

Hege Sandøy

DERFOR KAN IKKJE MATERIALET VÅRT BLI FULLSTENDIG!

1. Innleiing

I ei foreløpig orientering til denne konferansen stod det under det foreslåtte emnet "Det daglige" eit stikkord: "samlingarnas fullstændighet". Stikkordet kalla fram att ein viss irritasjon egg stundom har hatt over holdningar ein ofte møter i den nordiske språkvitenskapen, og ikkje minst i dialektologien, der vi har arkiv. Kanskje er ikkje irritasjonen alltid like rettkommen. Men de får ta dette innlegget som ei lita meiningstyring. Og i ei slik ytring gjør det seg best om ting er sett på spissen:

I arkiva samlar og samlar vi, for vi vil helst ikkje ha att kvite flekkar på kartet. Det er støtt eit ideal å ha mye materiale. For dermed kan framtidige forskarar få eit solid grunnlag å bygge på. — Slik verkar det som arbeidsfilosofien har vore.

Men frå arbeidet i arkiva veit vi at lite av det arkiverte materialet blir brukt — i alle fall i den tida vi har oversikt over. Det er dermed svært usikre investeringar vi gjør.

Og når vi så ein hende gong hentar fram arkivmateriale, oppdagar vi at det er vanskelig å få svar på det vi spør etter. Og vi får problem med å tolke opplysningane fordi vi ikkje ser kva materialet i si tid var målretta imot. Kanskje det ikkje var målretta i det heile. Dataa er både ufullstendige og upålitelige i forhold til spørsmåla vi stiller.

Alt arbeidet vi har lagt ned i datainsamling i tradisjonell dialektologi, er ikkje alltid like lett å forsvare. Men sosiolingvistikken har ikkje gjort dette forholdet betre. Heller tvert imot, kan det sjå ut til. For den har mest reagert imot andre sider ved arbeidsmåten i den tradisjonelle dialektologien, f.eks.

bruken av spørjelster. I sosiolingvistikken har ein hatt som ideal å få opptak av såkalla "naturlig tale" i "naturlige situasjonar". Resultaret har så blitt uendelige opptak i time etter time – med "naturlig tale". Og kanskje først når innsamlinga endelig er over, har ein begynt å tenke på analysen.

2. Induktiv arbeidsmåte

Dette er den induktive arbeidsmåten: Utifrå eit materiale skal ein finne ut noe. Og for ein har fått materialet, veit ein også lite om kva ein skal finne ut. Men di meir materiale vi har, di "fullstendigare" er det, og di meir solide resultat får vi. Trur ein. Kanskje er det slik ein driv arkeologiske museum: ein ser det som ein fordel å få inn stadig fleire spydspissar. Det er kan hende av interesse å rekne ut gjennomsnittslengd og gjennomsnittsbreidd på spydspissane, og svaret blir meir nøyaktig di større grunnlaget er. Det er jo òg teoretisk moglig å få inn alle spydspissane som har eksistert.

Ikke veit eg. Men så langt er det i alle fall tale om konkrete produkt. Og konkrete gjenstandar har ein *kvantitet*, og dei har eit totaltal, eller ein *populasjon*.

3. Forskingsobjektet 'språk'

Men slik er det ikkje med forskingsobjektet 'språk'. Det språklige som vi aller oftest er interessert i, er ikkje det konkrete produktet, men *normene* bak språkproduksjonen, dvs. normene for moglig språkbruk. Og normene har ingen populasjon. Dei blir brukt og kan brukas i det uendelige. Vi lagar eit finfint tal med setningar med hjelp av eit finfint tal med reglar, som det heiter i lingvistikken i dag. Derfor er det også meningslaust å snakke om 'fullstendighet' sett i forhold til ein populasjon eller ein kvantitativ størrelse. Materialet treng ikkje bli meir fullstendig om det blir større.

I språkstudiet er vi interessert i eit abstrahert system som gir ein adekvat beskrivelse av ei normert og regelstyrt åtfærd, nemlig talen (= realisert språk). Dette fundamentale utgangspunktet må vi ta omsyn til når vi vel oss forskingsmetodar. Utgangspunktet er at vi aldri kan ha forskingsobjektet framfor oss som eit objekt! Vi må nøye oss med bare resultatet av det. Og med å *tolke*

produktet eller resultatet av språknormene kan vi finne ut noe om det eigentlige forskingsobjektet vårt. Altså indirekte. Innafor kjernefysikken og astronomien er ikkje denne situasjonen ukjent.

