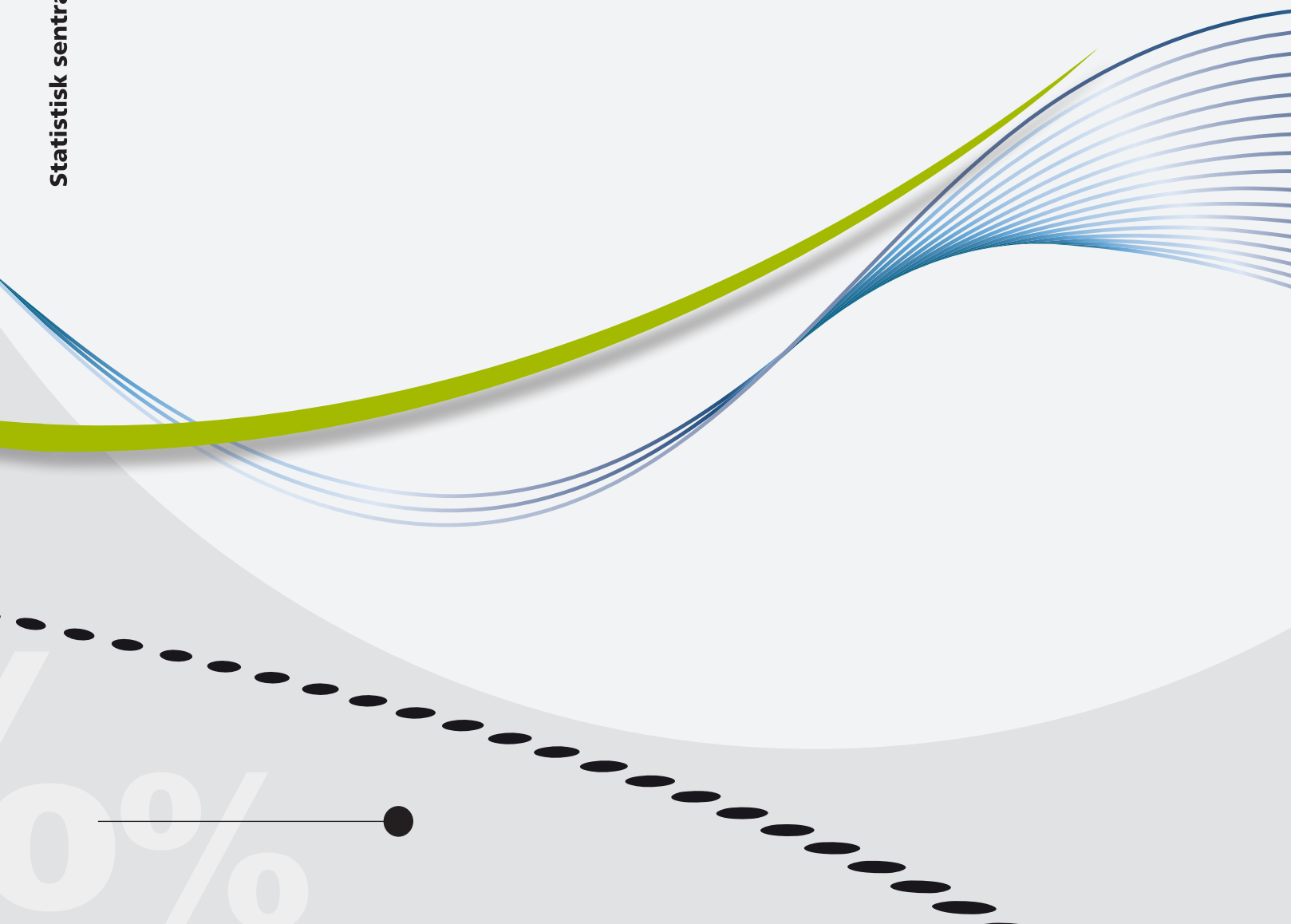




Bjørn Are Holth og Magnar Lillegård

Statistikk over samiske språkbrukere i Norge

En kartlegging av eksisterende datakilder og vurdering av fremgangsmåter for statistikk



Bjørn Are Holth og Magnar Lillegård

Statistikk over samiske språkbrukere i Norge

En kartlegging av eksisterende datakilder og
vurdering av fremgangsmåter for statistikk

I serien Notater publiseres dokumentasjon, metodebeskrivelser, modellbeskrivelser og standarder.

© Statistisk sentralbyrå
Ved bruk av materiale fra denne publikasjonen
skal Statistisk sentralbyrå oppgis som kilde.

Publisert 18. oktober 2017

ISBN 978-82-537-9605-5 (elektronisk)

Standardtegn i tabeller	Symbol
Tall kan ikke forekomme	.
Oppgave mangler	..
Oppgave mangler foreløpig	...
Tall kan ikke offentligjøres	:
Null	-
Mindre enn 0,5 av den brukte enheten	0
Mindre enn 0,05 av den brukte enheten	0,0
Foreløpig tall	*
Brudd i den loddrette serien	—
Brudd i den vannrette serien	
Desimaltegn	,

Forord

På oppdrag for Sametinget, har Statistisk sentralbyrå (SSB) påtatt seg å vurdere muligheter for å etablere statistikk over samiske språkbrukere i Norge.

Arbeidet med notatet er finansiert av Sametinget. Ved SSB har Bjørn Are Holth og Magnar Lillegård hatt ansvar for utredningen og utarbeiding av dette notatet som dokumenterer arbeidet.

Statistisk sentralbyrå, 5. oktober 2017

Jørn Leonhardsen

Sammendrag

Det finnes i dag ingen offentlig registrering av personer som snakker samisk i Norge, og dermed heller ingen offentlig statistikk over utbredelsen av de samiske språkene.

I dette notatet gjennomgår og vurderer SSB eksisterende datakilder som kan belyse utbredelsen av de samiske språkene, og hvordan de ulike kildene kan tenkes benyttet som grunnlag for etablering av fremtidig statistikk på området.

Metodikken som anbefales, er først å etablere en type statistisk populasjon over personer med høy sannsynlighet for enten å snakke samisk eller ha potensiale for å utvikle «sitt» samiske språk. En finner fram til denne statistiske samespråklige populasjonen ved å ta utgangspunkt i de største og beste foreliggende registre over personer av samisk ætt eller med forventet samisk tilknytning. Det samiske valgmanntallet, Folke- og bolig tellingen i 1970, Reindriftsforvaltningens register over personer med reinmerke og grunnskolenes register over personer med avgangskarakterer i samisk, er vurdert som de mest sentrale kildene.

Når alle relevante registerkilder er hentet inn, vil en ved å benytte Folkeregisteret som koblingsnøkkel, kunne supplere den statistiske populasjonen med familie- og slektsrelasjoner til disse personene og ende opp med et utvidet slektstre over personer med høy sannsynlighet for samisk tilknytning og språk.

En slik statistisk samiskspråklig populasjon vil i neste omgang kunne utnyttes til å trekke utvalg for spørreundersøkelser der tema er samisk språk og språkutvikling.

SSB har til slutt gitt noen foreløpige anbefalinger rundt hvordan en slik spørreundersøkelse kan utformes med sikte på å kartlegge utbredelsen av de samiske språkene i Norge. SSB mener at det vil være hensiktsmessig å gjennomføre en pilot i første omgang for å kartlegge de ulike registerkildenes faktiske kvalitet bedre. En pilot vil også gi viktige innspill til utforming og design for en fremtidig spørreundersøkelse om samisk språk, og dessuten gi oss nærmere svar på hvor god statistikk det vil være mulig å utvikle på grunnlag av en slik spørreundersøkelse.

Innhold

Forord	3
Sammendrag	4
Innhold	5
1. Bakgrunn	6
2. Hvordan måle språkferdigheter i befolkningen?	6
2.1. Metode/fremgangsmåte	8
2.2. Folkeregisteret og det samiske valgmanntallet	9
2.3. Folke- og boligtellingerne	11
2.4. Reindriftsforvaltningens register	13
2.5. Register over personer som har mottatt samisk språkopplæring	13
2.6. Medlemsregister i sameforeningene	14
2.7. Partilister til sametingsvalgene	14
2.8. Eventuelle andre kilder	14
3. Studier der etnisitet eller samisk språk er kartlagt	15
3.1. SAMINOR-undersøkelsene (2003-2004 og 2012-2014)	15
3.2. Finnmarksundersøkelsene	16
3.3. Ungdomshelseundersøkelsen i Nord-Norge, UHNN	16
3.4. Sør-Varanger-undersøkelsen	17
3.5. SLiCA-studien	17
3.6. Andre helseundersøkelser blant samer	18
4. Populasjon og utvalgsundersøkelse	18
4.1. Populasjonen	18
4.2. Utvalget	19
4.3. Frafall og utvalgsvekting	20
4.4. Spørreundersøkelse – design	21
5. Veien videre	22
5.1. Pilot	23
Referanser	24
Vedlegg A: Samisk språk. Guide med spørsmål til registereiere	25

1. Bakgrunn

Våren 2016 ble Statistisk sentralbyrå (SSB) kontaktet av Samisk språkråd med en forespørsel om SSB kunne vurdere muligheter for å kartlegge utbredelsen av samiske språkbrukere i Norge. Bakgrunnen for henvendelsen var Samisk språkråds arbeid i et offentlig oppnevnt utvalget satt til å utrede *lovverk, tiltak og ordninger for samiske språk* (NOU, 2016).

Det finnes ikke offisiell statistikk over hvem som snakker samisk her til lands, og utvalget ønsket derfor SSBs vurdering av om det kunne la seg gjøre å lage en slik statistikk over språkbrukere, og i tilfelle, hvordan det kunne gjøres.

Det var ikke mulig for SSB å gjennomføre en slik utredning innen fristen for utvalgets rapport, men *Sametinget* (deltaker i utvalget) ønsket allikevel å få SSBs vurdering av spørsmålet, og overtok derfor oppdragsgiverrollen ovenfor SSB.

