenerande" for testdeltakarane.

Det er elles ikkje vanskeleg å finne inkonsekvensar og manglande systematikk i måten testspørsmåla er skrivne på. Til dømes er preposisjonen *under* skriven på normert form, medan substantivet *stranda* er skrivi *stranna/strannja*. Alle slike tillempingar er gjorde i tråd med det eg reint intuitivt trur ein trøndersktalande nettavislesar vil oppleve som mest "naturleg" og lettfatteleg.

4 Kven har vi egentleg testa? Utfordringar knytte til validitet og representativitet

Etter at testen hadde vore aktiv i om lag fem månader, vart han avslutta, og eg sat att med totalt 27 798 utfylte elektroniske spørjelister. Eg kan med stor visse seie at trondheimsmålet aldri før har vore undersøkt med så mange informantar, og det er sjølvsagt lett å bli euforisk når ein tenkjer på alle dei moglege analysane ein kan få ut av eit slikt rekordstort materiale. Det er likevel nok av dei som vil helle kaldt vatn i blodet på ein eldhuga talemålsforskar. Mellom dei som manar til nøkternheit i utnyttinga av Internett, er statsvitarane Samuel J. Best og Brian S. Krueger (2004:3f.). Dei peikar på to hovudproblem relaterte til utnyttinga av denne kanalen til å samle inn data: For det første er tilgangen til og bruken av verdsveven til dels svært varierende, noko som gjer at nettbrukarane som gruppe berre delvis speglar folkesetnaden som heilskap. For det andre er det snakk om ein eksperimentsituasjon som er ekstremt vanskeleg å kontrollere, og som derfor lett kan gje ugyldig og ubrukande respons. Begge desse utfordringane er høgst aktuelle i samband med denne testen, og eg skal her gå litt nærare inn på dei.

4.1 Kven er på nett, og kva gjer dei der?

Dersom ein brukar typologien hjå Couper (2000), kan "Trondheimstesten" kallast ein "self-selected web survey" (dvs. ei nettundersøking der respondentane sjølve vel om dei vil delta), og eit slikt format har

reinhekle statistikarar vondt for å akseptere som vitskapleg haldbart. Couper har i det heile mykje å utsetje på datainnsamling som skjer på denne måten (eller rettare sagt bruken av dei innsamla dataa som grunnlag for generaliseringar), og han ser ikkje at ein slik test har stort anna enn underhaldningsverdi (op.cit.:479). Atteslander (2005:1072f.) kjem òg med kraftige åtvaringar mot å ha overdriven tru på datainnsamling gjennom Internett. Han peikar på at det framleis er langt ifrå alle språkbrukarar som har tilgang til dette mediet i det daglege, og at det derfor vil vere vanskeleg å oppnå ein dugande *representativitet* i nettbaserte undersøkingar. Ifølgje Hewson (2002:27f.) har representativiteten betra seg dei siste åra, men framleis er dette eit av dei svake punkta ved datainnsamling gjennom Internett. Heile språksamfunnet er enno ikkje til stades på verdsveven, og det er store skilnader i kva bruk folk gjer av dette mediet. I Noreg er desse ulikskapane mindre enn i mange andre land, men òg hjå oss skil enkelte grupper seg ut (Vaage 2007:62ff.). Den mest markante variasjonen er knytt til alder. I 2006 var 17 % av nordmenn mellom 67 og 79 på Internett ein gjennomsnittsdag, medan talet for aldersgruppa 16–24 år var heile 80 %. Ein kan dessutan sjå visse skilnader i måten dei ulike aldersgruppene nyttar Internett-tenester på. Vaksne brukar tilbod som bank- og tingingstenester, e-post og offentleg informasjon, medan born og unge er dei mest aktive brukarane av underhaldningstilbod. Det er òg skilnader i nettbruken dersom ein ser på utdanningsnivå. Dei med lang utdanning (universitet/høgskule) brukar Internett nær dobbelt så mykje som dei med ungdomsskule som høgaste utdanning.

Ein kan påvise tydelege utslag av desse aldersmessige skilnadene i materialet mitt. Av alle trondheimarar over 60 år var det berre noko slikt som 1,5 % som tok denne testen, medan om lag 16 % av alle i alderen 15–40 år er representerte. Datatilfanget har med andre ord ei vesentleg skeivfordeling som har direkte samband med mitt val av medium.