Ser vi for oss dette at vi ikkje studerer sjølve språkuttrykket, men heller *normene* som styrer språkuttrykket, er det lettare å innsjå at talesituasjonane våre, når vi gjor opptak eller oppskriftar, er ein type *eksperimentelle situasjonar*. Og poenget med eksperiment er at ein prøver å *styre vilkåra* for det som skjer.

Dette er sjølv sagt ikkje heilt nytt. Men det manglar noe i den eksperimentelle holdninga vår. Idealet om ein "naturlig talesituasjon" kan avspegle ei tru på at det fins ein type materiale som er det "verkelige" studieobjektet.

Også "naturlige samtalesituasjonar" er eksperimentelle situasjonar, men ein prøver å definere ein situasjon som er vanligere enn andre situasjonar. Og så styrer ein situasjonen deretter. Men vi oppdagar ofte at ulempa med såkalla "naturlige situasjonar" er for lita styring, i alle fall etter lingvistiske interesser. Jamfør f.eks. vanskane ein møter med å få belegg til visse syntaktiske konstruksjonar.

Situasjonane styrer ein systematisk ettersom: ein eksperimenterer med vilkår som ligg i *situasjon* eller *grammatikk*. Denne systematiske styringa skjer sjølv sagt etter gjetingar vi gjor oss om *kva* faktorar som betingar den språklige återda og *kortleis*, dvs. vi gjettar oss til ei formulering av norma. Vi lagar ein *hypotese*. Og den blir testa i eksperimenta.

På sine måtar er jo dette ein rein banalitet. Men i praktisk arbeid ser det ofte ut til at vi glemmer dette teoretiske og metodisk grunnleggande utgangspunktet. For vi arbeider med levande språk som om dei var daude språk. Daude språk kan ikkje få utvida korpuset sitt, vi kan sjå på det som ein avslutta populasjon. Og vi kan heller ikkje drive eksperiment. Derfor vil forskinga av slike språkstadium sjølv sagt bli meir prega av induktive arbeidsmåtar. Det er ein *prinsipiell* forskjell på daude og levande språk.

4. Hypotetisk-deduktiv arbeidsmåte

Resultata av eksperimenta våre kan fortelje oss om hypotesen er galen, dvs. om vi kan *avkrefte* hypotesen. Kan vi det, er neste steget å lage ein *ny*, men meir *presis* hypotese, som vi igjen kan teste. På den måten kan vi håpe stadig

å komme nærare sannheita. Men vi kan aldri stadfeste eller *verifisere* empirisk interessante hypotesar om språknormer, dvs. generaliseringar om språknormer. For det blir jo stadig produsert nye språklige ytringar, og snart kunne det jo dukke opp eksempel som avkrefar hypotesen likevel. Men greier vi ikkje å avkrefte ein hypotese, er sjølv sagt hypotesen *styrkt*.

Dette er viktig for utgangspunktet vårt, for det betyr at arbeidsmåten vår dermed må vere å lage stadig *meir presise* hypotesar ved at vi greier å avkrefte gale formulerte hypotesar eller spesifisere vidare allereie styrkte hypotesar. Kunnskapane våre om språket eller språknormene kan altså aldri få annen status enn som 'hypotesar', men dei kan bli stadig meir presise.

Dermed er vi òg komme til eit viktig punkt: Vi treng ikkje så mye materiale for å avkrefte ein hypotese dersom testen eller eksperimentet er gule konstruert. Sjølv sagt blir avkreftinga meir solid med eit f.eks. ti gonger større materiale, men det er feilinvesterte krefter å slite seg ut på å arbeide med det store materialet når det viktigaste likevel er å forbetre hypotesen for å kunne gå vidare i forskingsprosessen.

Eit levande språk utforskar vi altså hypotetisk-deduktivt — som vi lærte å kalle det til førebunde filosofii. Og min kritikk er da at altfor mye av arbeidet i nordisk språkvitenskap, og dialektologi spesielt, har i realiteten lagt ein induktiv arbeidsmåte til grunn. Derfor blir det lagt så stor vekt på data-innsamling. Vi slit oss ut på ei datainsamling som er altfor lite målretta.