Fra oppdragsgivers side ønsker man å kunne spesifisere hvilke av de samiske språkene det tales: nordsamisk, lulesamisk, sørsamisk, pitesamisk, østsamisk, og eventuelt andre. Nivået på språkferdighetene er også en sentral dimensjon. Herunder ligger også ønske om at en fremtidig statistikk skal inkludere personer med et *potensiale* for å utvikle sine språkferdigheter i et av de samiske språkene. Det vil typisk kunne være personer av samisk opprinnelse som selv, eller i tidligere familieslektsledd, har mistet sitt samiske «morsmål».

Det er også ønske om å kunne gi statistikk på kommunenivå, gjerne brutt ned på hensiktsmessige alderskategorier. Videre er det uttrykt et ønske om å kunne måle utviklingen i de samiske språkene over tid. Målinger hvert 4. år har vært antydning som et passende intervall.

Det er her viktig å understreke at SSB gjennom denne utredningen ikke tar stilling til om en i fremtiden skal gjennomføre en slik undersøkelse av samiske språk. Heller ikke om det skal eller bør etableres løpende statistikk på området. Utredningen har kun som siktemål å vurdere gjennomførbarheten, sammen med noen forslag til hvordan det kan la seg gjøre. Beslutninger om eventuell gjennomføring må komme på et senere tidspunkt, og vurderes i sammenheng med finansiering av en slik statistikk.

2. Hvordan måle språkferdigheter i befolkningen?

Det er først nødvendig å klargjøre noen premisser for en statistikk over språkbruk og -ferdigheter. Det å beherske et språk på et visst nivå, her samisk, er en egenskap som best kan kartlegges ved at språkbruker selv gjør en egen vurdering av sine språkferdigheter. En kan selvsagt også få informasjon om språkferdigheter gjennom avlagte eksamener o.l., men dette vil kun gjelde en mindre andel av språkpraktiserende.¹ Det er flere måter språkbruker kan gi denne informasjonen på, og dette vil vi også diskutere som del av oppdraget. Her er poenget at vi kun kan få hele språkbildet ved å spørre befolkningen om deres språkferdigheter. En nærliggende fremgangsmåte når en ønsker å få svar på slike egenskaper i hele befolkningen, er å spørre et representativt utvalg av den norske befolkningen, gjerne stratifisert på de bakgrunnsvariablene en ønsker seg som output, f.eks. kommune og alderskategorier, som her. En slik undersøkelse ville imidlertid bli svært ressurskrevende, og dessuten også være befengt med betydelige utfordringer både når det gjelder utvalgs- og målefeil. Siden egenskapene en er ute etter, språkferdigheter i samiske språk, har en forholdsvis lav utbredelse i befolkningen,

¹ I kapittel 2.5 vil vi også undersøke hvilke muligheter som finnes for kartlegging av språkferdigheter gjennom grunnskolen.

ville de aktuelle utvalgsstørrelsene som måtte benyttes, måtte bli svært store, gitt ønsket om detaljeringsgrad en har her (geografi, skille mellom flere samiske språk, alderskategorier). For et stort flertall av de respondentene en kontaktet ville dessuten tema for undersøkelsen (ferdigheter i samisk) ikke være relevant. Irrelevante spørsmål for en stor andel av de spurte, er alltid et dårlig utgangspunkt for oppslutning om en spørreundersøkelse, i tillegg til den manglende respekten for folks tid det ville innebære (oppgavebyrde). I praksis ville derfor et slikt design være lite aktuelt.²

Hvis vi skal slippe å spørre et representativt utvalg av *hele* den norske befolkningen, er eneste farbare vei å forsøke å avgrense målpopulasjonen til de personer som enten snakker samisk, eller som vi antar har et stort potensiale for å utvikle samiske språkferdigheter. Avgrensningen må foretas uten at betydelige andeler av samisktalende faller utenfor målpopulasjonen. Desto mer presist vi makter å kartlegge disse reelle og potensielle språkbrukerne i en slik *første fase* uten at for mange i målgruppen faller utenfor, desto lavere utvalgsstørrelser er det nødvendig å trekke ut for videre spørsmål om språk i den *andre fasen* (spørreundersøkelsen).

Hvilke personer er det så som har en stor sannsynlighet for enten å snakke samisk eller ha potensiale for å utvikle sitt samiske språk? Samisk etnisk bakgrunn vil selvsagt være sentralt. De aller fleste som snakker samisk vil også ha samisk familiebakgrunn. De betyr ikke nødvendigvis at en på direkte spørsmål vil svare at en oppfatter seg som same, men samisktalende vil allikevel oftest komme fra familier med samisk opprinnelse. Vi vet også at det finnes samisktalende personer helt uten kjent samisk opphav/ætt. Disse vil gjerne ha lært seg noe samisk fordi det har vært hensiktsmessig i kommunikasjonsøyemed, f.eks. knyttet til arbeid innenfor reindriften. Men dette vil neppe dreie seg om særlig mange personer. Siden disse få samisktalende personene uten noen samisk familietilknytning ikke vil utgjøre noen vesentlig andel av alle samisktalende, vil det ikke være regningssvarende å bruke mye ressurser på å kartlegge dem. Imidlertid vil en god del av dem allikevel komme med, hvis en i geografiske områder der andelen samer er høy, for eksempel rundt ti prosent, velger å inkludere alle personer i målpopulasjonen. Ved å trekke utvalget fra alle personer i det geografiske området, vil en også få med samisktalende uten samiske forfedre. Mer om dette senere.

På den annen side vet vi at det er mange samer som selv ikke snakker samisk. For å overleve over tid, må et språk holdes vedlike gjennom praktisering. Det være seg over generasjoner i en befolkningsgruppe eller gjennom en persons livsløp. Når vi skal forsøke å finne utbredelsen av samiske språkferdigheter i Norge, holder det ikke å gå ut fra at en person eller gruppe personer behersker samisk fordi de har en viss etnisk bakgrunn, eller bor i et visst geografisk område. Språkferdigheter må kartlegges på individnivå ved en eller annen type måling. I skolen gjøres dette gjennom tester og eksamener. Spørreundersøkelser er også en type måling, men i stedet for å teste faktiske ferdigheter, ber vi personen selv ta stilling til, og vurdere det vi ønsker svar på. Ved å spørre personen selv, vil vi ikke bare kunne få kunnskaper om deres ferdigheter i samisk, men også få svar på andre spørsmål om språket deres som vi mener er viktige for å belyse utviklingen av de samiske språkene. F.eks. vil vi kunne be personen svare på spørsmål om hvordan de har tilegnet seg språkferdighetene, hvilke språk som foreldre og besteforeldre snakker, og hva de mener skal til for at de kunne utvikle sitt samiske språk til et høyere nivå.

² Spørsmål om samisk språk ville allikevel kunne være aktuelt i en tenkt større kartlegging av hele befolkningens språkferdigheter. Siden en da ville spørre respondentene om ferdigheter i alle språk de behersket, ville temaet dermed være relevant for alle. Denne tematikken har tidligere vært diskutert av SSB i notatet, *Minoritetsspråk i Norge* (Wilhelmsen m.fl. 2013).

Når vi skal forsøke å ringe inn de personene i Norge som vi ønsker å stille spørsmål om deres forhold og praktisering av samisk språk, spiller det derfor mindre rolle om de faktisk snakker samisk eller ikke. Det viktige er at slike spørsmål er relevante for dem. Enten ved at en eller flere av deres foreldre/besteforeldre er samer, eller fordi de lever i et geografisk område med stor andel samer. Det siste premisset kan vise seg å være en utfordring for oss når det gjelder hvor relevant spørsmål om samisk språk faktisk vil oppfattes. Selv om en lever i områder med stor andel samer, og dermed som sådan forstår hvorfor en får spørsmål om samisk språk, kan det likevel tenkes at en ikke vil besvare undersøkelsen fordi en hverken regner seg som same, eller snakker samisk, og dermed går ut fra at undersøkelsen ikke gjelder dem. Vi ender dermed opp med stort *frafall* i undersøkelsen ved at mange avstår fra å svare. Problemet for oss er at vi ikke kan vite hva mangel på svar skyldes. Det kan være at de ikke snakker samisk eller har ønske om det, at de ikke har noen samisk bakgrunn, eller det kan være fordi de ikke ønsker å «flagge» sin samiske herkomst. Utfordringen med høyt *frafall* må vi ta hensyn til i det videre arbeidet med å designe en undersøkelse som tar sikte på å kartlegge samisk språk.

Så langt har vi dermed kommet fram til at vi på en eller annen måte må forsøke å avgrense utvalget vårt mest mulig til de personene der spørsmål om samisk språk er relevant, enten fordi de praktiserer språket selv, eller fordi de har forfedre og slekt som snakker eller snakket samisk, mens de selv ikke gjør det. Samisk opprinnelse blir derfor helt sentralt i avgrensningen.

I Norge har vi imidlertid ikke lenger noe register over personer av samisk etnisk bakgrunn. Sist gang det ble spurt om etnisitet i folketellingene var i 1970, da det ble spurt etter både samisk språk hos personen selv og dennes slekt, samt egen vurdering av om en oppfattet seg som samisk (Aubert 1978). Disse spørsmålene gikk imidlertid kun til befolkningen i enkelte områder i de nordlige fylkene i landet, der en fra før visste at det var en betydelig andel av samer. Landet for øvrig fikk ingen slike spørsmål. FoB70 var den siste i sitt slag der det ble spurt etter samisk etnisitet. Mer om folketellingene under kapittel 2.3.

Selv om det ikke er foretatt offisielle tellinger av den samiske befolkningen etter 1970, finnes det allikevel en rekke ulike kilder over personer med samisk bakgrunn og/eller samisk språk som vi kan dra nytte av i vår leting etter en samisktalende populasjon. Disse skal vi gjennomgå i det videre arbeidet. Ingen av disse kildene vil alene være omfattende og gode nok til å kunne beregne antallet samer i Norge, langt mindre hvor mange som taler språket. Imidlertid tror vi, og vil her argumentere for, at de ulike kildene tilsammen, vil kunne danne et ganske brukbart utgangspunkt for å etablere en type *statistisk samiskspråklig* populasjon, som vi i neste omgang kan spørre nærmere om deres praktisering av det samiske språket.