Representativiteten er truleg òg påverka av måten rekrutteringa av respondentar har skjedd på. Testformatet mitt legg opp til det som Taylor (2000:53f.) kallar "volunteer" eller "convenience sampling" i staden for "probability sampling" (utval basert på sannsyn). Det vil

seie at dei som til sist hamnar i databasen, gjer det fordi dei slumpa til å støyte på testen i løpet av nettavislesinga si, og ikkje fordi dei vart utplukka med særleg tanke på å oppnå eit representativt utval av populasjonen.

4.2 Kva skjedde i testsituasjonen?

Den andre hovudinnvendinga statistikarane reiser mot denne typen av nettundersøking, dreier seg om at testsituasjonen er for dårleg kontrollert frå forskarane si side (jf. Couper 2000). Eit testkonsept som ligg ope tilgjengeleg i ein sterkt trafikkert mediekanal¹⁷, vil sjølvstøtt vere utsett for fleire typar manipulering. Det er til dømes svært vanskeleg å fastslå kven som faktisk står bak kvar respons, slik at ein kan sikre at den enkelte nettbrukar berre leverer eitt svar. I "Trondheimstesten" vart IP-adressene¹⁸ registrerte saman med dei andre dataa, og det er dermed mogleg å sjå om det er levert fleire svar frå éi nett-tilkopling. Men doble eller mangedoble oppføringar av same IP-adresse treng sjølvstøtt ikkje vere noko prøv på at enkelte språkbrukarar har "fuska" og delteke fleire gonger. Eg kjenner til at heile skuleklassar har teke testen som ein del av undervisinga, og det er ikkje utenkjeleg at 10–15 elevar har levert svar gjennom same pc innanfor kort tid. Ein kan òg tenkje seg mange andre grunnar til at fleire personar går inn på nettet gjennom same datamaskin, og det er derfor ikkje noko alarmerande i at same IP-adresse dukkar opp ei rekkje gonger i materialet.

Det opne og anonymiserte opplegget til "Trondheimstesten" gjer det samstundes umogleg å finne ut om personar med tilgang til fleire pc-ar (med unike IP-adresser) har utnytta dette til å svare fleire gonger. No er det sjølvstøtt måte på kor sterkt vi skal mistru publikum, men ein kan likevel ikkje sjå bort frå at nokre testdeltakarar har gått aktivt inn for å manipulere resultatane. Respondentane har naturleg nok mange ulike motivasjonar og

¹⁷ Nettutgåva til *Adresseavisen* hadde i 2006 gjennomsnittleg 108 000 lesarar kvar dag (jf. målingar frå *medienorge*: <http://medienorge.uib.no/?cat=statistikk&medium=avis&queryID=253>).

¹⁸ Kort sagt er ei IP-adresse eit unikt nummer som identifiserer datamaskiner via *Internet Protocol*, og som dermed gjer det mogleg å spore kva eining innafor eit datanett som står bak informasjon som blir overført.

intensjonar bak si deltaking, og ifølgje Hauptmanns (1999:28) må ein rekne med å støyte på ein del "aktivistar" som responderer gjentekne gonger, og slik "Trondheimstesten" er lagt opp, er det lite ein kan gjere for å eliminere denne feilkjelda. Eit anna korrumperande innslag er dei som leverer det vi kan kalle "hyperkorrekte" svar. König (2004) har avdekka korleis enkelte informantar til språkgeografiske studiar rapporterer "det rette og beste" for å "hjelpe" forskaren, sjølv om responsane deira ikkje stemmer med røynda. Det er umogleg å seie noko om korleis dette kan ha slått ut på "Trondheimstesten".

Ei anna vesentleg utfordring når det gjeld å kontrollere kvaliteten på responsen, er at forskaren er maksimalt fråverande i intervju-situasjonen (jf. t.d. Pötschke & Simonson 2001:16). Dette gjer det umogleg å kome med oppfølgingsspørsmål eller på andre måtar validere den informasjonen respondenten kjem med. Eventuelle misforståingar eller feiltolkingar av sjekkpunkta vil ikkje bli avdekte, og som Taylor (2000:53f.) viser, fører dette i mange høve til fleire svar i "veit ikkje"/"usikker"-kategorien.