Eit materiale må vere tilpassa slik at det gir oss data som er relevante for hypotesen vår. Dermed må 'pålitelig' og 'fullstendig' òg tolkas i forhold til denne hypotesen. 'Pålitelig' må bety at vilkåra for språkproduksjonen *verkeilig tilfredsstillende* dei vilkåra hypotesen reknar som relevante. Og 'fullstendig' må vel tolkas dit at *det* betyr at materialet *dekker* alle dei kriteriuma og variablane som må testas.

5. Talemålsprosjekt

Ikkje reint lite materiale er samla inn på måfå. Kanskje gjør vi det fordi vi er redde for at dialekten og informantane skal dø ifrå oss? Sjølv har eg kjent meg i den situasjonen. Men vi får ein bakvent arbeidsmåte: vi samlar *først*, og tenker på analysen *etterpå*. Og feltarbeidet kan jo vere så slitsamt i seg sjølv at ein ikkje har overskott til å konsentrere seg om noen analyse før innsamlingsfasen er over.

Neste fase viss ein vil utsette enda lenger analysearbeidet, er å *transkribere* alt det innsamla materialet.

Forestillinga om at fonetisk transkripsjon er så nøyktig, sit svært godt i. Det kan ho sjølv sagt vere, men *treng* ikkje vere det. For er ikkje denne transkripsjonen målretta, veit vi heller ikkje kva distinksjonar som er relevante for eitt eller anna. Poenget med *fonematisk* skrift er nettopp at den har definert *kva* som er relevant og ikkje. Skal vi over til det fonetiske, må vi definere nye kriterium for relevans alt etter formålet, for sjølv lydmassen viser bare gradsforskjellar. I all transkripsjon må ein støtt tolke og gjøre val. Er ikkje formålet med transkripsjonen avklart, kan ein gjøre val som egentlig vil vise seg å føre til feil i samband med ei framtidig problemstilling.

Før var det svært vanlig å skrive fonetisk. Eg er sikkert ikkje aleine om å ha erga meg over eldre arkivmateriale — og for ein del òg eldre dialektlitteratur — som bruker slik transkripsjon at vi ikkje veit kva verdi og status lydskriftteikna har. Å finne fonologiske system i slikt materiale er nesten umoglig.

Kjem vi nå på at vi skal bruke det tilfeldig innsamla materialet vårt til ein fonologisk studie, oppdagar vi snart kor dårleg det er, dersom det består av bare "naturlig tale". Nettopp i fonologien er vi opplært til å drive ei form for eksperiment, både når vi skal etablere fonemsystem og finne allofonisk veksling. Sjølv om materialet vårt nå er omfangsrikt, kan det vere både *upålitelig* og *ufullstendig*.

Den eksperimentelle innstillinga som er heller sjølv sagt i fonologien, må vi òg prøve å overføre til andre delar av talemålsstudiet, til både morfologi, syntaks og sosial variasjon.

Det nye med sosiolingvistikken er interessa for variasjon i språket, og interessa for å tilfredsstillende statistiske krav til materialet, både om representativitet og statistisk signifikans. Men dette fører til ein enorm auke i materialet fordi med bare få klassifiseringskriterium blir gruppene eller sellene i ein undersøking svært mange. Idealet er ofte minst fire i kvar selle. Bare enkle problemstillinger fører lett til eit stort tal med informantar. Skal vi da bruke så tradisjonelle og urasjonelle innsamlingsmåtar som såkalla "naturlig samtale" for alle informantane, blir brått materialmengda uhorvelig stor. Kanskje alt også skal transkriberas? Slikt kan ta knekken på mange vitenskaplege assistentar. Og det som verre er: resultatet treng ikkje bli pålitelig og fullstendig.

Men slik har i alle fall den sosiolingvistiske interessa dei siste tiåra skapt fleire mammutprosjekt: Prosjekt med store medarbeidarstabar som arbeider i mange år.

6. Prosjektet "Talemål hos ungdom i Bergen"

Eg kan illustrere dette teoretiske utgangspunktet med å referere diskusjonar vi har hatt i Bergen, men som eg òg går utifrå er nett dei same som går igjen i andre fagmiljø.