En forutsetning er imidlertid at kildene foreligger på individnivå. I praksis må registrene inneholde fødselsnummer på enhetene for at vi skal kunne innlemme dem i dette populasjonsarbeidet. Det vil allikevel være mulig å dra nytte av kilder der enhetene ikke er registrert med fødselsnummer, enten til å forbedre stratifiseringen i utvalgsarbeidet, eller til å øke presisjonen i tallene vi sitter igjen med etter en gjennomført spørreundersøkelse.

2.1. Metode/fremgangsmåte

For å kartlegge de foreliggende kildene som vi kan dra nytte av i vårt arbeid med en statistisk samiskspråklig populasjon, har metoden gått ut på å innhente mest mulig relevant informasjon om deres beskaffenhet, og muligheter for tilgang til kildene for SSB. Vi har benyttet litt ulike fremgangsmåter overfor de ulike dataeierne. For noen av dem har vi funnet tilstrekkelig svar gjennom

organisasjonens egne hjemmesider på nettet. For andre har det vært nødvendig å kontakte dem for nærmere avklaringer. Vi har både sendt e-post og benyttet telefon for å få svar på våre spørsmål. Sametinget har også bistått oss for enkelte av datakildene. Vedlegg A viser den guiden som ble utviklet for å bistå Sametinget med kildene de hadde ansvar for. Spørsmålene er i hovedsak de samme som SSB tok sikte på å få svar på i vår kartlegging. De ulike spørsmålene vil ha forskjellig relevans, avhengig av type kilde.

Vi har fått svar på spørsmålene vi hadde knyttet til registrene på litt ulike måter. For flere av kildene har det vært enkelt å hente inn denne informasjonen gjennom deres nettsider og/eller fra publikasjoner fra dem, som dokumenterer dataenes beskaffenhet. For andre kilder har det også vært nødvendig å kontakte registreierne/ansvarlige/dataeiere direkte for å få avklart enkelte av våre spørsmål. Sametinget har dessuten hjulpet SSB med å innhente informasjon fra noen av kildene.

Nedenfor vil vi gjennomgå de mest hensiktsmessige kildene for vårt formål, og skissere kort hvordan vi ser for oss at vi kan utnytte dem. Vi starter med Folkeregisteret, som vil være vår nøkkel til å kartlegge samiske språkbrukere.

2.2. Folkeregisteret og det samiske valgmanntallet

Gjennom det 11-sifrede fødselsnummeret gir Folkeregisteret oss oversikt over alle personer som er, eller har vært, bosatt i Norge. Fødselsnummeret vil være nøkkelen som gjør det mulig for oss å koble sammen opplysninger fra øvrige register, forutsatt at fødselsnummer er benyttet som identifikasjonsmerke. Som vi skal se, er ikke det alltid tilfelle. Noen har kun registrert person med navn, adresse og eventuelt fødselsdato.

Skatteetaten, som eier og er ansvarlig for Folkeregisteret, får også jevnlig tilsendt oppdaterte lister over det samiske valgmanntallet, slik at personer med rett til å stemme ved sametingsvalget, er registrert med et merke om dette i Folkeregisteret. Valgmanntallet vil være vår grunnstamme for etablering av en samiskspråklig populasjon, siden samisk språk også er en del av kriteriet for å registrere seg i manntallet. La oss derfor tittle litt nærmere på det samiske valgmanntallet og kravene som stilles for å bli registrert.

Sametingets valgmanntall

Sametingets valgmanntall er som nevnt et register over personer som har mulighet til å stemme eller stille som kandidat i valg til Sametinget (Sametinget.no 2017). Kriteriet for å kunne melde seg inn er at man oppfatter seg som same, og i tillegg enten:

- a) har samisk som hjemmespråk, eller
- b) har eller har hatt forelder, besteforelder eller oldeforelder med samisk som hjemmespråk, eller
- c) er barn av person som står eller har stått i manntallet.

I tillegg må man i likhet med det ordinære manntallet, være over 18 år eller fylle 18 år i valgåret.

Begge de grunnleggende kriteriene, oppfatning av egen etnisitet, og språk for en selv eller sine foreldre/besteforeldre/oldeforeldre, er altså basert på egenerklæring.

Som vi allerede har vært inne på, finnes det ingen registre som kan verifisere disse opplysningene.³

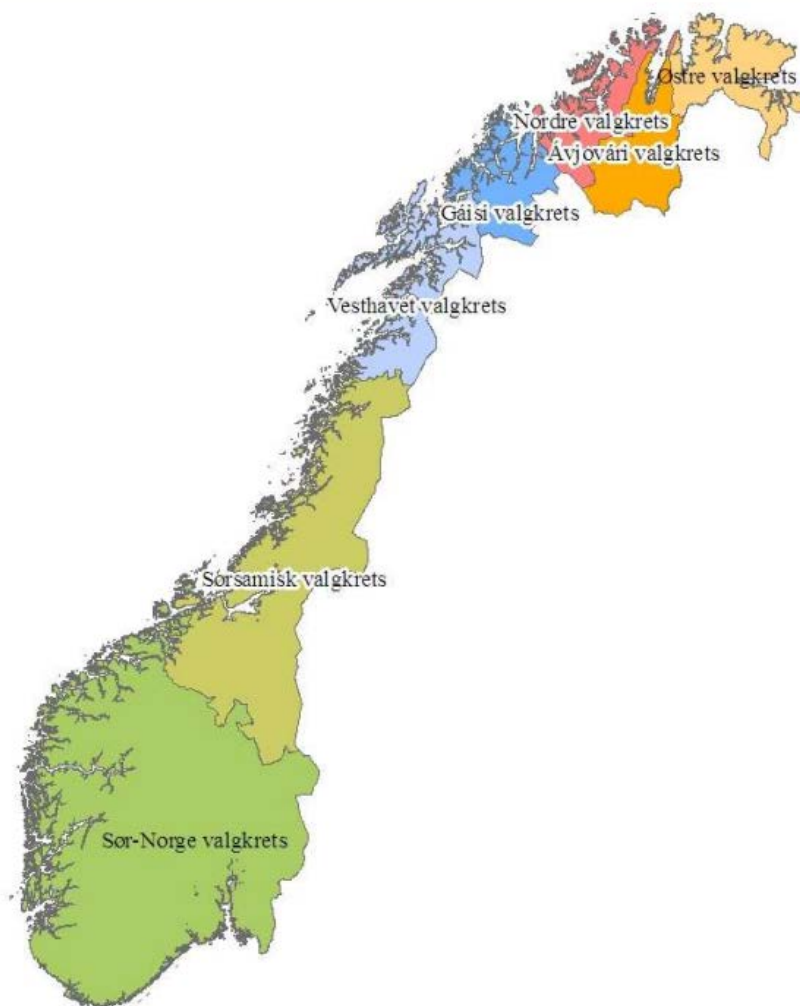
Per september 2017 består det samiske valgmanntallet av til sammen 16 958 personer som har rett til å stemme ved sametingsvalget 2017 (se tabell 2.1). De stemmeberettigede er fordelt på de sju valgkretsene på denne måten, ifølge Sametinget:

Tabell 1 Stemmeberettigede etter valgkrets. Sametingsvalget 2017

1	Østre valgkrets	2 303
2	Ávjovári valgkrets	3 738
3	Nordre valgkrets	2 540
4	Gáisi valgkrets	2 547
5	Vesthavet valgkrets	1 788
6	Sørsamisk valgkrets	1 148
7	Sør-Norge valgkrets	2 894
	Totalt	16 958

Valgkretsenes geografiske plassering er vist i Figur 2.1.

Figur 2.1. Kart over sametingets valgkretser



³ Det understrekes her at SSB ikke har noen indikasjoner på at personer melder seg inn i valgmanntallet på falskt grunnlag. Det er dessuten mulig å melde fra dersom man mener at en person feilaktig er oppført i manntallet (Sametinget.no 2017).

Det er viktig for vårt formål at det samiske valgmanntallet har som sentralt kriterium at en selv, eller en av ens forfedre har snakket samisk. En knytter altså samisk identitet og rettighet til å stemme, tett sammen med språket. Dette er noe vi har stor nytte av i vår søken etter kilder for samiske språkbrukere i Norge. Valgmanntallet gir oss kjennskap til personer som enten snakker samisk, eller har slekt som snakker samisk. De sistnevnte er like interessante for formålet vårt, i og med at de er aktuelle for spørsmål om hvorfor de selv ikke snakker samisk, og hva som eventuelt skal til for at de vil lære seg språket som deres forfedre praktiserte.

Selv om disse nær 17 000 personene er en god start for vår samespråklige populasjon, er vi på langt nær i mål med disse. Det finnes mange personer som taler samisk eller har forfedre som talte samisk, og som ikke står oppført i dette frivillige registeret. Dessuten omfatter det bare personer over 18 år med stemmerett.

Det er her Folkeregisteret kan benyttes som koblingsnøkkel. I Folkeregisteret vil vi kunne supplere populasjonen vår med familie- og slektsrelasjoner til personer registrert i valgmanntallet, men også supplere med andre individbaserte datakilder der samer eller samisk språk er inkludert. Denne fremgangsmåten er tidligere lansert av SSB som en mulig farbar vei for å forbedre det generelle datagrunnlaget for statistikk over samer og samiske forhold (Severeide 2008). Det ble den gang pekt på problemer med tilgang til datakildene for denne type individbasert statistikk. I den grad det har latt seg gjøre, har vi i denne utredningen derfor hatt særlig fokus på å kartlegge forutsetningene for tilgang til kildene.

Når det gjelder det juridiske grunnlaget, har SSB hjemmel i Statistikkloven, både for tilgang til foreliggende register over individdata, og for å benytte Folkeregisteret som koblingsnøkkel for å etablere en slik utvidet statistisk populasjon. Imidlertid er det enkelte begrensninger i dette grunnlaget som vi skal se. Tilgang, og mulighet for å utnytte data for vårt formål, vil vi derfor kommentere særlig under vår gjennomgang av de ulike datakildene nedenfor.