Den same problematikken er sjølvsagt òg relatert til alle andre former for *indirekte* utspørjing, men ein gjer seg kan hende ekstra sårbar når ein nyttar eit nettbasert massemedium som kanal. Det er grunn til å tru at ein nettavislesar assosierer dette mediet meir med *underhaldning* enn med *vitskaplege* eller *faglege* aktivitetar. Eit testkonsept som dukkar opp i ein slik kontekst, vil derfor vonleg bli teke mindre alvorleg enn eit spørjeskjema som kjem som personleg e-post, og ein kan tenkje seg at terskelen for å vere uærleg og unøyaktig dermed blir monaleg lågare. Den temmeleg populære framstillinga av temaet kan òg ha vore med på å underbygge dette inntrykket, og ein må derfor rekne med at avislesarane har nærma seg testen med ulik grad av seriøsitet.¹⁹

¹⁹ Som eit døme på denne humoristiske vinklinga kan vi sjå illustrasjonen som fungerte som det sentrale blikkfanget til testen. Dette var ein kollasj der ansikta til fem kjente trondheimarar var monterte inn på "fint"- "breitt"-skalaen med Flettfriid Andrésen og Odd Reitan som ytterpunkt (jf. vedlegg 2).

5 Oppsummering

Som eg allereie har vore inne på, utgjer ikkje datainnsamling gjennom Internett i seg sjølv noko nytt *metodisk prinsipp*; grunnteknikken er den same som i t.d. Georg Wenkers "Wenker-Bögen" frå 1870-åra (jf. Knoop, Putschke & Wiegand 1982) og Johan Storms "Kortere Ordliste" frå 1884 (jf. Bandle 1962:296). Ein kan heller snakke om ei radikal *vidareutvikling* gjennom at eit heilt nytt medium eller instrument blir teke i bruk i innsamlingsarbeidet, og som vi har sett ovanfor, har dette mediet både overbevisande fordelar og graverande ulemper. Heilt kort kan vi seie at denne framgangsmåten gjennomgåande gjev oss eit datatilfang med *suveren kvantitet*, men med noko *tvilsam kvalitet*. Til ein viss grad kan den veldige storleiken på materialet vege opp for nokre av dei skeivskapane som nødvendigvis vil oppstå når det gjeld representativitet og validitet, men det er òg grunn til å ta monalege atterhald om kor mykje ein kan generalisere ut frå denne typen data.

Det formatet som "Trondheimstesten" byggjer på, har i statistikarens auge altfor mange veike punkt til å kunne seiast å vere påliteleg (jf. t.d. Best & Krueger 2004, Couper 2000). Ein del av desse kritiske merknadene er eg samd i, men eg er ikkje utan vidare villig til å forkaste materialet frå *adresa.no*-prosjektet som verdilaust; eg er viss på at dei dataa som ligg føre, vil kunne omsetjast til nyttig innsikt i fleire aspekt ved trondheimsmålet. Innvendingane om at testsituasjonen er sårbar for manipulasjon og feilaktig respons, må eg sjølvsagt ta omsyn til i analysane, men på variabelnivå vil ein utan tvil kunne sjå interessante tendensar som kan seiast å vere gyldige for store delar av det trondheimske språksamfunnet.

Det har elles vore ei interessant oppleving å sjå kor stort engasjementet er blant folk når det gjeld talemål og ikkje minst nemningar for talemål. Eg fekk ei lang rekkje tilbakemeldingar med både ros og ris frå testdeltakarar, og testen avfødde fleire dialektdebattar i ulike nettforum. Dette innsynet i *folkelingvistikken* meiner eg er ein svært sunn og viktig *bieffekt* som meir "usynlege" datainnsamlingsmåtar ikkje evnar å oppnå.

Eg blir neppe omtalt som nokon dristig spåmann når eg seier at Internett kjem til å bli eit stadig viktigare verktøy for datainnsamling i framtidig talemålsforskning. Store, kvantitativt orienterte studiar der

direkte utspørjing er ein uaktuell framgangsmåte, vil ha mykje å vinne på å utvikle gode nettbaserte spørjeverktøy. Gjennom det eksperimentet eg har omtalt her, har vi fått kartlagt noko av potensialet til dette mediet, og med visse justeringar og forbetringar vil vi etter kvart nå fram til eit testformat som er raffinert og pålitande nok til at det kan bli ein hovudreiskap for indirekte datainnsamling innanfor talemålsforskinga. Dette føreset sjølvstøtt utbreiinga av nettilgang vil halde fram med å auke (jf. Atteslander 2005:1075), men det er vanskeleg å sjå for seg at Internettet vil bli mindre daglegdags og allminneleg i åra som kjem.