Prosjektet "Talemål hos ungdom i Bergen" (TUB) har som mål å undersøke — som namnet seier — først og fremst talemålet hos unge, korleis det har forandra seg i forhold til det klassiske bergenske talemålet beskrive i monografien til Larsen og Stoltz ("Bergens bymål") frå 1911-12, og korleis bergensmålet spreier seg til dei tidlegare kommunane som nå er lagt inn i den nye storkommunen Bergen. Materialet er opptak med 104 informantar. Opptak i 68 timar og 15 minutt i alt.

Det blei i starten ført diskusjonar i det vide og det breie om transkripsjonen: Skal vi transkribere, og ev. korleis? Alle synsmåtar var framme, og prosjektleiinga vedtok at alt materialet skulle transkriberas. Men da det blei lagt fram tal som viste kor stort arbeidet eigentlig ville bli, blei vedtaket endra, slik at knapt 3/4 er blitt transkribert. I det ikkje-transkriberte materialet har vi gått inn og ekserpert direkte frå bandet.

Ein stor *praktisk* fordel med ekserpering direkte frå lydband i staden for å transkribere først er sjølv sagt at ein sparer mye tid. Men det *prinsipielt* viktige er at ei ekserpering *med* nødvendiggvis vere meir målretta, fordi ein må ha definert kva ein leitar etter. Ein må tenke på eit tidligare stadium. Avlyttinga kan da konsentrere seg om dei distinksjonane som er definert som interessante, og det utskrivne materialet blir dermed meir nøyaktig enn eit transkribert materiale som ikkje er utskrive målretta. Direkte ekserpering blir på den måten meir pålitelig.

For den delen som skulle transkriberas, blei det valt ei rimelig løysing, nemlig å skrive fonematisk. For denne dialekten kjente vi foneminventaret og stort sett verdiane. Men det blei transkribert subfonematisk for eit par fenomen, ma. for nyansar mellom palatale og palatoalveolare frikativar, pga. at språkutviklinga ser ut til å føre til samanfall mellom desse to gamle fonema.

Dermed er finare nyansar av interesse, og dei er definert og gitt egne teikn. Det store talemålmaterialet vårt kan dermed fortelje oss noe om *korleis* det foregår ei omlegging av *uttalen* av desse fonema. — Men det er også alt. Skal vi undersøke eit ev. *fonematisk* samanfall, er denne framgangsmåten *prinsipielt* galen. Da må vi lage egne eksperimentelle undersøkelser. For bare avlytting av band kan ikkje gi oss anna enn eit tilnærma rett bilde av uttalen: den kan ikkje seie oss noe om ev. *endå relevante* distinksjonar for språkbrukaren. Materialet vårt er gale innsamla, fordi det ikkje er laga eksperiment som høver til ei slik interessant problemstilling. Materialet er stort, men *ufullstendig* og *upålitelig*.

I prosjektet vårt i Bergen er det innlagt ein variabel som gjeld stil-, ariasjon, eller situasjonsbestemt variasjon. Det er tale om meir og mindre formell stil. Rett nok er ikkje analysen komme så langt ennå at vi kan legge fram endelige tal og resultat, men det ser nokså tydelig ut til at vi ikkje får noe utslag. Det er gjort opptak med 61 informantar i ulike situasjonar. Materialet er transkribert og ekserpert, dataført og korrekturlese med stort slitt.

Ej vurdering av resultatet kan så vere: Her har vi eit solid grunnlag for å påstå at ungdommen i Bergen *ikkje* har stilvariasjon i språket sitt.

Men kva vitenskaplig empirisk status har ein slik påstand? Han kan ikkje bevisas, bare avkreftas. Alle mislykka forsøk på avkrefting: styrker hypotesen, men han blir altså aldri bevist, for det kunne jo finnast stilvariasjon om korpuset var større! Og framfor alt: Det kan jo hende vi ikkje har greidd å lage situasjonane så godt vi burde, eller vi burde ha testa i *andre* situasjonar. I staden for å slite med eii kjempemateriale, burde vi ha laga meir presise hypotesar — som kanskje *kunne* avkreftas. Fann vi ikkje forskjellar etter å ha testa f.eks. 4 informantar, burde tidspunktet vere inne til å presisere hypotesen vår til f.eks.: det er ikkje forskjell på språket til gutar og jenter i *svært* jrmell stil (f.eks. telefonstil, foredragsstil eller radiointervju) og i svært uformell stil, og det er ikkje forskjell i språket om emnet er praktiske kvardagsspørsmål, fotball, seksuelle tabuområde, seriøse filosofiske spørsmål osv. Har vi greidd å avsanne noen slike hypotesar, veit vi straks noe meir. Greier vi å avsanne slike hypotesar med 10 informantar, vinn vi meir innsikt enn om vi slitt med 100 for å styrke ein annen *minde* presis hypotese!