Modernisering av Folkeregisteret

Det er også verd å merke seg at det for tiden pågår et arbeid med å utarbeide nytt folkeregister. Gjennom informasjonsmodellen for revisjonsprosjektet, er det tatt høyde for å inkludere variabelen: «bruker samisk språk», og med tre verdier, *nordsamisk*, *lulesamisk*, *sørsamisk*. Endelig status for hvordan dette blir seende ut var ikke tilgjengelig for oss når denne rapporten ble avsluttet, men det er et arbeid med høy relevans for utarbeiding av statistikk over samiske språkbrukere, og som bør følges tett. Det er også viktig å være klar over det sentrale spørsmålet om hvordan informasjon på denne språkvariabelen i et modernisert Folkeregister eventuelt skal hentes inn. Hvis en tar som utgangspunkt at språkbruker selv frivillig oppgir om en benytter samisk, vil det allikevel være langt igjen før informasjonen kan utnyttes statistisk som en kilde over samisktalende.

2.3. Folke- og boligtellingerne

I folke- og boligtellingerne har det historisk vært inkludert enkelte spørsmål om samer og kvener. Ved alle folketellingene fra 1845 til 1930 ble antallet samer og kvener registrert (Severeide 2008). Ved tellingen i 1865 ble det også spurt om språk, og for alle tellinger mellom 1890 og 1930 ble det spurt både om avstamning og om språk.

Folketellingene i 1946 var den første etter krigen. Forestillingen om «rase» hadde av kjente grunner falt sterkt i miskreditt, og en fant det derfor ikke tjenlig å ta med spørsmål om samer og kvener i denne tellingen. I forkant av tellingen i 1950 ble SSB fra flere hold oppfordret om å gjenoppta tellingen av samer og kvener. Som et

kompromiss besluttet SSB kun å innhente opplysninger om språkforholdene i de tre nordligste fylkene.

Folke- og boligtellingsen i 1970 var siste gang spørsmål om etnisitet og språk ble inkludert i tellingene. Denne gangen ble det gjennomført en tilleggsundersøkelse i en del utvalgte tellekretser i Nordland, Troms og Finnmark. Formålet var å gi statistikk over den samiske tilknytningen til de totalt 113 874 personene som bodde i kretsene (Aubert 1978). Tilleggsundersøkelsen inneholdt de fire spørsmålene:

1. **Var samisk det første språk som personen snakket?** Svaralternativ: **Ja/Nei.** *For barn som ennå ikke har lært å snakke, skal det settes kryss for Ja dersom samisk antas å bli det første talespråket.*
2. **Var samisk det første språk som minst én av personens foreldre snakket?** Svaralternativ: **Ja/Nei/Vet ikke.**
3. **Var samisk det første språk som minst én av personens besteforeldre snakket?** Svaralternativ: **Ja/Nei/Vet ikke.**
4. **Regner personen seg selv som same?** Svaralternativ: **Ja/Nei/Usikker/Ønsker ikke å svare.** *Foreldre eller andre foresatte avgjør om barn under 15 år skal regnes som same.*

Vilhelm Aubert analyserte statistikken fra tilleggsundersøkelsen som et oppdrag på vegne av SSB. Resultatene ble publisert under tittelen «Den samiske befolkningen i Nord-Norge» (Aubert 1978). Aubert forsøkte her å avgrense og beskrive den samiske befolkningen, i tillegg til å sammenlikne den samiske befolkning i ulike deler av Nord-Norge. Her vil vi konsentrere oss om funn og analysene av språkforholdene.

Totalt var det 10 535 personer som svarte at de hadde samisk som første språk, mens 16 808 personer svarte at samisk var første språk som minst én av personens foreldre snakket. 19 635 personer svarte at samisk var det første språket som minst én av personens besteforeldre snakket.

Ved å analysere forskjeller i svarene mellom de fire spørsmålene, og sammenstille med tidligere tellinger i 1930 og 1950, forsøker Aubert også å beskrive ulikheter i hvordan språksituasjonen i de nordlige fylker og kommuner har utviklet seg. Han finner bl.a. støtte i tallene for at fornorskingsprosessen slår betydelig svakere ut i områder der samene utgjør en større andel av befolkningen, og som på den måten gjorde det mulig for samisk (språk)kultur og identitet å stå bedre imot det ytre presset fra storsamfunnet.

Dette er interessant lesning og metodikken er kreativ. Det er imidlertid ikke rom for å gå nærmere inn på disse analysene her, men vi bør ta med oss at det kan være gode grunner for å inkludere de samme spørsmålene (nevnt over) hvis det skulle bli aktuelt å gjennomføre en ny spørreundersøkelse om samisk språk. På den måten vil en trolig kunne få mye nyttig informasjon ved å sammenstille svarene før og nå for de utvalgte områdene som mottok tilleggsundersøkelsen i 1970.

Når det gjelder utnytting av data, er det også gledelig å vite at SSB fremdeles har tilgang til de individuelle svarene fra tellingen i 1970, registrert med fødselsnummer. Disse kan derfor kobles mot Folkeregisteret og utgjøre et viktig supplement til vår statistiske samespråklige populasjon. Disse dataene gir i tillegg direkte svar på spørsmål om språkkompetanse for svarpersonen selv, samt for eventuelle barn, foreldre, besteforeldre og oldeforeldre. Dette er nyttig tilleggsinformasjon som kan utnyttes både i det videre populasjonsarbeidet, utvalgsarbeidet, og i fremtidig analyseformål, som antydnet over.

2.4. Reindriftsforvaltningens register

Det er kun samer som har rett til å drive med reindrift i Norge. En forutsetning for å bli registrert med såkalt reinmerke, og dermed rett til å drive med reindrift, er at enten foreldre eller besteforeldre har drevet med reindrift. Det er Reindriftsforvaltningen som er satt til å føre register over personer tilknyttet reindriften. Det er mulig for ikke-samer å få tildelt reinmerke, men kun gjennom inngifte i en reindriftsfamilie. Per 5. april 2017 er det registrert 3 185 personer med rett til å drive reindrift i Norge (Meld.St.32 2016-2017).

Reinmerkeregisteret er et offentlig register der personene er registrert med fødselsnummer. SSB har tilgang til registeret, og vi vil kunne utnytte det som supplement til vår samiskspråklige statistikkpopulasjon.

2.5. Register over personer som har mottatt samisk språkopplæring

Elever som har mottatt språkopplæring i grunnskolen vil være en god kilde til å belyse samespråklige forhold og gi en god pekepinn på hvordan utbredelsen av de samiske språkene står. Fra grunnskolens informasjonssystem (GSI) kan vi ta ut statistikk over elever som har fått opplæring i samisk. Vi kan se antallet elever fordelt på årstrinn (1.–10.) og fordelt på de tre samiske språkene, nordsamisk, lulesamisk og sørsamisk. Vi kan også se om opplæringen har vært gitt som førstespråk eller andrespråk. For skoleåret 2016/17 var det til sammen 814 elever her til lands som mottok opplæring i et av de samiske språkene (GSI 2017). Dette ut av et samlet elevtall på 629 275. Vi skal ikke gå nærmere inn på den geografiske fordelingen og hvordan opplæringen fordeler seg på de tre offisielle samiske språkene. For vårt formål er det i denne omgang tilstrekkelig å vite at fordelingen finnes og vil være nyttig for det videre arbeidet.

Imidlertid er det en viktig begrensning i disse tallene: De foreligger ikke som individdata, dvs. at de ikke inneholder fødselsnummer. Vi vil derfor ikke kunne inkludere hele dette materialet i vår statistiske samisktalende populasjon.

Av individdata på utdanningsområdet er det avgangskarakterer i samisk fra avsluttet grunnskole (10. årstrinn) som vil være den beste kilden. I tillegg foreligger det individdata for elever som har avlagt nasjonal prøve på samisk (5., 8. og 9. trinn). For skoleåret 2015/16 er det 203 elever som er registrert med standpunkt eller eksamenskarakter i et av de samiske språkene, mens det for skoleåret 2016/17 er 219 elever som avla nasjonal prøve på samisk (GSI 2017).

Ved også å gå tilbake i tid, vil vi gjennom disse kildene få oversikt over et stort antall elever som har avlagt eksamen på samisk i grunnskolen. Vi har ikke tatt stilling til hvor langt tilbake i tid vi vil innhente data fra, men det er naturlig å gå så langt tilbake som vi har tall av noenlunde god kvalitet. Vi har ikke gått nærmere inn på kvaliteten bakover i tid i denne utredningen.

Avgangskarakterer og eksamener i samisk fra grunnskolen vil være en meget viktig og god kilde for vårt formål. I tillegg til å gi oss direkte tilgang til personer som snakker samisk på et visst nivå, har vi i tillegg også mulighet til å få en pekepinn på nivået på deres språkferdigheter gjennom karakterene de har fått og om de har deltatt i undervisning i samisk som første- eller andrespråk.

Statistikkloven gir SSB anledning til å utnytte denne type utdanningsdata på individnivå.

Samiske språksentre

Det gis undervisning i samisk ved de ulike samiske språksentrene. Problemet her er at deltakerne i liten grad registreres med fødselsnummer og at registreringen for øvrig er av litt varierende kvalitet, i alle fall for vårt statistikkformål. Et unntak er E-skuvla (navn på språksenter) som årlig sender inn lister til SSB med navn og fødselsnummer på alle deltagerne. Listene gir grunnlag for refusjon for deltagere som har gått på språkkurs, og bør derfor være av god kvalitet.

Sametinget vil også ha mulighet til å gi tilgang til lister over personer som har mottatt stipend, både til videregående opplæring og høyere utdanning. På videregående får de tilskudd til samiskundervisning, og på høyere utdanning er kravet at de er samisktalende. Antallet som mottar stipend til høyere utdanning ligger i størrelsesorden 150 personer pr år, får vi opplyst av Sametinget. Stipend til videregående opplæring i omtrent samme omfang.