Litteratur

- Atteslander, Peter (2005): "Schriftliche Befragung" i Ammon et al. (red.): *Sociolinguistics: An International Handbook of the Science of Language and Society*. Berlin–New York: Walter de Gruyter. Band 2, 1063–1076.
- Bandle, Oskar (1962): "Die norwegische Dialektforschung" i *Zeitschrift für Mundartforschung*, 29, 289–312.
- Batinic, Bernad (2001): *Fragenbogenuntersuchungen im Internet*. Aachen: Shaker Verlag.
- Bergens Tidende*, 18. januar 2005.
- Best, Samuel J. & Krueger, Brian S. (2004): *Internet Data Collection. Quantitative Applications in the Social Sciences* 141. Sage Publications.
- Chambers, J.K. & Trudgill, Peter (1980): *Dialectology*. Cambridge: Cambridge Univ. Press.
- Comley, Pete (1996): "The use of the Internet as a data collection method" i *Proceedings of the ESOMAR/EMAC Symposium on Research Methodologies for the New Marketing*, ESOMAR Publication Series, Vol. 204, 335–346. [Nettpublikasjon: http://www.virtualsurveys.com/news/papers/paper_9.asp]
- Couper, Mick P. (2000): "Web surveys. A review of issues and approaches" i *Public Opinion Quarterly*, 64, 464–494.
- Crystal, David (2005): "Johnson and the Internet". *Hilda Hulme Memorial Lecture* holden ved University of London, 21. april 2005. [Nettpublikasjon: <http://ies.sas.ac.uk/Publications/hulme/crystal-johnson.pdf>]
- Dalen, Arnold (1990): "Talemålet i Trondheim" i Jahr (red.): *Den store dialektboka*. Oslo: Novus forlag. 205–209.
- Eichhoff, Jürgen (1982): "Erhebung von Sprachdaten durch schriftliche Befragung" i Besch et al. (red.): *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Band 1. Berlin: de Gruyter. 549–554.

- Elsaß, Stephan (2005): "Zum Wandel im Gebrauch regionalsprachlicher Lexik" i *Zeitschrift für Dialektologie und Linguistik*, 72(1), 1–51.
- Elsaß, Stephan & Robert Möller (2006): "Internet-Exploration: Zu den Chancen, die eine Online-Erhebung regional gefärbter Alltagssprache bietet" i *Osnabrücker Beiträge zur Sprachtheorie*, 71, 143–158.
- Fintoft, K.E. & Mjåvatn, P.E. (1980): *Språksosiologiske forhold i Trondheim bymål*. Trondheim: Tapir forlag.
- Frick, A., Bächtiger, M.T. & Reips, U.-D. (1999): "Financial incentives, personal information and drop-out rate in online studies" i Reips et al. (red.): *Current Internet Science. Trends, techniques, results*. Deutsche Gesellschaft für Online-Forschung. [Nettpublikasjon: http://gor.de/gor99/tband99/pdfs/a_h/frick.pdf]
- Gräf, Lorenz (1999): "Optimierung von WWW-Umfragen: Das Online Pretest-Studio" i Batinic et al. (red.): *Online-Research. Methoden, Anwendungen und Ergebnisse*. Göttingen: Hofgreffe. 156–177.
- Grefenstette, Gregory (2002): "The WWW as a Resource for sLexicography" i Corréard (red.): *Lexicography and Natural Language Processing*. EURALEX. 199–215.
- Hauptmanns, Peter (1999): "Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet" i Batinic et al. (red.): *Online-Research. Methoden, Anwendungen und Ergebnisse*. Göttingen: Hofgreffe. 21–38.
- Hewson, Claire (2002): *Internet Research Methods. A Practical Guide for the Social and Behavioral Sciences*. London: Sage Publications.
- Hoff, Ingeborg (1968): "Norwegische Mundartforschung" i Schmitt (red.): *Germanische Dialektologie. Festschrift für Walther Mitzka zum 80. Geburtstag*. Wiesbaden. 398–458.
- Knapp, F. & Heidingsfelder, M. (1999): "Drop-Out-Analyse: Wirkungen des Untersuchungsdesigns" i Reips et al. (red.): *Current Internet Science. Trends, techniques, results*. Deutsche Gesellschaft für Online-Forschung. [Nettpublikasjon: http://www.gor.de/gor99/tband99/pdfs/i_p/knapp.pdf]
- Knoop, Ulrich, Putschke, Wolfgang & Wiegand, Herbert E.: "Die Marburger Schule: Entstehung und frühe Entwicklung der Dialektgeographie" i Besch et al. (red.): *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Band 1, 38–91. Berlin: de Gruyter.
- König, Werner (2004): "'Stimmen tut's nicht, aber es ist was Besonderes!' Die Motivation/Intention von Informanten als Faktor für 'hyperkorrekte' Antworten bei Sprachgeographischen Erhebungen" i Greule, Hochholzer & Wildfeuer (red.): *Die bairische Sprache. Studien zu*