Det eg nå har sagt, verkar svært negativt om forskinga som er gjort, både generelt og om prosjektet vårt i Bergen. Det betyr ikkje at prosjektet treng

vere så dårlig. Men vi skal lære av feila våre òg. Prosjektet vårt i Bergen skal gi like mange interessante resultat som mammutprosjektet før oss, f.eks. i Lund og Oslo. Og som naturlig er i forskingsprosessen, kan våre konklusjonar bli hypotesar i nye prosjekt. Poenget i denne sjølvkritikken er at opplegget har vore prega for sterkt av induktivisme, og det har vore ein lite effektiv måte å nå fram til vårt presisjonsnivå på. Kostnadene òg slitet har vore for store i forhold til vinninga.

7. Avslutning

Vi må revurdere arbeidet i store prosjekt, som er karakterisert av lite presis. Hypotesar i utgangspunktet, kjempeinsamling av materiale, fleire års åndelaust slit med materialet før ein i det heile kan få ut noe resultat, som til slutt gir bare liten auke i presisjonsnivået i hypotesane.

Det treng ikkje nødvendigvis vere noe gale med store prosjekt, men dei bør heller splittas opp i mange småe som heng saman. I Bergen burde vi konsentrert oss først om ei lita gruppe informantar, f.eks. bare ein bydel, gjort eit lite feltarbeid, bearbeidd materialet, laga nye og meir presise hypotesar, gjort nytt feltarbeid, bearbeidd det materialet osv. Det vil seie at vi kunne ha komme inn i ein sirkel der både vi sjølve og forskinga kunne stadig ha modnas. Prosjektet kunne ha produsert fleire spørsmål. Og det er det viktigaste.

I vitenskapen vår er det spørsmål, dvs. hypotesar, som det er mangel på, ikkje materiale. Og hypotesane er primære, for ulike hypotesar krev ulikt materiale. Materialet er fullstendig og påliteleg bare i forhold til hypotesane.

Med å samle inn mye materiale får vi vite *mindre* om språket fordi vi druknar i banalt slit. Med mindre materiale får vi vite *meir*, fordi vi da er nøyvde til å spekulere. Og det er spekulasjonane som fører oss framover.

Den sosiolingvistiske forskinga har gitt oss mange verdifulle spørsmål, ved at vi nå spør oss om språklig variasjon og om korleis eit språksamfunn eigenlig fungerer. Men i kombinasjon med nordiske filologar har vi fått ei slagside ved at vi har tatt med oss den induktive arbeidsmåten over til studiet av levande språk. Vi druknar i materiale. Dette er litt paradoksalt når eitt av forbildra, Labov, har gjort ein så elegant undersøkelse som å studere variasjon i +/- realisasjon av /r/ i utlyd og føre konsonant i New York-engelsk med å

springe gjennom fleire varehus og spørje ekspeditrisene om kva etasje han kunne finne det eller det vareslaget på, slik at ekspeditrisene var nøyvde å svare "the fourth floor". Utifrå dette materialet kunne han finne mange interessante variasjonar. Og materialet inneheldt ikkje noe overflødig.

Vi kunne ha dette som forbilde i arbeidet vårt i arkiv og prosjekt òg. Det stiller krav til oss ved at vi må bli meir oppfinnsame. Vi må utvikle repertoaret av innsamlingsmåtar til å omfatte meir enn bare spørjeliste og "naturlig tale". Vi må tenke før vi samlar. Når forskarar i ettertida vil kikke i arkiva våre, vil dei òg sette mest pris på materialet vårt om vi har arbeidd slik. Det er ikkje kvantitet dei har bruk for.