2.6. Medlemsregister i sameforeningene

Denne kilden viste seg å være noe mer krevende å få klarhet i gjennom denne utredningen. På henvendelsen fra Sametinget og SSB, svarte de fleste av registereierne (sameforeningene), at før de kunne ta stilling til eventuell utlevering av data, måtte dette tas opp i deres respektive styrer. Siden vi i denne utredningen kun ønsket å sondere terrenget mht. hvilke kilder vi ville kunne trekke veksel på, ble det ikke fra SSBs side bedt om en formell utlevering av register. Det får eventuelt komme ved en senere anledning hvis det blir aktuelt å gjennomføre en undersøkelse.

Vi har imidlertid avdekket at mange av disse foreningene ikke har registrert medlemmene med fødselsnummer, men kun med navn, fødselsår og eventuelt adresse. Dermed er det ikke like enkelt å utnytte registrene direkte i oppbyggingen av vår samiskspråklige populasjon. Det vil kanskje være mulig å utlede fødselsnummer ved hjelp av de foreliggende personopplysningene. Dette må vi vurdere nærmere når vi eventuelt får tilgang til registrene og ser hvilke variabler som er registrert.

Det er nok også grunn til å spekulere rundt om ikke de fleste av de som er medlemmer i en sameforening, også vil være de samme personene som vi fra før har funnet fram til gjennom de registrene vi tar utgangspunkt i (det samiske valgmannstallet, folketellingen fra 1970, reindrifregisteret, utdanningsregisterne o.a.). Særlig hvis vi også inkluderer de personene vi identifiserer etter å ha foretatt påkobling av personer med familie- og slektskap til de vi finner i registrene.

2.7. Partilister til sametingsvalgene

Personer som stiller til sametingsvalgene vil sannsynligvis også selv være registrert som stemmeberettigede, og vil således fremkomme gjennom det samiske valgmannstallet. Vi har derfor ikke gjort noen ytterligere forsøk på å undersøke utnytting av dette registeret.

2.8. Eventuelle andre kilder

Det finnes sikkert en del flere kilder enn vi her har nevnt som kan tenkes å ha relevante personopplysninger om enten samisk bakgrunn eller språk. Abonnenter på samiske tidsskrifter og aviser har vært nevnt som en slik kilde. Når en skal ta stilling til hvor mange slike mer perifere register en skal søke om tilgang til, er det viktig å veie nytten nøye opp mot ressursbruken. I den grad informasjonen er av bra kvalitet og en kan få enkel tilgang, er det selvsagt hensiktsmessig å inkludere slike kilder. Hvilke kilder en konkret ender opp med å søke tilgang til, må vi ta stilling til hvis det eventuelt blir aktuelt å gjennomføre en spørreundersøkelse.

3. Studier der etnisitet eller samisk språk er kartlagt

3.1. SAMINOR-undersøkelsene (2003-2004 og 2012-2014)

SAMINOR-undersøkelsene er et omfattende forskningsprosjekt som ble gjennomført som et samarbeid mellom Senter for Samisk Helseforskning og Folkehelseinstituttet. Prosjektet ble utført i to runder, den første i 2003-2004, og den andre i 2012-2014. Et viktig siktemål for prosjektet var å undersøke samiske helse- og levekårsforhold, og en inkluderte derfor områder hvor den samiske befolkningsandelen var rundt 10 prosent, basert på folketellingen i 1970 (Severeide 2008). For SAMINOR 1 ble utvalgte kommuner i Finnmark, Troms, Nordland og Trøndelag valgt ut. Enkelte kommuner der andelen samer (igjen basert på FoB 1970) var en del lavere enn 10 prosent, ble også tatt med. Aldersmessig rettet studien seg mot alle 30-åringene, og personer i alderen 36-79 år. «Same» ble operasjonalisert som «minst én besteforelder som snakket samisk», basert på svarene fra FoB1970. (Andersdatter Siri 2015).

SAMINOR 2 ble gjennomført i det samme geografiske området som den første studien. I tillegg ble Sør-Varanger kommune inkludert. Aldersgruppen studien rettet seg mot denne gangen var 18-69 år.

For vårt formål er SAMINOR-undersøkelsene nyttige fordi det ble inkludert spørsmål om egen og foreldre/besteforeldres språk, i tillegg til spørsmål om egen og foreldres etniske bakgrunn. Det ble også spurt om en selv «regnet seg som same». De samme spørsmålene rundt etnisitet og språk ble benyttet i Saminor 1 og 2, som vist i Figur 2. I de områdene der undersøkelsen ble gjennomført (og andelen samer var høyest) gir SAMINOR oss positiv identifikasjon av et betydelig antall samer og samisktalende.⁴

Resultatene fra disse undersøkelsene foreligger som individdata, dvs. med fødselsnummer. Dermed ville det vært nyttig om vi kunne bruke disse direkte i vår oppbygging av en statistisk samisktalende populasjon. Imidlertid vil ikke dette la seg gjøre, i og med at det dreier seg om en frivillig undersøkelse, der respondentene har gitt sitt samtykke til deltagelse basert på en viss bruk av resultatene, og slik denne bruken er kommunisert i forkant til respondentene. Det er naturlig nok ikke beskrevet en bruk i retning av å utnytte svarene til å opprette en samisk språkpopulasjon mange år senere. SSB har tidligere søkt om tilgang til mikrodata fra SAMINOR, og fått avslag (Severeide 2008). I forbindelse med en eventuell fremtidig undersøkelse i SSBs regi rundt utbredelse av samiske språk, er det selvsagt mulig at en på ny kan søke om tilgang til disse personopplysningene, men resultatet vil trolig bli det samme. SSBs tilgang til foreliggende register, slik dette er hjemlet i Statistikkloven, omfatter normalt ikke svar gitt i slike frivillige spørreundersøkelser når bruken fraviker beskrivelsen som er formidlet til respondenten i forbindelse med deres samtykke. Kun i helt spesielle tilfelle kan Datatilsynet fravike dette grunnprinsippet og gi særskilt dispensasjon for annet bruk.

Resultatene fra undersøkelsen vil imidlertid kunne utnyttes som statistisk grunnlag i stratifiserings- og analysefasen av en fremtidig undersøkelse av samiske språk.

⁴ Fra SAMINOR 1 ble 5 867 personer identifisert med en samisk tilknytning (Severeide 2008).

Figur 3.1. Spørsmål om etnisitet og språk fra SAMINOR-undersøkelsene

Hvilket hjemmespråk har/hadde du, dine foreldre og besteforeldre? (sett ett eller flere kryss)				
	Norsk	Samisk	Kvensk	Annet, beskriv
Morfar:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mormor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Farfar:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Farmor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Far:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jeg selv:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hva er din, din fars og din mors etniske bakgrunn? (sett ett eller flere kryss)				
	Norsk	Samisk	Kvensk	Annet, beskriv
Min etniske bakgrunn er:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fars etniske bakgrunn er:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mors etniske bakgrunn er:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Hva regner du deg selv som? (sett ett eller flere kryss)				
	Norsk	Samisk	Kvensk	Annet, beskriv
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3.2. Finnmarksundersøkelsene

Finnmarksundersøkelsene er en annen stor helsestudie som er gjennomført i til sammen fem runder i Tromsø kommune og seks runder i Finnmark fylke i tidsrommet 1974–2000 (Severeide 2008). Undersøkelsene har vært gjennomført som et samarbeid mellom Institutt for Samfunnsmedisin (ISM) og Statens helseundersøkelser (SHUS), og består av en spørreundersøkelse i kombinasjon med kliniske undersøkelser. En benyttet de samme spørsmål om etnisitet i alle runder, formulert som: «Er to eller flere av Deres besteforeldre av samisk ætt?» (Andersdatter Siri 2015). Samme spørsmål ble stilt for finsk ætt. Til sammen 3 705 personer svarte ja på spørsmålet om samisk ætt i disse undersøkelsene (Severeide 2008).

Når det gjelder utnytting av disse dataene, blir resonnetet det samme som for SAMINOR-undersøkelsene beskrevet over. Mest sannsynlig vil det altså ikke la seg gjøre å utnytte mikrodata fra disse undersøkelsene siden respondentenes samtykke i utgangspunktet ikke omfatter annen bruk enn beskrevet. Imidlertid kan data utnyttes statistisk til kalibrering under utvalgs- og analysearbeidet.

3.3. Ungdomshelseundersøkelsen i Nord-Norge, UHNN

UHNN var en omfattende studie som ble gjennomført blant alle tiendeklasser i Nordland, Troms og Finnmark i tidsrommet 2002–05 (Andersdatter Siri 2015). Undersøkelsen var et samarbeid mellom Senter for samisk helseforskning (SSHF), Nasjonalt folkehelseinstitutt, Institutt for samfunnsmedisin, UiT og Nasjonalt folkehelseinstitutt. Etnisitet var et sentral anliggende, og det ble stilt spørsmål om egen og foreldres oppfatning av etnisk tilhørighet, samt spørsmål om eget og foreldre/besteforeldres språk. Det ble også spurt om hvilket språk en har benyttet i skolen. Figur 3 viser spørsmålene om etnisitet og språk som ble benyttet.