- ihrer Geographie, Grammatik, Lexik und Pragmatik. Festschrift für L. Zehetner.* Regensburg: Vulpes. 189–193.
- Möller, Robert (2003): "Das rheinische tschö" i *Rheinische Vierteljahrsblätter*, 67, 333–339.
- Pötschke, Manuela & Simonson, Julia (2001): "Online-Erhebungen in der empirischen Sozialforschung: Erfahrungen mit einer Umfrage unter Sozial-, Markt- und Meinungsforschern" i *ZA-Information*, 49, 6–28.
- Storm, Johan (1884): *Kortere Ordliste med Forklaring af Lydskriften.* [Kristiania].
- Taylor, Humphrey (2000): "Does internet research work?" i *International Journal of Market Research*, 42(1), 51–63.
- Trondheimsstatistikken* (2001). Trondheim kommune: Byplankontoret.
- Vikør, Lars S. (2004): *Dialektar som skriftspråk i tre norske distrikt.* Oslo: Novus.
- (1997): "Dialekt som skriftspråk i Trøndelag" i Reinhammar (red.): *Nordiska dialektstudier.* Språk- och folkminnesinstitutet. 341–347.
- Vaage, Odd Frank (2007): *Norsk mediebarometer 2006.* Oslo–Kongsvinger: Statistisk sentralbyrå.

Vedlegg 1: Skjermbilete av testen

Spørsmål 1	<input type="radio"/> Vi krøyp under bordet <input type="radio"/> Vi kraup under bordet <input type="radio"/> Vi krøp under bordet
Spørsmål 2	<input type="radio"/> Huset står på stranna <input type="radio"/> Huset står på strannja
Spørsmål 3	<input type="radio"/> En gris / to grise <input type="radio"/> En gris / to grisa <input type="radio"/> En gris / to griser
Spørsmål 4	<input type="radio"/> Flisk <input type="radio"/> Fesk
Spørsmål 5	<input type="radio"/> 100 = hunnjer <input type="radio"/> 100 = honnjer <input type="radio"/> 100 = hundre
Spørsmål 6	<input type="radio"/> Kvar <input type="radio"/> Kvær <input type="radio"/> Vær
Spørsmål 7	<input type="radio"/> Vi har spist sillj <input type="radio"/> Vi har spist sill
Spørsmål 8	<input type="radio"/> Jel <input type="radio"/> Æ
Spørsmål 9	<input type="radio"/> Her / Der <input type="radio"/> Hær / Dær
Spørsmål 10	<input type="radio"/> Vi har fri hele året <input type="radio"/> Vi har fri heile året.
Spørsmål 11	<input type="radio"/> <i>'problem</i> <input type="radio"/> <i>pro'blem</i>
Spørsmål 12	<input type="radio"/> Det ble mørkt <input type="radio"/> Det vart mørkt <input type="radio"/> Det blei mørkt
Spørsmål 13	<input type="radio"/> Vi sjøpe brus i butikken <input type="radio"/> Vi kjøpe brus i butikken
Spørsmål 14	<input type="radio"/> Redd <input type="radio"/> Redd

Vedlegg 2: Uttdrag frå testframsida

Hva slags trondheimsmål snakker du?



Test deg selv og finn ut om du snakker fint eller bredt

Trondheimsdialekt kan være så mangt, og enhver som oppholder seg i byen en stund, vil legge merke til en viss variasjon i måten trondheimere snakker på. Tradisjonelt har det vært betydelige forskjeller mellom for eksempel "iling", "lamo-dialekt" og "singsakersk", men det er mer uklart hvordan dette er i dag.