Figur 3.2. Spørsmål om etnisitet og språk. Fra spørreskjema til UHNN

7.2 Hvordan ser du på deg selv? (Sett ett kryss for hver linje)

Jeg oppfatter meg selv som:	Helt enig	Delvis enig	Delvis uenig	Helt uenig
Norsk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Samisk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kvensk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Finsk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Annet (hva)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4 3 2 1

Min etnisitet er (skriv ett eller flere av tallene nedenfor):
 1=Norsk, 2=Samisk, 3=Kvensk, 4=Finsk, 5=Annet
 (skriv hvilken):

Fars etnisitet er (bruk tallene ovenfor):

Mors etnisitet er (bruk tallene ovenfor):

7.3 Hvilket språk snakker du og familien din? (Sett ett eller flere kryss)

	Norsk	Samisk	Kvensk/finsk	Annet språk
Hjemme har jeg lært	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1. språk på skolen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. språk på skolen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Far snakker	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mor snakker	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Farmor snakker(t)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Farfar snakker(t)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mormor snakker(t)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Morfar snakker(t)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3.4. Sør-Varanger-undersøkelsen

Som navnet tilsier ble undersøkelsen gjennomført i Sør-Varanger kommune i perioden 1994–95, og alle innbyggere i alderen 18–69 år ble invitert til å delta (Andersdatter Siri 2015). I tillegg til et spørreskjema, ble det tatt blodprøver fra alle deltakere. 60 prosent av befolkningen deltok. Etnisitetsspørsmålet var formulert som: «Er én eller flere av dine besteforeldre av samisk slekt?». I tillegg ble det spurt om samisk språk, med formuleringen: «Har én eller flere av dine besteforeldre hatt samisk som hjemmespråk?». De samme to spørsmålene ble også stilt for finsk (kvensk) slekt og språk.

For Sør-Varanger gir disse dataene åpenbart et godt bidrag til personer med enten samisk eller kvensk slekt og språk i området. Også for disse dataene er vi avskåret fra utnytting på mikronivå av samme grunn som beskrevet over.

3.5. SLiCA-studien

Survey of Living Conditions in the Arctic (SLiCA) er en internasjonal levekårsundersøkelse gjennomført i Alaska, Grønland, Sverige, Russland og Norge (Poppel, 2015). Her til lands ble undersøkelsen utført i 2003 og i perioden 2006–08 i områdene Finnmark, Troms, Nordland og Trøndelag. Aldersgruppen var personer over 16 år. Det ble benyttet standardiserte spørreskjema administrert av intervjuere gjennom ansikt til ansikt-intervju. Utvalget av personer som ble invitert til å delta i områdene Karasjok, Kautokeino og Nesseby ble også delvis rekruttert blant de som hadde vært deltakere i SAMINOR 1-undersøkelsen og som hadde gitt tillatelse til å

bli kontaktet igjen. I tillegg benyttet man seg også av en noe mer utradisjonell rekruttering, ved at personer som allerede hadde blitt intervjuet ble spurt om de kjente til andre aktuelle kandidater, den såkalte snøballmetoden (Andersdatter Siri 2015)⁵.

Det ble stilt spørsmål om mor og fars etnisitet («betrakter seg som same eller norsk/finsk/svensk») og hvilket språk som ble lært og benyttet i hjemmet.

På grunn av den noe utradisjonelle fremgangsmåten for rekruttering av respondenter (snøballmetoden, kun for deler av utvalget), er det noe mer problematisk å benytte resultatene fra SLiCA-studien som et godt mål for utbredelse av samisk språk.

3.6. Andre helseundersøkelser blant samer

Som vi har sett har det blitt gjennomført en rekke studier med mål om å belyse den nordligste befolkningens levekår og helse. Samers livsvilkår har gjerne vært en sentral dimensjon en har vært ute etter å si noe om. Undersøkelsene har til dels gått ut til alle personer i en viss aldersgruppe i områder med kjent høy andel samer. Tromsøundersøkelsen i 1974, Ungdom trivsel og adferd i 1989–93 og Miljøgifter i svangerskapet i 2007–09, er alle omfattende forskningsprosjekter gjennomført i Nord-Norge, i tillegg til de vi har gjennomgått i denne utredningen (Andersdatter Siri 2015). I alle disse studiene har en gjerne spurt om både egen, foreldre og besteforeldres språk og oppfatning av etnisitet. Til sammen gir studiene et grundig og bredt bilde av samisk etnisitet og språk i Nord-Norge. Som inntak til andelen samiskspråklige i disse områdene, gir det en god pekepinn, men for samisktalende i resten av landet sier disse studiene ingen ting. En annen svakhet for vårt formål er at ingen av disse helsestudiene ser ut til å skille mellom de ulike samiske språkene.

4. Populasjon og utvalgsundersøkelse

4.1. Populasjonen

Vår undersøkelse har som mål å kartlegge utbredelsen av det samiske språket i Norge. For å kunne gjøre dette, må vi som nevnt starte med å lage en populasjon som inkluderer både de som er samiskspråklige, dvs. kan snakke et av de samiske språkene, og de som har et potensiale for å utvikle sine språkferdigheter i samisk. En slik populasjonsdefinisjon har den ulempen at den ikke gir helt klare grenser for om en person tilhører populasjonen eller ikke. Men det gir samtidig en viss frihet i hvordan populasjonen lages.

Som tidligere nevnt, er det i prinsippet mulig å lage en populasjon over en del av samiske språkbrukere ved å ta utgangspunkt i noen relevante registre hvor vi har fødselsnummer. Her vil dette i første omgang være det samiske valgmanntallet og FoB 1970. Men andre registre kan også være aktuelle. For eksempel kan vi bruke befolkningsregisteret og inkludere alle som bor innenfor det samiske språkområdet (NOU, 2016, s. 96), eventuelt innenfor STN-området. Hvilke registre vi til slutt bruker, vil avhenge av a) om vi får tilgang til dataene, b) om enhetene er identifisert med fødselsnummer, og c) relevans.

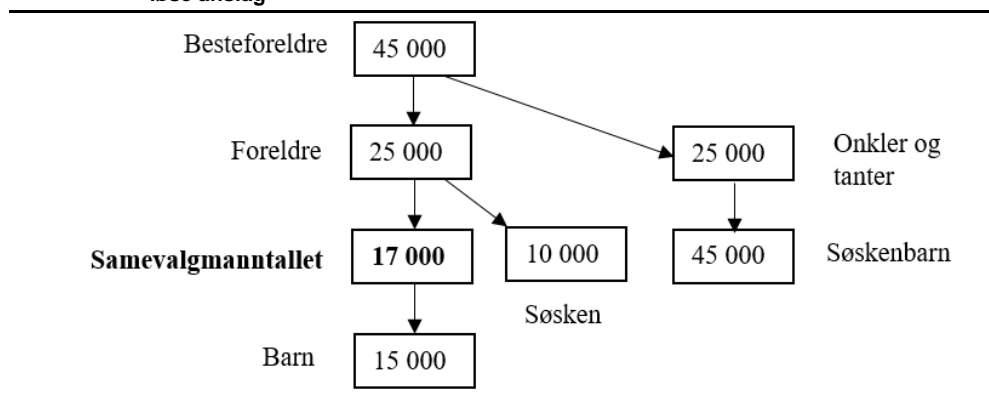
De registrene vil til slutt ender opp med, vil helt sikkert overlappe en god del. Vi slår derfor sammen registrene på en slik måte at hver person kun opptrer én gang. Dette sammenslåtte registeret vil høyst trolig være mangelfullt både med hensyn til

⁵ Det er noe uklart for oss om en med begrepet «aktuelle kandidater» her mener personer med samisk, ev kvensk avstamning, eller om det rett og slett var personer bosatt i de angitte områdene en var ute etter.

å dekke de samiskspråklige i Norge og når det gjelder å dekke de som kan tenke seg å lære språket. Ideen er nå at de sistnevnte gjerne vil være i slekt med noen som finnes i et av de nevnte registrene. Vi etablerer derfor et slektstre med utgangspunkt i personene i registeret. Her vil det være noen praktiske begrensninger i hvor langt ut i slektstreet det er mulig å gå. Og metodikken bør trolig avhenge av hvilket register en person er hentet fra. For eksempel kan det være hensiktsmessig å se på alle etterkommere til besteforeldrene til de som står i dagens samemantall, se eksempel i figur 4. For en person som vi finner i FoB70, virker ikke dette like relevant. Her kan vi f.eks. nøye oss med å se på direkte etterkommere, eventuelt også etterkommere til søsken.

Det vi til slutt står igjen med er en populasjon av personer som enten snakker samisk eller som vi tror kan ha et potensiale for å lære seg samisk pga. slektskap.

Figur 4.1 Eksempel på slektstre basert på samvalgmanntallet. Antall i hver slektsboks er kun løse anslag



4.2. Utvalget

Det ideelle vil nå være å gjøre en undersøkelse som omfatter hele populasjonen, en fulltelling, men det vil det neppe finnes ressurser til. For eksempel var det over 10 000 som sa de hadde samisk som førstespråk i FoB 1970 (Aubert, 1978) og ca. 17 000 personer som hadde stemmerett ved Sametingsvalget 2017 (Sametinget.no 2017). Videre har STN-området over 55 000 innbyggere. Når vi i tillegg kobler på slektninger, se eksempel Figur 4, sier det seg selv at populasjonen raskt blir for stor til at en fulltelling er praktisk gjennomførbart. Altså gjør vi den for kun et utvalg av populasjonen.

Stratifisering

For å sikre at utvalget gir best mulig informasjon om populasjonen som utvalget trekkes fra, gjør vi først en stratifisering. Det er å dele inn populasjonen i et gitt antall ikke-overlappende grupper (strata) som til sammen dekker hele populasjonen. Stratifisering gjør at vi luker bort lite representative utvalg allerede i trekkingen, og vil derfor gi oss mer presise estimater av totaltall i populasjonen.

Vi prøver å bestemme strataene på en slik måte at personer i ett og samme stratum blir mest mulig like med hensyn til det vi er interessert i å undersøke, i dette tilfellet tilknytning til, eller potensiale for, samisk språk. Etersom vi ikke vet veldig mye om dette før vi har gjort undersøkelsen, må stratuminndelingen baseres på variabler som vi tror har tilknytning til språk. Det kan være geografi, for eksempel om man bor innenfor det samiske språkområdet eller innenfor STN-området, om man står i et de viktigste registrene (samisk valgmanntall, FoB 1970, osv.), eller om man har kommet med på grunn av et slektsforhold. Kombinasjoner av variable er også mulig. I tillegg tas det gjerne hensyn variabler som brukes i publisering av statistikken, som f.eks. kjønn og alder. Et enkelt eksempel på inndeling i fire strata kan være:

1. Hadde samisk som førstespråk i FoB70 eller står i samisk valgmanntall
2. I slekt med noen i punkt 1 og bor i et typisk samisk område
3. I slekt med noen i punkt 1 og bor ikke et typisk samisk område
4. Andre som bor i et typisk samisk område

Når populasjonen er delt inn i strata, er en vanlig fremgangsmåte å trekke et såkalt enkelt tilfeldig utvalg innen hvert av strataene. At et utvalg er enkelt tilfeldig, betyr at alle personene i stratumet har samme sannsynlighet for å bli trukket i utvalget. Det medfører også at alle mulige utvalg av en gitt størrelse blir like sannsynlige. Vi sier da gjerne at utvalget er selvveiende, m.a.o. at alle personene i utvalget skal tillegges samme «vekt» når vi lager statistikken. Både andel og antall personer som trekkes kan variere mellom strataene.

Størrelse

Det er en rekke forhold som er med på å bestemme hvor stort utvalget skal være: ressurser, statistikkens detaljeringsgrad, krav til nøyaktighet, forventet frafall og hvor homogen populasjonen er. I praksis vil det imidlertid alltid ressursene være det viktigste. Man ønsker alltid et størst mulig utvalg, samtidig som utvalget trolig aldri blir stort nok til å kunne dekke alle statistikkbehov med en akseptabel grad av nøyaktighet.

Det er alltid mulig å lage noen enkle talleksempler som viser egenskapen til ulike utvalgsstørrelser. La oss si at vi ønsker å bestemme hvor stor andel, p , som har en bestemt egenskap i populasjonen. Dette kan for eksempel være andelen som behersker et bestemt språk. Anta så at den faktiske andelen er på 20 prosent ($p = 0,2$), og vi ønsker at statistikken for denne andelen, \hat{p} , skal ha en standardfeil på én prosent, dvs. $SE(\hat{p}) = 0,01$. Da viser enkel beregning at utvalget må ha størrelse $n = (0,2 \times 0,8)/0,01^2 = 1600$. Dette kan kanskje virke som et lite tall, men da må vi være klar over at vi med dette kun greier å måle en enkel størrelse i populasjonen med god nøyaktighet. En detaljering av statistikken ned på alder, kjønn eller geografi vil kreve et mye større utvalg for å opprettholde samme nøyaktighet.

Og i tillegg må vi regne med en stor andel frafall i undersøkelsen. Et frafall på 60 prosent eller mer, er ikke utenkelig. Særlig hvis en sikter seg inn på et undersøkelsesdesign med selvutfylling på web. For at vi skal sitte igjen med et nettoutvalg på 1600 etter frafall, må vi, med et frafall på 60 prosent, ha et bruttoutvalg på $1600/0,4 = 4000$. Og dette vil, som nevnt, være for lite for å kunne gi statistikk på et mer detaljert nivå enn populasjon.

4.3. Frafall og utvalgsvekting

Det nettoutvalget vi ender opp med, vil ha litt andre egenskaper enn det bruttoutvalget vi opprinnelig trakk. Frafall har en tendens til å være skjev, dvs. at tilbøyeligheten til å svare på en undersøkelse vil variere mellom forskjellige grupper. Konsekvensen av dette er at nettoutvalget *ikke* lenger blir selvveiende innen hvert stratum. Dette må rettes opp ved hjelp av frafallsvekting, noe som betyr at personer som tilhører en gruppe vi antar er underrepresentert i nettoutvalget, gis en høyere vekt enn personer som tilhører en gruppe vi antar er overrepresentert. Hvilke grupper som er under- og overrepresentert, finnes enkelt sagt ut ved å sammenligne fordelingen av ulike variabler i nettoutvalget med fordelingen av de samme variablene i populasjonen, eventuelt i bruttoutvalget.

Grunnen til at vi er særlig bekymret for skjevheter i frafallet for akkurat denne undersøkelsen, skyldes tematikken i seg selv. Det å være same, har historisk vært sett på som en belastning for enkelte, og som kjent ble det i tidligere tider ført en

aktiv fornorskningsspolitikk fra statens side overfor personer av samisk ætt (NOU 2016). Det er naturlig å tenke seg at noe av dette fremdeles kan sitte igjen hos enkelte personer med samisk opprinnelse, og dermed føre til at de avstår fra å delta i undersøkelsen. Når det en skal måle har direkte sammenheng med tilbøyeligheten for å delta, som det har i dette tilfellet, er det særlig problematisk mht. skjevhet i utvalget. En må forsøke å vekte opp for dette etter beste evne som beskrevet over, men det er umulig å sikre seg mot at estimatene kan bli påvirket av slike frafallsgrunner.

Variabler brukt i frafallsvektingen

En god og effektiv vekting skal være slik at statistikken vi ender opp med, er mest mulig lik den statistikken vi ville fått dersom vi hadde gjort en fulltelling og fått svar fra alle. For å oppnå en god vekting, må det vektes med hensyn på størrelser som er kjent i både nettoutvalget og populasjonen, og som har en sammenheng med det vi skal lage statistikk om. Merk at for å kunne bruke en variabel i vektingen, trenger vi kun å kjenne dens verdi på individnivå i nettoutvalget. For populasjonen er det tilstrekkelig å kjenne totalsummen av variabelen.

Variablene som blir brukt til å konstruere strata, er opplagte kandidater, men det er også vanlig å inkludere andre. Blant annet vil vi vurdere å utnytte eksisterende statistikk over samisk språk i vektingen, som referert under gjennomgangen av kilder over. I tilfelle vil frafallsvektene få den egenskapen at resultatene fra utvalgsundersøkelsen blir identisk med den statistikken vi har vektet mot. Forutsetningen for å bruke en slik metode, er at statistikken er relevant og av bra nok kvalitet. Den må også antas å gjelde for den populasjonen som utvalget er trukket fra. Det siste betyr i praksis at den bør være av nyere dato, slik som for eksempel den siste Tromsøundersøkelsen.

4.4. Spørreundersøkelse – design

Med utgangspunkt i vår statistiske populasjon over samiske språkbrukere, reelle og potensielle, kan det trekkes et utvalg etter de retningslinjer som er beskrevet i kapittel 4.2, med siktemål om å gjennomføre en spørreundersøkelse blant de utvalgte personene. Hvor stort utvalg en ender opp med å trekke er altså i stor grad et ressurs spørsmål. Valg av design og kommunikasjonskanal for undersøkelsen er i den sammenheng høyst sentralt. Velger vi å gjennomføre undersøkelsen med en eller annen form for selvutfyllingsskjema, enten på papir eller web, er kostnadene vesentlig lavere enn om en bestemmer seg for å intervju respondentene på telefon eller ansikt-til-ansikt. Til gjengjeld vil selvutfylling på web, som kanskje er det mest hensiktsmessige selvutfyllingsmodus i våre dager, innebære stort frafall. En regner det gjerne som svært krevende å oppnå svarandeler over 15–20 prosent på web-undersøkelser.

Vi har nevnt skepsis mot å «flagge» samisk etnisitet som en utfordring når det gjelder svarvilligheten. På den andre siden vil trolig samisktalende som ikke opplever et slikt stigma knyttet til sin samiske herkomst, oppleve tema for undersøkelsen som høyst relevant og interessant. Det taler sterkt for høy svarvillighet. Så her er det krefter som trekker i ulike retninger. Sannheten rundt hvor stor svarvilligheten faktisk vil være, er det kun empirien som vil kunne fortelle oss. Vi må gjennomføre en prøveundersøkelse for å få en pekepinn på dette. Hvilke svar piloten gir oss vil i stor grad bestemme hvilken eller hvilke kanaler en bør benytte seg av for undersøkelsen. En kombinasjon, der en først går ut med en spørreundersøkelse på web, for dernest å følge opp frafallet ved hjelp av intervjuere, kan være en vei å gå.

Om en skal benytte intervjuer på telefon eller ansikt-til-ansikt vil i stor grad være et spørsmål om kostnader. Besøksundersøkelser er svært kostbare, og med de

utvalgsstørrelser og den geografiske stratifisering i til dels krevende tilgjengelige områder vi her snakker om, vil det også by på en del praktiske utfordringer.

Gjennom prosjektets gang har vi også diskutert om vi bør benytte ulike fremgangsmåter for de nordligste fylkene, med høy andel samer, og for landets øvrige fylker. En kunne f.eks. tatt utgangspunkt i forvaltningsområdet for samisk språk, eller STN-området, og inkludert alle personer bosatt i disse områdene i populasjonen. Hvorvidt en bør gjøre dette eller ei, er imidlertid avhengig av hvor god dekning vi oppnår i disse områdene når vi etablerer den beskrevne statistiske samiskspråklige populasjonen. Hvis dekningen viser seg å bli god, skulle det ikke være andre grunner for å inkludere alle innbyggerne, enn at vi dermed også vil få med de relativt få personene som snakker samisk uten noen annen tilknytning til det samiske enn nettopp denne egenskapen. Det vil som før nevnt, neppe være regningsssvarende. Men så vet vi ikke akkurat om fremgangsmåten med vår statistiske samepopulasjon viser seg å være riktig så god. Her er det fristende å trekke inn alle de øvrige studiene som er gjennomført i de nordligste fylkene, og som vi har sett har data på individnivå som viser både samisk språk og etnisk bakgrunn. Hvis det hadde latt seg gjøre å sammenlikne vår statistiske same-språklige populasjon med disse kildene på individnivå, hadde det gitt oss et meget nyttig verktøy for å bestemme kvaliteten på dekningen i vår statistikkpopulasjon. Hvis det skulle vise seg å være få personer som har oppgitt samisk språk eller etnisitet i noen av disse forskningsstudiene, og som ikke fremkommer gjennom vår statistikkpopulasjon, kunne vi ha konkludert med at registeret vi har konstruert sannsynligvis har god dekning i Nord-Norge. Det ville også indirekte være en god indikasjon på registerets kvalitet i resten av landet. Som nevnt er det høyst usikkert om det på en eller annen måte ville ha latt seg gjøre å utnytte disse forskningsdatabasene på denne måten, men vi lanserer det allikevel her som en mulighet som det kan være interessant å sjekke ut nærmere.

Vi finner det ikke hensiktsmessig å si noe mer konkret om utforming og design på en eventuell fremtidig undersøkelse med sikte på å kartlegge samisk språkbruk. Det samme gjelder kostnadsbildet, som i stor grad vil være avhengig av designet, som nevnt. Det er allikevel sikkert at en slik studie vil bli relativt kostbar å gjennomføre med de ønsker som er signalisert rundt å kunne gi tall for ulike samiske språk, brutt ned på kommune og gitte alderskategorier.

5. Veien videre

Vi konkluderer med at det bør være gjennomførbart å frambringe brukbare tall over utbredelsen av samiske språk i Norge. Hvor gode tall vi vil kunne regne med å få, er i stor grad avhengig av hvor mye ressurser en er villig til å benytte under datafangsten, men store ressurser er heller ikke nødvendigvis tilstrekkelig. Det er nærmest umulig å vite om det kan være personer med tilknytning til det samiske språket som vi ikke makter å komme i kontakt med for å spørre dem om deres forhold til språket. En sjekk mot tidligere gjennomførte undersøkelser vil som nevnt over være en spennende teoretisk mulighet, men altså med høy grad av usikkerhet knyttet til om en slik tilgang vil være mulig.

I tillegg har vi usikkerheten rundt hvordan tematikken i seg kan tenkes å påvirke svarvilligheten, med tilknyttet skjevhetsproblematikk, beskrevet over.

Vi argumenterer for at fremgangsmåten med å ta utgangspunkt i kjente register over personer med en eller annen forbindelse til det samiske, og ut fra dette grunnlaget nøste seg videre gjennom familie- og slektskap, trolig vil sette oss i forbindelse med de aller fleste personer i målgruppa. Vi kan selvsagt ikke garantere at noen ikke klarer å «gjemme seg unna», men denne muligheten minskes ved at de

samiske samfunnene er så pass bundet sammen geografisk, særlig var de det i tidligere tider. Når vi nøster oss tilbake noen generasjoner, tror vi derfor at vi vil få med de aller fleste i vår samespråklige statistikkpopulasjon. For statistikkens presisjon spiller det mindre rolle om vi ikke makter å få med absolutt alle. Vi tenker allikevel ikke å foreta noen fulltelling, men kun benytte grunnlaget til å trekke et representativt utvalg.

Dessuten er fremgangsmåten vi har skissert her den eneste farbare veien vi har klart å komme fram til. Det er godt mulig at det finnes enda flere kilder enn de vi har gjennomgått her, som vi ville kunne trekke veksler på. I tilfelle er det bare å inkludere dem i det endelige designet, i den grad en har ressurser til det. Kanskje ville en slik ny kilde kunne gi oss noen flere personer for videre nøsting av familie- og slektskap. Men for hver ny kilde vi legger til, vil det bli færre personer som ikke allerede er inkludert gjennom de kildene vi allerede har benyttet. Nytteverdien må derfor veies opp mot den krympende merverdien for hver ny kilde. Sannsynligvis vil det være et økonomisk spørsmål, der en ender opp med å inkludere de største og beste kildene, samt de som ikke innebærer altfor store ressurser å hente inn.

5.1. Pilot

For å oppnå bedre kunnskap om kvaliteten på de ulike kildene og en mer detaljert beskrivelse av fremgangsmåte for kartlegging av familierelasjoner, er det slik vi ser det, nødvendig å gjennomføre en type pilot. Her går en fram som om det skulle gjennomføres en faktisk undersøkelse. Vi søker tilgang til de kildene vi ønsker å benytte og undersøker kvaliteten på dem. Dette kan først gjøres fullt ut når vi har de faktiske registrene med fødselsnummer tilgjengelig og kan foreta kjøringer opp mot Folkeregisteret for å kontrollere beskaffenhet og kvalitet. Deretter sammenstiller vi alle kildene på fødselsnummer i ett nytt sammenslått register og kobler på familie- og slektskapsrelasjoner som beskrevet over. Deretter trekkes et mindre utvalg, stratifisert etter de grunnprinsipper vi har beskrevet ovenfor, ev. med de justeringer vi finner hensiktsmessige basert på den lærdom vi har gjort oss i første fase av pilotprosjektet. I en slik pilot settes utvalgsstørrelsen gjerne lavt, f.eks. til 50 personer. Disse mottar spørreundersøkelsen med de spørsmål vi ønsker å stille og gjennom den eller de kommunikasjonskanaler en har bestemt seg for. Som nevnt kan det være en fordel å prøve ut litt ulike kanaler og kontaktstrategier for å finne den optimale avveiningen mellom kvalitet og kostnader.

Svarene fra respondentene og andre erfaringer en har gjort i gjennomføringsfasen analyseres så i etterkant for å avdekke mangler ved spørsmål og design. Først etter at en slik pilot er ferdig analysert og man har konkludert mht. hensiktsmessige justeringer, kan en starte planlegging og utforming av den faktiske studien med mål om å kartlegge utbredelsen av de samiske språkene i Norge.

Grunnet stor usikkerhet har vi ikke sett det formålstjenlig å forsøke oss på en prissetting av en spørreundersøkelse med angitt tema. Svarene en får i piloten vil imidlertid sette oss i bedre stand til å beregne kostnadsbildet for gjennomføring av studien og hvor gode tall det vil være mulig å sitte igjen med, gitt ressursene en er villig til å bruke på prosjektet.

Referanser

- Aubert, V. (1978). *Den samiske befolkning i Nord-Norge. Sámi ál'bmut Davvi-Norgas*. Artikler 107, Statistisk sentralbyrå.
- GSI (2017). Grunnskolens informasjonssystem. Utdanningsdirektoratet.
- Meld.St.32 2016-2017. *Reindrift. Lang tradisjon – unike muligheter*. Det Kongelige landbruks- og matdepartement. 2017.
- NOU (2016). *Hjertespråket: Forslag til lovverk, tiltak og ordninger for samiske språk*. Norges offentlige utredninger, 2016:18.
- Poppel, B. (Red.) (2015). *SLiCA: Arctic living conditions – Living conditions and quality of life among Inuit, Saami and indigenous peoples of Chukotka and the Kola Peninsula*. TemaNord 2015:501, DOI: 10.6027/TN2015-501.
- Andersdatter Siri, S. R. (2015). *Bruk av samisk etnisitet i forskningsdatabaser og helseregistre*. Rapport fra Senter for samisk helseforskning/Sámi dearvvasvuodadutkama guovddás. UiT
- Severeide, P.I. (2008). Datagrunnlaget – begrensninger og muligheter. Samiske tall forteller 1. Kommentert samisk statistikk 2008. Sámi allaskuvla.
- Wilhelmsen, M., Holth, B.A., Kleven, Ø., Risberg, T. (2013). *Minoritetsspråk i Norge. En kartlegging av eksisterende datakilder og drøfting av ulike fremgangsmåter for statistikk om språk*. Notater 8/2013. Statistisk sentralbyrå.

Vedlegg A: Samisk språk. Guide med spørsmål til registreiere

Bakgrunn

Statistisk sentralbyrå (SSB) gjennomfører for tiden en metodeutredning på oppdrag for Sametinget. Oppdraget går ut på å undersøke muligheter for å kartlegge utbredelsen av samiske språkbrukere i Norge. Det er ønskelig å spesifisere hvilke av de samiske språkene det tales, Nordsamisk, Lulesamisk, Sørsamisk, Pitesamisk, Østsamisk, ev. andre. Oppdragsgiver ønsker også å inkludere personer med et *potensiale* for å utvikle sine språkferdigheter i et av de samiske språkene. Det vil typisk kunne være personer av samisk opprinnelse som har mistet sitt samiske «morsmål».

Som del av oppdraget er det nødvendig å kartlegge hvilke aktuelle registerkilder som finnes fra før, og som enten direkte kan fortelle noe om språkferdigheter i samisk, eller indirekte kan gi et inntak til en populasjon over personer med samisk tilknytning og språk (reelle eller potensielle språkferdigheter). Vi har derfor laget noen spørsmål om egenskaper ved aktuelle registre som vi håper at registreiere/ev. ansvarlige vil hjelpe oss med å besvare så godt det lar seg gjøre.

Hvis vi også kan få opplysninger om kontaktperson(er) med kjennskap til registeret som ved behov kan kontaktes ved en senere anledning, ville vi satt stor pris på det.

Spørsmål rundt registerets beskaffenhet

Spørsmålene besvarer i den grad de er relevante for aktuelt register.

1. Hva heter registeret?
2. Hva er enhet i registeret?
3. Hva er identifikasjonsmerke? Er det fødselsnummer (11 siffer), fødselsdato (6 siffer), kun navn og adresse eller noe annet?
4. Hvor mange enheter består registeret av i dag?
5. Hvilke variable er registrert?
 - a. Bakgrunnsvariabler, f.eks. alder, bosted, osv.
 - b. Språkvariabler, type samisk språk (Nordsamisk, Lulesamisk, Sørsamisk, Pitesamisk, Østsamisk, ev. andre).
 - c. I tilfelle det er relevant: Er det noen form for gradering av språkferdigheter?
6. Hvordan foregår registreringen, og hvordan holdes registeret oppdatert?
7. I hvilket format ligger dataene lagret?
8. Hva er registreiers generelle vurdering av styrker og svakheter ved registeret?
9. Hvordan vurderes kvaliteten med tanke på vårt behov: kartlegging av samiske språkbrukere/personer med et potensiale for å utvikle sitt samiske språk?
10. Har registreier noen motforestillinger knyttet til ev. fremtidig tilgang til registeret (mikrodata) for SSB?
11. Hva er navn, e-post og telefonnummer på kontaktperson for senere henvendelser vedr. registeret?

Statistisk sentralbyrå

Postadresse:
Postboks 8131 Dep
NO-0033 Oslo

Besøksadresse:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-post: ssb@ssb.no
Internett: www.ssb.no
Telefon: 62 88 50 00

ISBN 978-82-537-9601-7 (elektronisk)



Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist