



Forskingsetiske normer og juridiske rammer i norsk språkforsking med data frå sosiale medium

Geir Bakkevoll og Anne-Mette Somby

Denne artikkelen bygger på ei undersøking av seks nyare, norske språkstudier med data frå sosiale medium, og han handlar om forkingsetiske normer og juridiske rammer i slike prosjekt. Artikkelen diskuterer først skiljet mellom anonyme og aidentifiserte data og informantar. Når desse omgrepa blir brukte rett, kan språkforskaren sjeldan hente data frå sosiale medium utan å melde prosjektet og få det vurdert frå Norsk senter for forskningsdata (NSD). Vidare utforskar artikkelen omgrepet personopplysing og argumenterer for at språklege element kan vere personidentifiserande. Når ein behandlar slike personidentifiserande opplysingar i forking, står norma om det informerte samtykket fram som hovudregelen i den norske konteksten. Dette er berre eitt av dei lovlege grunnlaga for å behandle personopplysingar. I diskusjonen av desse lovlege grunnlaga viser artikkelen korleis teknologien kan gjere det vanskeleg å informere deltakarane i studien. Eit siste spørsmål er korleis språkforskinga i sosiale medium definerer og behandlar tredjepartar, og artikkelen argumenterer for at desse eigentleg er informantar i sosiolingvistisk forking.

Nøkkelord: CMC, sosiale medium, forkingsetikk, personopplysing, reidentifisering, anonymisering, aidentifisering, informert samtykke, tredjepart

1. Innleiing

Det norske feltet av språkforskarar som hentar data frå sosiale medium blir stadig større. Hårstad (2020: 142) seier at digitale arenaar er «eit domene som enno er ganske nytt i forkingssamanheng» og at dei etiske problemstillingane

derfor er særleg viktige. I dei nyaste etiske retningslinjene for forskning på internett peikar den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH) på at den teknologiske utviklinga skapar nye forsknings-etiske utfordringar (NESH 2018: 4). Nesvold, Sjurseike og Wadel (2014: 107) held fram den menneskelege faktoren i ei potensiell blindsona for forskaren: «Det er også noe ved internetts karakter som skaper en avstand mellom forsker og debattant som kan bidra til å redusere oppmerksomheten omkring disse etiske spørsmålene sammenlignet med tradisjonelle møter mellom forsker og informant.» Denne artikkelen er derfor vigd etikk og jus når norske språkforskarar brukar data frå sosiale medium.¹ Artikkelen brukar sosiale medium som samlenemning for digitale plattformer der brukarane sjølve kan skape og dele innhald og som slik legg til rette for sosial interaksjon (Enjolras, Karlsen, Steen-Johnsen & Wollebæk 2013: 11).

1.1 Forskingsspørsmål og inklusjonskriterium

Nokon etiske dilemma for språkforskinga i sosiale medium har blitt meir utforska enn andre dei seinare åra. For eksempel har Hårstad (2020), Lüders (2015) og Bolander og Locher (2014) diskutert det uklare skiljet mellom privat og offentleg innhald i sosiale medium. Desse diskusjonane viser seg blant anna i den nye *Forskingsetisk veileder for internettforskning* (NESH 2018). I versjonen frå 2014 seier NESH (2014: 4) at «forskere som hovedregel fritt [kan] benytte materiale fra åpne fora uten å innhente samtykke fra dem opplysningene gjelder». I den nye versjonen blir det tydelegare at skiljet mellom det opne og det lukka ikkje er eindimensjonalt, men at det er «viktig å skille mellom *yttringens offentlighet* og *informasjonens sensitivitet*: Ytringen kan være fremsatt privat *eller* offentlig, og innholdet kan være personlig *eller* generelt» (NESH 2018: 9-10).

-
1. Her skal vi takke Øystein Vangsnes for kunnig rettleiing av masteroppgåva til Bakkevoll (2019) og for oppmuntringa til å skrive ein sjølvstendig artikkel om dei etiske vurderingane og dialogen med NSD undervegs i prosjektet. Ei liknande oppmuntring frå Unn Røyneland og Åse Mette Johansen i eksamenskommissjonen ligg òg til grunn for denne artikkelen. Redaktørane Elin Gunleifsen, Gro Renée Rambø og Magnhild Selås og dei tre (anonyme) fagfellane skal òg ha ein stor takk for innspela og lesetipsa som har forbetra artikkelen monaleg. Alt som kunne og burde vore annleis, derimot, tar vi sjølve ansvar for. Dei seks som står bak studiane i denne analysen, fortener utan tvil den aller største takka, for spennande lesing og grundige vurderingar i eit relativt nytt og komplisert forskningsetisk landskap. Deira refleksjonar og val har altså vi gjort til studieobjekt. Slik er dei seks med på å bringe den språkvitskaplege forskinga vidare. Takk for at vi kan trø opp desse stiane saman.

Denne undersøkinga, som legg seks nyare, norske språkstudiar under lupa, viser at andre forskingsetiske spørsmål treng meir merksemd. Det første forskingsspørsmålet i denne artikkelen er *korleis nyare, norsk språkforskning i sosiale medium definerer ei personopplysing*, sagt på ein annan måte er spørsmålet korleis ein vurderer om ein informant kan identifiserast. Det andre spørsmålet er *korleis det same forskingsfeltet forvaltar den normative forventinga om at informantane i ein studie skal gi eit informert samtykke*. Til slutt undersøker studien *korleis nyare, norsk språkforskning i sosiale medium definerer og behandlar tredjepartar*.

Eit forskingsprosjekt kan delast inn i fasane innsamling av data, analyse, publisering og eventuell lagring eller gjenbruk av forskingsdataa. I denne artikkelen krinsar diskusjonane om innsamling, analyse og publisering. Vidare nemner forskingsspørsmåla over tre inklusjonskriterium for dei seks språkstudiane som er undersøkte, kriterium som kjem i tillegg til at studiane har data frå sosiale medium.

For det første er avhandlingane *norske* forskingsbidrag. Artikkelen ser etiske vurderingar opp mot normative tekstar som er relevante for forskarar. Det gjeld først og fremst rettleingane frå dei nasjonale forskingsetiske komiteane (FEK), der mandatet er å bidra til at forskinga i Norge skjer etter anerkjente forskingsetiske normer (FEK 2018, forskingsetikkloven §9). Særleg *Forskingsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi* (NESH 2016) og *Forskingsetisk veileder for internettforskning* (NESH 2018) er relevante normative tekstar. Ein annan normativ instans er Norsk senter for forskningsdata (NSD), som er eit kompetansesenter for personvern og sørger for lovleg tilgang til persondata ved at forskaren melder inn prosjektet.² Råda om personvern og meldeskjemaet på nettsidene til NSD må reknast blant dei normative tekstane for språkforskaren som skal hente data i sosiale medium.^{3,4} For at samanlik-

-
2. Frå juli 2018 gjeld den nye personopplysingslova der den europeiske personvernforordninga, kjend som GDPR, er inkludert. Med det er òg rolla til NSD endra frå å vere personvernombodet for forskning til å vere eit kompetansesenter for personvern, mens den enkelte forskingsverksemda etter lova skal ha sitt eige personvernombod. NSD leverer framleis personverntenester til dei aller fleste norske forskingsinstitusjonane. Praksisen med å melde prosjekt som behandlar personopplysingar, er derfor ikkje vesentleg endra. Standpunktet i denne artikkelen er at ein vanskeleg kan tenke seg eit forskingsprosjekt om språk i sosiale medium som ikkje må ha ei vurdering etter personvernregelverket, meir om dette i del 3.1.
 3. <https://www.nsd.no/personverntjenester/oppslagsverk-for-personvern-i-forskning/personvernulempe>
 4. <https://meldeskjema.nsd.no/test/>

ningane mellom det deskriptive og normative planet skal vere relevante, held diskusjonane i denne artikkelen seg innanfor den norske forskingsverda.

For det andre drøftar artikkelen korleis etiske normer og reglane for personvern blir praktiserte i *nyare* forskning. Under lupa ligg derfor berre studiar publiserte dei siste fem åra. Målet med ein kort historisk horisont er at diskusjonane skal vere ferskvare for språkforskarar som planlegg nye undersøkingar med data frå sosiale medium. Sjølv med ein så kort horisont er utfordringa at normative tekstar og relevante lover kjem i stadig nye versjonar. Alle dei seks undersøkte studiane er gjennomførte etter at NESH la fram nye *Forsknings-etiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi* (NESH 2016). To av forskarane (Folkvord 2020, Morland 2020) har hatt den siste versjonen av *Forskningsetisk veileder for internettforskning* (NESH 2018) tilgjengeleg, mens dei fire andre (Bröske 2017, Lande 2017, Hersdal 2018, Strand 2018) har hatt den eldre versjonen (NESH 2014). Sameleis har dei to nyaste studiane (Folkvord 2020, Morland 2020) data som er samla inn under den nye personopplysingslova (2018), inkludert den europeiske personvernforordninga GDPR. Dataa for dei fire eldste studiane (Bröske 2017, Lande 2017, Hersdal 2018, Strand 2018) er samla inn under den gamle personopplysingslova (2000). Her forenkla vi perspektivet og presiserer dette: Fordi artikkelen skal vere eit verkøy for *framtidig forskning*, brukar vi dei siste etiske retningslinjene og den nye personopplysingslova (2018) i diskusjonen. Det tillet vi oss særleg fordi alle dei seks studiane leverer gode eksempel på korleis ein kan løyse dilemma i tråd med dei siste etiske retningslinjene og den nye personopplysingslova. For eksempel skal vi i del 3 sjå korleis ein av dei eldste studiane (Lande 2017) reflekterer på ein måte som møter definisjonen av ei personopplysing i den nye lova svært godt.

Det tredje kriteriet i valet av avhandlingar som kjem under lupa, handlar òg om relevans. Språkforskaren er jo ikkje er den einaste på jakt i sosiale medium. Det er ikkje vanskeleg å tenke seg at for eksempel helsevitenskapen og valforskninga vil finne interessante data på desse digitale plattformene. Forskarar frå ulike fag må planlegge prosjekta sine etter både overordna juridiske praksisar, forskningsetiske prinsipp og rettleiingar for fagområdet. For at denne artikkelen skal bli eit godt verkøy for språkforskarar, er han snevra inn til å handle om dette fagfeltet spesielt. Av same grunn er juridiske termar og utgreiingar gjort så enkle som råd er, når dei samtidig skal vere presise nok.

Den neste delen av artikkelen presenterer metoden og datagrunnlaget for denne studien. Dei tre forskningsspørsmåla får deretter kvar sin del av artikkelen, altså del 3 til 5, og desse blir samanfatta i del 6. Når denne artikkelen presenterer

forskingsetisk refleksjon og praksis, ligg vekta på vurderingar og val som ligg tettast på normene frå NESH (2016, 2018) og måten NSD praktiserer personvernreglane på. Artikkelen vurderer altså ikkje systematisk den enkelte studien på kvart av forskingsspørsmåla. Målet er nemleg å «bidra til refleksjon som skal føre til forsvarlege framgangsmåtar og gode grunngevingar for metodeval» (Hårstad 2020: 142). Då har det størst verdi å vise fram dei stødigaste stega på dei etiske snublesteinane.

2. Metoden og datagrunnlaget

Datagrunnlaget for denne studien er avhandlingane frå Bröske (2017), Lande (2017), Hersdal (2018), Strand (2018), Folkvord (2020) og Morland (2020). Dei seks avhandlingane har sjølvsagt langt fleire etiske refleksjonar og interessante funn enn dei vi kan presentere innanfor rammene for denne artikkelen. I kollegialt sinnelag kan vi derfor berre oppfordre interesserte til å lese meir av avhandlingane. Dei korte presentasjonane under har eit selektivt utval av informasjon som er relevant for dei vidare diskusjonane om forskingsetikk.

Skriftlege dokument som ei forskingsavhandling har ifølge Jacobsen (2005: 164) «mer gjennomtenkt og bearbeidet» informasjon enn munnlege kjelder. Lesaren får gjennomarbeidde refleksjonar og konklusjonar frå forfatarane bak dokumentet, i dette tilfellet etiske diskusjonar og val frå seks forskarar på språkjakt i sosiale medium. Derfor er dokumentundersøkinga ein eigna metode for å svare på dei tre forskingsspørsmåla over.

Jacobsen (2005: 165) minner likevel om at dokumenta er «ferdig transformerte data», og det er ei potensiell svakheit med dokumentundersøkinga som metode. Sjølvsagt kan òg tankane, tvilen og motargumenta som ikkje er formulerte i avhandlinga, vere interessante data i ein studie om forskingsetikk. Derfor kan det vere interessant å supplere denne dokumentundersøkinga med kvalitative intervju med dei same eller fleire forskarar om vegen fram til vala som dei presenterer i avhandlingane sine. Det fell likevel utanfor rammene for denne undersøkinga.

2.1 Innhaldet i private og offentlege profilar på Instagram

Bröske (2017) undersøker korleis innhaldet i private og opne innlegg på Instagram skil seg frå kvarandre. På denne billedelingstenesta kan nemleg brukarane ha ein open eller ein lukka profil. Forskjellen er om andre brukarar må be om å få følge profilen før dei kan sjå og respondere på bilda eller videoane (Bröske 2017: 6). For å dele bilde eller videoar, eller for å respondere

på andre sine bilde, må brukarane ha sin eigen profil der brukarnamnet avgjer kor anonym ein er. Undersøkinga til Bröske (2017) har to kvinnelege informantar, begge med ein privat og ein offentleg profil på Instagram. Datamaterialet er fire multimodale tekstar henta frå Instagram, ein tekst frå den lukka og ein annan frå den opne profilen til kvar av dei to informantane. Bröske (2017: 56) samanfattar studien blant anna i omgrepet «sjølvframstilling» og seier at «bevissthet med tanke på publikum og mottaker, er med på å påvirke det som publiseres på Instagram» og at praksisen er ulik på dei offentlege og private profilane.

2.2 Samtalestilen i ein lukka gruppechat

Studien til Lande (2017) er derimot frå ei heilt lukka plattform: Facebook Messenger.⁵ Lande undersøker samtalestilen i ein gruppechat og korleis denne stilen bygger opp under den sosiale funksjonen til gruppechatten (Lande 2017: 5). Dei ni informantane, inkludert forskaren sjølv, har skrive 4 945 meldingar i løpet av eit halvt år i 2016, og det er desse meldingane som blir datagrunnlaget for undersøkinga. Til forskjell frå det ein Facebook-brukar publiserer på den såkalla veggen hos seg sjølv eller andre, er meldingane på Messenger skjulte for alle andre enn dei du samtalar med. Lande (2017: 101) konkluderer med at språkbruken i gruppechatten er variert og at «[d]eltagerne bruker ulike stilistiske ressurser med forskjellige intensjoner», blant andre å uttrykke ekspressivitet eller identitet. Lande brukar vidare termene «talespråknær» og «uformell» om den språklege stilen og «uregulert» og «liberalt» om det vennegruppegenererte normfeltet i praksisfellesskapet (Lande 2017: 102). Mot slutten poengterer Lande (2017: 103) at det er «mange fellestrekk mellom både språkbruken og det sosiale livet på og utenfor nettet. En naturlig slutning blir at språkbruken online og offline overordnet sett bør sees på som ekvivalenter i forhold til hverandre.»

2.3 Talemålsnær skriving og identitet

Studien til Strand (2018) er frå den same digitale plattformen og spør kva plass den talemålsnære skrivinga har på Messenger hos fire unge trondheimarar. Dei fire informantane i studien, tre kvinner og ein mann, er mellom 20 og 25 år og har vakse opp i Trondheim. Informantane har sjølve valt ut og sendt forskaren

5. Lande (2017) brukar ikkje sjølv nemninga Facebook Messenger om plattformen der gruppechatten går føre seg. Facebook har ingen andre applikasjonar for lukka samtalar som dei Lande har studert. Derfor er det meir enn rimeleg å tru at dataa er henta frå Facebook Messenger, og det er denne plattformen som blir definert over.

til saman 30 samtalar frå november 2017. Datagrunnlaget for studien er dei 224 meldingane i desse samtalan. Studien viser at informantane brukar pronomen, spørjeord og apokope for å markere dialekt i skrift på Messenger, men samtidig held dei seg i stor grad til bokmålstandarden. Strand (2018: 55) konkluderer blant anna med at dialektskriving kan vere eit verktøy for å opptre autentisk eller vise tilhøyrse til den trønderiske gruppa, sameleis som bokmål kan verke til «å tone ned sin trønderiske identitet.»

2.4 SKAM-prat i ei Facebook-gruppe

Hersdal (2018) undersøker om ungdomsserien SKAM på NRK har hatt ein positiv effekt for forståinga av norsk hos dei danske sjåarane. Studien viser særleg korleis medlemmane i Facebook-gruppa *Kosegruppa DK* «forholder seg til, forhandler om og diskuterer det norske språket» i og utanfor serien (Hersdal 2018: 2). Ei viktig presisering her er at ei Facebook-gruppe ikkje er det same som ei chattegruppe på Facebook Messenger. Privatpersonar og verksemder kan gå saman i ei gruppe på Facebook. Ei slik Facebook-gruppe kan vere open for alle eller lukka inntil nokon i gruppa aksepterer ein førespurnad om å få vere med i gruppa. I gruppa utvekslar medlemmane innhald om ei felles sak eller interesse, og innhaldet er synleg for alle i gruppa.

Facebook-gruppa *Kosegruppa DK* hadde i april 2018 fleire enn 43 000 medlemmar. Det kjem ikkje fram hos Hersdal (2018) kor mange av dei desse som blir informantane for studien. Etter eit tekstsøk med bestemte søkeord blir datagrunnlaget for studien til Hersdal (2018) 13 innlegg og 30 kommentarar. «I tillegg kommer en del direkte sitat henta fra kommentarfeltet i gruppa», skriv Hersdal (2018: 35). Derfor er det uråd å avgjere den endelege storleiken òg på datamaterialet. Hersdal (2018: 69) samanfattar studien med at «danske nasjonens fasinjon og interesse for å lære det norske språket er minst like gjen-nomtremgengende som nordmennenes engasjement og iver etter å lære bort og for-klare.» Ein av konklusjonane er derfor at studien gir «en solid støtte for antakelsen om at SKAM har påvirket den norske språkforståelsen i Skandinavia generelt, og kanskje Danmark spesielt» (Hersdal 2018: 88).

2.5 Normer og funksjoner for talemålsnær skiving på Messenger

Folkvord (2020) hentar òg data frå Facebook Messenger i studien som spør kva normer og funksjonar ein kan finne for den talemålsnære skivinga hos unge trondheimarar og til kven desse språkbrukarane skriv talemålsnært og standardnormert. Til slutt spør studien om talemålsnær skiving varierer med kjønn og utdanningsnivå. Studien har åtte informantar mellom 21 og 26 år og

som har vakse opp og framleis bur i Trondheim. Halvparten er menn og halvparten kvinner. Vidare har halvparten høgare utdanning, mens den andre halvparten ikkje har det. Kvar av informantane har sendt forskaren 12 samtalar frå Messenger, fordelte på til saman 375 skjermdumpar. Folkvord (2020: 38) anslår datamaterialet til å omfatte 1 375 meldingar frå informantane. Ein av konklusjonane frå Folkvord (2020: 103) er at talemålsnær skriving er vanleg hos informantane. Til liks med Bröske (2017) knyter Folkvord (2020) den talemålsnære skrivinga til identitet og autentisitet både på individ- og gruppenivå. Dei sosiale variablane kjønn og utdanningsnivå ser derimot ikkje ut til å påverke bruken av talemålsnære trekk.

2.6 Kodeveksling, emojiar og emotikon på Facebook Messenger⁶

Morland (2020) er den siste av dei seks avhandlingane i datagrunnlaget for denne studien og den fjerde med data frå Facebook Messenger. Den todelte problemstillinga i studien spør først «i hvor stor grad engelsk blir brukt i Facebook Messenger-samtaler og om det er andre språktrekk som er karakteristiske for disse tekstene» og deretter «hvilke funksjoner kodevekslingen, emojiene og emotikonene har i utsagnene [...]». Studien analyserer 11 samtalar med to samtalepartar i kvar. Sidan éin av språkbrukarane er med i to samtalar, er det til saman 21 informantar i studien. Desse er mellom 19 og 30 år, femten kvinner og seks menn. Dei 11 samtalane har til saman 13 519 ord, som blir datagrunnlaget for studien. Kodevekslinga mellom norsk og engelsk utgjer omlag tre prosent av datamaterialet i undersøkinga til Morland (2020: 88). Han finn òg mange forkortingar, utelatingar, emojiar, emotikon og eksempel på talemålsnært skriftspråk i materialet. Morland konkluderer blant anna med at informantane kodevekslar for å gjere meldingane kortare og raskare å skrive eller for å gjere budskapet meir presist. Emojiene og emotikona har såkalla paralingvistiske funksjonar ved at dei markerer eller forsterkar haldningar eller følelsar hos informantane.

Dei seks forskarane som no er presenterte, har implisitt eller eksplisitt definert ei personopplysing. Forskarane har òg avgjort om informantane skal gi eit informert samtykke, og dei har behandla informasjon om tredjepartar. Det første av desse tre spørsmåla er tema for den neste delen av artikkelen.

6. Emojiar er dei små teikningane, særleg av ansikt, som er blitt vanlege i det digitale språket (sjå for eksempel Bakkevoll 2019). I dei noko eldre emotikona er ansiktet og følelsen det skal uttrykke, bygd opp av skiljeteikn. Eit godt, norsk omgrep er etter vårt syn 'uttrykksikon', som blant andre Røyneland og Vangnes (2020) brukar.

3. Identifisering av informantar

For å forvalte sentrale forskingsetiske prinsipp på ein god måte treng språkforskaren ein presis definisjon av det grunnleggande omgrepet *personopplysing*. Kunnskap om dette avgjer kva for etiske og juridiske rammer forskaren tilpassar prosjektet etter. Ei personopplysing er *alle opplysingar* om ein fysisk person eller som gjer at ein kan kjenne ein fysisk person igjen (personopplysingslova 2018).^{7, 8} Namn, bilde, e-postadresse og IP-adresse er eksempel på slike opplysingar, men lista er ikkje uttømmende.

3.1 Omgrepa anonym og aidentifisert

I starten av prosjektet er det viktig å avklare om datamaterialet vil innehalde personopplysingar, som forskaren eventuelt kan aidentifisere undervegs. Denne avklaringa avgjer kva etiske normer og ikkje minst juridiske rammeverk prosjektet sorterar under.

Det går nemleg eit viktig skilje i forskingsetikken og jusen mellom *anonyme* og *avidentifiserte* data og informantar. Når sjølv ikkje forskaren kan kople personar og data i innsamlinga eller analysen, er desse *anonyme*. NESH (2016: 17) presiserer at dette ikkje er det same som at forskaren *avidentifiserer* opplysingane under innsamlinga, i analysen eller før publisering. Då tar forskaren vekk for eksempel namnet eller erstattar det med ein kode eller eit pseudonym. Også NSD presiserer at slike aidentifiserte data ikkje er det same som anonyme data og at prosjektet er anonymt berre «dersom det ikke er mulig å identifisere enkeltpersoner *på noe tidspunkt av datainnsamlingen*» (vår utheving).⁹ Berre i slike tilfelle treng ikkje forskaren å melde prosjektet til og rådføre seg med NSD.¹⁰

Denne diskusjonen er meir enn teoretisk avklarande. Det har praktiske konsekvensar og etisk verdi at norske språkforskarar som vil jobbe i sosiale medium, kjenner denne viktige forskjellen mellom anonyme og aidentifiserte

-
7. Her og vidare er (personopplysingslova) brukt for å vise både til den norske lova og den europeiske personvernforordninga (GDPR), som gjeld som norsk lov og er integrert i den norske personopplysingslova. Referansane til paragrafar, kapittel og artiklar kan bli så lange at dei går ut over flyten i teksten. Derfor har artikkelen fotnotar for å vise til den nøyaktige staden i lova og personvernforordninga.
 8. <https://lovdata.no/lov/2018-06-15-38/gdpr/a4>
 9. <https://www.nsd.no/personverntjenester/oppslagsverk-for-personvern-i-forskning/hvordan-gjennomfore-et-prosjekt-uten-a-behandle-personopplysninger>
 10. Meldinga til NSD oppfyller kravet i personopplysingslova om at forskingsinstitusjonar og andre verksemder skal ha protokoll for behandling av personopplysingar. Nokon verksemder har interne løysingar for å oppfylle dette kravet, men dette ein del av personverntenestene som NSD leverer til dei fleste norske forskingsinstitusjonane.

data og informantar. Presiseringa frå NSD inneber at det vil vere vanskeleg for ein språkforskar å gjennomføre eit prosjekt i sosiale medium med anonyme informantar slik dette er definert over. Så lenge plattformar opnar for språkleg interaksjon innanfor eit fellesskap (Enjolras, Karlsen, Steen-Johnsen & Wollebæk 2013: 11, Hårstad 2021: 24) vil òg anonyme språkbrukarar vere i dialog med språkbrukarar med namn og bilde på profilane. Forskaren kan avidentifisere også desse språkbrukarane før publisering. Studien er likevel ikkje fullstendig anonym sidan forskaren på eit tidpunkt i prosjektet kan identifisere enkeltpersonar og kople bestemte data til bestemte informantar.

I denne studien samlar alle dei seks forskarane inn opplysingar som kan identifisere informantar og tredjepartar, hovudsakeleg med namn og bilde på profilane i sosiale medium. Ingen av prosjekta behandlar altså anonyme data, sjølv om alle heilt eller delvis avidentifiserer dataa før publisering. Så har også fem av dei seks forskarane meldt prosjektet og gjennomført det etter rettleiing frå NSD (Brøske 2017, Lande 2017, Strand 2018, Folkvord 2020, Morland 2020). Studien tyder altså på at dei fleste norske språkforskarar *i praksis* er kjende med forskjellen mellom anonyme og avidentifiserte informantar, og konsekvensen dette har for forholdet til NSD. Tre av forskarane *formulerer* også dette skiljet presist.

Eit eksempel på ei presis, etisk formulering kjem frå Lande (2017: 36). Han konkluderer med at «[s]iden direkte og indirekte personidentifiserende opplysninger *er samlet inn*, måtte prosjektet meldes inn til NSD – Norsk senter for forskningsdata» (vår utheving). Som diskusjonen i del 3.3 vil vise er det ikkje mogleg å identifisere informantane i studien til Lande om ein limer inn tekstutdrag på søkemotorar på internett. Likevel har altså Lande ikkje vurdert prosjektet som anonymt, men meldt det inn til NSD. Eit anna eksempel på presis bruk av omgrepa kjem frå Morland (2020: 44): «Jeg vurderte informantene som avidentifisert selv om jeg beholdt alder og kjønn i materialet.» Det siste eksempelet kjem frå Brøske (2017: 34) som brukar omgrepet «avidentifisere» om å sladde ansikt og delar av brukarnamna til informantar og tredjepartar.

Til samanlikning skriv Bakkevoll (2019: 101) dette til informantane i studien sin: «[...] presiseringen er at jeg verken bruker navn, bilde eller noen annen informasjon som identifiserer deg. Tvert imot blir alle slike opplysninger anonymisert gjennom hele forskingsprosjektet.» Sjølv om omgrepet «avidentifisert» ville vore meir korrekt her, skriv NSD i prosjektvurderinga til Bakkevoll (2019: 100) at «[i]nformasjonsskrivet vi mottok på epost 06.02.2018 er godt utformet.» Konklusjonen i denne diskusjonen kan derfor vere at både Kong Salomo og Jørgen Hattemakar, altså både NSD og masterstudenten, kan tene

på å vurdere om dei eigentleg meiner *avidentifiserte* når dei skriv om anonyme data og informantar.

3.2 Personidentifikasjon når personopplysingar blir kombinerte

I den tredje fasen av eit forskingsprosjekt, altså publiseringa, treng forskaren å ta høgde for nokon særtrekk ved informasjonen på internett, samanfatta av Boyd (2008: 27). Det første av desse særtrekka er at det er lett å søke seg fram til innhaldet i den nettbaserte offentlegheita. Eit anna er varigheit, altså at ytringar automatisk blir registrerte og lagra. Reduplikasjon er eit tredje trekk, noko som betyr at det digitale innhaldet svært enkelt kan bli kopiert. For å ta høgde for desse særtrekka må forskaren vurdere to viktige presiseringar til definisjonen av ei personopplysing.

Den eine presiseringa er at forskaren må ta seg i vare for å *kombinere* opplysingar slik at ein kan kjenne igjen ein person.¹¹ To av dei seks forskarane i studien (Folkvord 2020, Morland 2020) viser ei slik utvida forståing av kva som kan vere identifiserande personopplysingar. Folkvord (2020: 48) seier at «[i] enhver studie hviler det et ansvar på forskeren om å ivareta personvernet til informanten» og at dette betyr «at det ikke er opplysninger i det publiserte som gjør at man gjennom å sette dem sammen kan identifisere informantene.»

Sjølv om han ikkje formulerer det like eksplisitt, ser særleg Morland (2020: 45) ut til å ha tatt høgde for dette i avhandlinga si. Han går ikkje lengre enn å behalde kjønnet og alderen på informantane i materialet han publiserer, sjølv om han har samla inn og kunne brukt fleire opplysingar. Berre tre av informantane hans er ikkje studentar ved Universitetet i Bergen. Den eine går på NTNU i Trondheim og dei to andre på Noroff. I tillegg kjem ti av informantane frå andre fylke enn Vestland, men få kjem frå kvart fylke: tre frå Agder, to frå Oslo, ein frå Rogaland, ein frå Trøndelag og ein frå Viken. Han har derfor ikkje med opplysingar om kor informantane studerer eller kor dei kjem frå. Slik innskrenkar han moglegheitene for å sette saman opplysingar slik at kombinasjonen berre passar til få eller eitt individ.

3.3 Personidentifikasjon på grunn av språklege element

Den andre presiseringa til definisjonen av ei personopplysing kjem frå lovverket, som seier at ei personopplysing òg kan vere «elementer som er spesifikke for nevnte fysiske persons [...] genetiske, psykiske, økonomiske,

11. <https://lovdata.no/lov/2018-06-15-38/gdpr/a4>

kulturelle eller sosiale identitet» (personopplysningslova 2018).¹² Det kan vere relevant for sosiolingvistar å vurdere om element frå språkstilen, altså «a repertoire of linguistic forms associated with personas or identities» (Bucholtz & Hall 2005: 597) aleine eller saman med andre opplysingar kan vere med på å identifisere enkeltindivid.

Det er for eksempel lett å tenke seg at syntaktiske konstruksjonar i talemålsnært, skriftleg språk kan peike på at informanten kjem frå eit område med språkkontakt (sjå for eksempel Nilsen 2000). Sameleis kan kodeveksling vise kva språkleg eller kulturell bakgrunn informanten har. Eit sentralt poeng her er at ein ikkje treng å vere språkforskar for å fange opp slike trekk. For eksempel har Nilsen (2000: 129-130) vist at innbyggerane i Reisadalen sjølve er bevisste på at dialekta deira har ein særskilt syntaks.

For å illustrere denne etiske utfordringa ytterlegare kan vi gå frå dei hypotetiske eksempla til eit konkret der Bakkevoll (2019: 2-3) siterer språkbrukaren i den politisk orienterte Facebook-gruppa «Bevar Andøya Flystasjon». Kanskje i affekt skriv språkbrukaren «Forbanna mannskit bekkola» som ein kommentar til at fleirtalet på Stortinget blei einige om å legge ned flybasen med om lag 300 arbeidsplassar på Andenes i Nordland. Ordet «bekkola» er i avhandlninga forklart som ein variant av det finske banneordet perkele. Ytringa er berre tre korte ord som kommenterer eit anna innlegg, så ein søkemotor på internett gir inga hjelp om ein vil finne identiteten til denne språkbrukaren. I eit språksamfunn langt frå områda med finsk språkkontakt har dette særeigne ordet likevel eit stort potensial til å identifisere språkbrukaren, dersom det er slik at ho eller han held til på Andenes.¹³ Det er ikkje uproblematisk at Bakkevoll (2019) på denne måten brukar eit språkleg element, altså eit finsk banneord, som kan identifisere personen bak eit følelsesladda innlegg i ein betent politisk debatt.

Både dei hypotetiske og det konkrete eksempelet over gir eit vink om at forskarar i norske språksamfunn bør vere oppmerksame på at språklege element kan identifisere informantane. Valfridom har ifølgje Vikør (2007: 178) vore eit særmerke for norsk språkplanlegging, slik at standarden i begge dei to skriftspråka våre er vide. I tillegg kjem den lange tradisjonen for å respektere og bruke dialekt i mange eller alle brukssituasjonar (Røyneland 2017: 97, Vikør 2018: 348). Det kan vere årsaken til den digitale dialektskrivinga som stadig

12. <https://lovdata.no/lov/2018-06-15-38/gdpr/a4>

13. Sidan kommentaren er i ei Facebook-gruppe, kan det sjølvsagt hende at språkbrukaren bak ordet ikkje høyrer til på Andøya. Likevel vil sitatet vere problematisk fordi språkbrukaren har meldt seg inn i den lukka gruppa og slik antyda tilhørsle til staden.

fleire norske studiar dokumenterer (Evjen 2011, Rotevatn 2014, Brøske 2017, Lande 2017, Strand 2018, Vangsnes 2019, Bakkevold 2019, Folkvord 2020, Røynealand & Vangsnes 2020).

Norske språkbrukarar kan altså velje mellom fleire former, både frå dei vide standardane i skriftspråka og i den ustandardiserte, talemålsnære skrivinga. Aleine eller i kombinasjon kan desse formene bli individuelle markørar, vise tilknytning til geografiske stadar eller på annan måte bli «elementer som er spesifikke for nevnte fysiske persons [...] identitet» (personopplysingslova 2018).¹⁰

Det er viktig å skilje denne *språklege moglegheita* (A) for at sjølv korte språkelement kan identifisere ein person frå den *teknologiske moglegheita* (B) som søkemotorar gir til å knytte tekst til språkbrukaren. Språkdataa frå sosiale medium er elektronisk mediert språk (Hougaard & Balleby 2019: 8). Denne kombinasjon av språk og teknologi i sosiale medium gir nokon premisser for språkforskaren som brukar slike data. Vi skal raskt sjå på fire av dei som blir viktige for den vidare diskusjonen.

- 1) Moglegheit B baserer seg på teknologi og kan oppstå berre når forskaren brukar tekst frå opne forum.
- 2) Moglegheit A baserer seg derimot på språk og er alltid til stades, altså anten forskaren brukar tekst frå opne eller lukka forum.

Av 1 og 2 følger den viktige premissen 3:

- 3) Sjølv om teknologien sperrar for moglegheit B i lukka forum, må språkforskaren framleis vurdere moglegheit A i publisering av forskning på språk frå slike forum.

Ein siste premiss er kanskje sjølv sagt, men derfor òg lett å gløyme:

- 4) Forskaren kan ikkje påverke teknologien, altså tilgangen til forumet. Av omsyn til validiteten i prosjektet bør forskaren heller ikkje påverke språket i datamaterialet. Den einaste måten å ta høgde for dei tre første premissane er dermed å vurdere *mengda* av og *innhaldet* i sitat ein tar med i ein publikasjon.

Premiss 2 seier at moglegheit A alltid er til stades. Likevel er ho overraskande lite diskutert i den norske metodelitteraturen og dei generelle normative tekstane som denne studien krinsar rundt (NESH 2016, 2018).

Mens moglegheit B får god merksemd, kan det sjå ut som moglegheit A framleis ligg i ei forskingsetisk blindsona. Denne ubalansen inneber at premiss 1 på ulike måtar blir artikulerte i forskingsetiske refleksjonar, mens ein finn svært få formuleringar som handlar om premissane 2 og 3. For eksempel skriv NESH (2018: 16) at «[i]nformasjon og ytringer på nettet lagres digitalt, slik at deltakere og informanter ofte kan identifiseres *ved søk*» (vår utheving). Nesvold, Sjørseike og Wadel (2014: 116) skriv at «ved bruk av sitater fra diskusjoner på nettet bør forskeren ta hensyn til at det kan være mulig å spore informantens identitet ved å foreta fulltekst søk på sitatene.» Hårstad (2020: 151) presiserer at det er

viktig å hugse at ein tekststreng i seg sjølv kan vere identifiserande òg etter at han vore anonymisert, fordi han kan søkast opp i digitale søkeverktøy. Dermed vil altså teknologiske føresetnader styre kva data som bør attgjevast når forskingsfunna blir formidla.

Det siste eksemplet kjem frå Segadal (2015: 43): «This may be a challenge if one wishes to publish direct quotes, as these will be searchable on the Internet.» Som vi ser opnar ingen av desse for at eit språkleg element kan vere ei identifiserande opplysing òg om ein ikkje kan finne sitata på internett.

Lande (2017) formulerer fleire bidrag til at den norske språkforskinga kan komme ut av denne blindsona. Han peikar på at det ikkje gir meining å skilje mellom online og offline liv (sjå elles Stæhr 2015: 154, Lüders 2015: 77 og Hårstad 2021: 27) og at det òg gjeld språkbruken: «Siden jeg selv er deltager i gruppechatten, ser jeg at det i mange tilfeller er tydelig at konkrete språklige valg blant deltagerne, samsvarer med typisk muntlig språkbruk offline» (Lande 2017: 23). Til liks med NESH (2018: 16), Nesvold, Sjørseike og Wadel (2014: 116) og Segadal (2015: 43) set Lande ord på premiss 1. Det nye er at han forlenger resonnementet til å inkludere premissane 2 og 3. I sitatet under er det ikkje berre ein grupped medlem som kan kjenne ein informant på språket. Her ser vi at Lande opnar for at òg utanforståande lesarar kan gjere det:

En fordel med at gruppa er privat, i forskningsetisk sammenheng, er at det ikke er mulig å søke opp eksempler som er brukt i oppgaven ved hjelp av Google eller andre søkemotorer. Søkemotorer kan dermed ikke benyttes for å koble

meldinger til bestemte informanter eller gruppa som helhet. Likevel er det en risiko for at en leser vil kunne være i stand til å koble språkbruken til gruppechatten eller et av medlemmene dersom leseren kjenner én eller flere deltakere svært godt. Dette har blant annet sammenheng med den nevnte gjensidigheten mellom språkbruk på og utenfor nettet [...]. Det kan også ha sammenheng med at de ulike deltakerne deltar i andre praksisfellesskap på nettet hvor språket til dels brukes på lignende måter. (Lande 2017: 35)

Her plasserer Lande (2017) seg sjølv inn i tabell 1. Denne tabellen gir eit skjematisk oversyn over kva av moglegheitene A og B (over) som forskarane i denne studien bør ta høgde for når dei publiserer resultatane.¹⁴

Tabell 1: *Moglegheiter for identifisering av informantane.*

	Aktuell moglegheit for identifisering	
	A	B
Brøske (2017)	•	•
Lande (2017)	•	
Hersdal (2018)	•	•
Strand (2018)	•	
Folkvord (2020)	•	
Morland (2020)	•	

Moglegheit A er altså felles for alle avhandlingane i denne studien, og ho kan sjå ut til å ligge i den forskingsetiske blindsona. Då blir det spesielt interessant å diskutere korleis forskarane har møtt denne utfordringa. To av dei (Lande 2017, Folkvord 2020) grip tak i dilemmaet sterkare enn dei fire andre. I diskusjonen rundt desse to publikasjonane blir premiss 4 sentral. Denne premissen sa at språkforskaren kan vurdere både *mengda* av og *innhaldet* i sitat ein tar med i ein publikasjon, sidan dette påverkar korleis forskaren tar vare på personvernet og privatlivet til informantane.

Av eit samla datamateriale på 4 945 meldingar siterer Lande (2017) 156 (3 prosent) av dei som eksempel. Resultata blir først og fremst samanfatta i tabellar, og sitata supplerer desse. Det samla materialet ligg heller ikkje med som vedlegg til publikasjonen. Folkvord (2020) har valt den same strategien: Han brukar først og fremst aggregerte resultat i tabellar for å framstille funna, og han illustrerer dette ved å vise 8 (2 prosent) av 375 skjermdumpar. Desse

14. Til liks med Hersdal (2018: 39) og NESH (2018: 10) reknar vi ei Facebook-gruppe med 43 000 medlemmar som eit ope forum, sjølv om språkbrukarane må spørje om å få bli med.

skjermdumpane inneheld 20 (1,5 prosent) av dei anslåtte 1 375 meldingane i datamaterialet. Heller ikkje i publikasjonen til Folkvord ligg heile datamaterialet som vedlegg.

Dei andre publikasjonane i denne studien publiserer mellom 27 og 100 prosent av det samla datamaterialet som eksempel i teksten eller som vedlegg. Til den vidare diskusjonen om sensitivt innhald kan det vere greitt å ha med seg at sitata i desse publikasjonane handlar om blant anna one night-stands, haldningar til nabospråk og kva renter informantane har på bustadlåna sine.

På dette punktet i diskusjonen er det verdt å minne om at forskning er ein legitim aktivitet med stor verdi for samfunnet (NESH 2016: 9, Elgesem 2015: 14). Det gjeld òg forskning som hentar datamaterialet sitt frå internett. Slike data er følge Segadal (2015: 35) «increasingly valuable resources for researchers across many disciplines.» Nesvold, Sjurseike og Wadel (2014: 117–118) minner derfor om at målet ikkje kan vere at talet på sitat i publikasjonar frå slik forskning skal vere lågast mogleg:

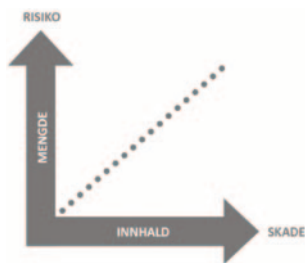
dersom [...] alle sitater og alle nettdiskusjoner anonymiseres og gjøres ugjenkjennelige, reduseres også troverdigheten til og relevansen av innholdet i prosjektet. Utgangspunktet var kanskje å si noe om en samfunnsprosess, en gruppe mennesker eller et tema innenfor en særskilt kontekst, men ved å anonymisere i for stor grad ender man opp med vage og lite utsagnskraftige data.

Alle forskarar må dokumentere funna sine. Det gjeld uansett problemstilling, talet på informantar og tekniske tilhøve på språkbruksarenaen dataa er henta frå. Sitat og eksempel frå datamaterialet er ein essensiell del av dokumentasjonen for at språkforskning i sosiale medium skal vere truverdig og kunne etterprøvast. Det er som Segadal (2015: 44) seier: Mellom dei klare eksempla ligg det ein spekter av gråsoner som må vurderast kvar for seg.

Som eit kompass i gråsonene seier premiss 4 (over) at forskaren ikkje berre må vurdere mengda av sitat i publikasjonen, men òg innhaldet i desse. Her tar Folkvord (2020: 48, 111) eit interessant grep når han ber informantane unngå tema som helse, seksuell legning, straffedommar, politisk oppfatning og religiøs tilhøyrsløse i samtalanane dei sender han. Vi har allereie sett at Folkvord samtidig held talet på sitat lågt i publikasjonen. Dermed har han redusert både risikoen for at språket skal identifisere informanten og den potensielle belastinga om det likevel skjer. Det siste er ikkje mindre viktig enn det første: «God og ansvarlig forskning innbefatter også vurderinger av utilsiktete og uønskete konsekvenser» (NESH 2016: 11).

Ei rimeleg innvending til grepet som Folkvord (2020) tar når han ber informantane unngå visse tema, er at han påverkar sitt eige datamateriale.¹⁵ I ein studie om korleis talemålsnær skrivning kan markere individuell identitet og gruppetilhøyrse er det relevant å spørje korleis dette hadde sett ut dersom informantane òg hadde snakka om for eksempel politikk eller religion. Folkvord (2020) kunne ha redusert den potensielle belastinga for informantane på ein annan måte: Om tema for samtalen var opp til informantane, kunne Folkvord framleis brukt heile datamaterialet til å vise dei aggregerte resultatane i alle tabellane (sjå for eksempel Folkvord 2020: 52, 65, 70). Når han deretter skulle velje sitat for å illustrere faglege poeng, kunne han unngått eventuelle tema som har potensial for å bli ei belastning eller til skade for informanten.

Figur 1 samanfattar premiss 4 og diskusjonen så langt. Fordi språket i eit datamateriale frå sosiale medium kan identifisere informantane, bør språkforskaren jobbe langs to aksar i dei etiske refleksjonane. Langs den vertikale aksa ligg mengda av sitat i publikasjonen. Her kan forskaren justere risikoen for at informanten blir identifisert. Samtidig bør ikkje mengda vere så lita at det går ut over truverda til forskinga. Langs den horisontale aksa ligg innhaldet i dei utvalde sitata til publikasjonen. Her kan forskaren redusere skaden om informanten likevel blir identifisert. I arbeidet langs den horisontale aksa bør forskaren først og fremst *vurdere* innhaldet i sitata ein tar med i publikasjonen, i staden for å påverke *produksjonen* av dei.



Figur 1: Språkforskaren kan vurdere både mengda og innhaldet i sitata som blir med i publikasjonen frå undersøkinga i sosiale medium.

15. Det er ikkje usannsynleg at Folkvord har tatt dette grepet for å unngå særlege kategoriar av personopplysingar som etter personopplysingslova (2018, §9) utløyser fleire plikter til å vurdere og rådføre seg med personvernombodet om konsekvensane for personvernet og integriteten til informantane.

3.4 Reidentifisering av informantar

Artikkelen har så langt diskutert om kombinasjonen av opplysingar eller språket i sitata kan identifisere informantane bak ein studie. Når språkforskaren vurderer desse presiseringane til definisjonen av ei personopplysing, vurderer ho eller han om ein informant kan bli *reidentifisert* når studien blir publisert: «I noen tilfeller er det mulig å reidentifisere (bakveis-identifisere) individer selv om alle opplysningene i utgangspunktet er aidentifisert» (NESH 2018: 16).

Brøske (2017) er i ei særstilling i denne studien, ettersom brukarnamnet kan reidentifisere i alle fall den eine av dei to informantane (Brøske 2017: 34).¹⁶ NESH (2018: 16) minner nettopp om at brukarnamn ikkje er ein garanti mot reidentifisering, fordi «slike kallenamn ofte kan spores på tvers av ulike data og i ulike sammenhenger.»

Når Brøske (2017) i avhandlingen refererer frå samtalar med informantane sine, kjem det fram at ho er bevisst på dette: «[...] vi snakket flere ganger om at jeg ville bruke materiale som kunne være identifiserbart» (Brøske 2017: 32). I samtykkeskjemaet til Brøske (2017: vedlegg 1) kommuniserer ho det same bodskapet til informantane: «På grunn av oppgavens formål vil du som deltaker i dette forskningsprosjektet kunne gjenkjennes i oppgaven.» Slik blir Brøske (2017) den einaste forskaren i denne studien som i dialogen med informantane sine tar eit ope atterhald om at aidentifiseringa kan vere utilstrekkeleg. Brøske (2017) er aleine om å følge NESH (2018: 16) på dette punktet:

[...] det begrenser samtidig hva slags konfidensialitet forskerne faktisk kan garantere deltakerne. Ved informering og innhenting av samtykke bør forskeren i størst mulig grad forsøke å forklare disse mulige begrensningene ved konfidensialiteten der dette er aktuelt.

4. Samtykke og/eller informasjon

Forskingsetikken (for eksempel NESH 2016: 14-16) held det frie og informerte samtykket fram som ei hovudnorm. Fossheim og Ingierd (2015: 11) kallar det informerte samtykket ein gullstandard, og ifølge Eckert (2013: 14) er det grunnsteinen i etisk forskning. Samtidig opnar forkingsetikken for at forskaren kan vike av frå denne hovudnorma (Segadal 2015, Fossheim & Ingierd 2015: 11). NESH (2016: 16) seier at «unntak kan være aktuelt i situasjoner der forskningen

16. Når ein ved hjelp av brukarnamnet søker opp den private profilen til denne informanten, finn ein personnamnet like under brukarnamnet.

ikke innebærer direkte kontakt med deltakerne, der opplysningene som behandles er lite sensitive, og der forskingen samtidig har en nytteverdi som klart overstiger ulempene som påføres deltakerne.»

Trass denne normative rangeringa i dei forskingsetiske retningslinjene skal forskaren legge merke til at samtykket og retten til å trekke seg frå studien er eitt av fleire likestilte, lovlege grunnlag for å behandle personopplysningar i den nye personopplysingslova (2018).¹⁷ Forsking er eit formål «i allmennhetens interesse» med same status i lovverket som eit samtykke.¹⁸

Uavhengig av kva lovleg grunnlag behandlinga av personopplysingane har, presiserer personopplysingslova at informanten likevel skal ha informasjon om blant anna formålet med behandlinga.¹⁹ Sameleis opnar både Fossheim og Ingierd (2015: 11) og Segadal (2015: 43-44) for at ein kan redusere ulempene for individet ved å informere og gi informanten moglegheit til å trekke seg, men utan at informanten svarar med eit uttrykkeleg samtykke.

Personopplysingslova (2018) opnar òg for eit tredje alternativ der informanten heller ikkje får informasjon, dersom personopplysingane ikkje er samla inn direkte frå informanten og det «viser seg umulig å gi nevnte informasjon eller det vil innebære en uforholdsmessig stor innsats, særlig i forbindelse med behandling for [...] formål knyttet til vitenskapelig eller historisk forskning.»²⁰ Sameleis seier forskingsetikken (NESH 2018: 14) at «det [er] flere utfordringer forbundet med å informere eller innhente samtykke ved internettforskning.» Utan informasjon aukar likevel den potensielle ulempa eller skaden for individet, og det gir auka krav til forskaren om å vurdere risikoen og argumentere for samfunnsnyttan i prosjektet (Seriss 2019: 18). Segadal (2015: 43) presiserer òg at forskaren ikkje bør forlate hovudnorma utan først å vurdere blant anna kor sensitiv informasjonen er, om han er publisert på eit ope eller lukka forum

17. <https://lovdata.no/lov/2018-06-15-38/§8>, <https://lovdata.no/lov/2018-06-15-38/§9>, <https://lovdata.no/lov/2018-06-15-38/gdpr/a6> og <https://lovdata.no/lov/2018-06-15-38/gdpr/a7>

18. «Allmennhetens interesse» som lovleg grunnlag for å behandle personopplysningar må ha eit såkalla supplerande rettsgrunnlag. Det vil seie at behandlinga òg er heimla i eit anna lovverk, og det vil vere personopplysingslova (2018 §§ 8 og 9).

19. <https://lovdata.no/lov/2018-06-15-38/gdpr/a13> og <https://lovdata.no/lov/2018-06-15-38/gdpr/a14>

20. <https://lovdata.no/lov/2018-06-15-38/gdpr/a14>. Standpunktet i denne artikkelen er altså at personopplysningar som er henta inn via sosiale medium, ikkje er samla inn frå den registrerte sjølv og at denne paragrafen då kan komme i bruk. Grunngevinga for dette er todelte. For det første er det inga kontakt mellom forskaren og informanten, og for det andre er dataa publiserte for eit anna formål enn forskning.

og kor sårbare informantane er. Lova og forskningsetikken gir altså forskaren desse tre alternativa:

- 1) Forskaren innhentar eit informert og dokumentert samtykke, som òg gir informantane rett til å trekke seg frå studien til ei kvar tid. Her seier lova at «[d]et skal være like enkelt å trekke tilbake som å gi samtykke.»²¹
- 2) Forskaren informerer informantane om studien og gir dei moglegheit til å utøve rettane dei etter personopplysingslova har til blant anna innsyn, retting, sletting og protest.
- 3) Forskaren informerer ikkje informantane.

Som utgangspunkt kan vi seie at fem av seks forskarar bak arbeida i denne studien har følgd hovudnorma om det informerte samtykket (Bröske 2017, Lande 2017, Strand 2018, Folkvord 2020, Morland 2020). To av desse grunngir valet i tråd med prinsippa om å vurdere tilgangen til plattformar og sensitiviteten i innhaldet. Strand (2018: 12) seier at plattformar «må regnes som privat», mens Lande (2017: 35) seier det slik: «Siden forskningsobjektet mitt er en lukket gruppe, er det liten tvil om at språkbruken som produseres på denne arenaen må regnes som privat, både med tanke på tilgang og innhold.» Det er òg verd å legge merke til at Morland (2020: 39, 100) har delt opp samtykket. Slik gir han informantane moglegheit til å sende inn eit tekstmateriale, men reservere seg frå at desse skal bli brukte i publikasjonen eller lagra i eit korpus for vidare forskning. Dette er i tråd med personopplysingslova (2018) som seier at «[e]t samtykke bør omfatte alle behandlingsaktiviteter som utføres med henblikk på samme formål. Dersom det er flere formål med behandlingen, bør det gis samtykke til alle.»²²

Den sjette studien (Hersdal 2018) har valt alternativet der informanten verken gir eit samtykke eller får informasjon om prosjektet. Hersdal (2018: 38-39) grunngir valet dels med at forumet kan vurderast som ope og dels med «vanskeligheter knytta til det å innhente samtykke i digitale fora». Som Bakkevoll (2019: 33) tar Hersdal opp at folk i sosiale medium kan gi seg ut for å vere ein annan enn dei verkeleg er. Derfor skriv òg NESH (2014: 6) at «forskaren ikkje kan være sikker på at samtykke er innhentet fra den man ønsker å innhente samtykke fra.» I tillegg problematiserer Hersdal (2018: 38) at informasjonen på internett er varig, men han kan vere eit resultat av at ein har deltatt

21. <https://lovdata.no/lov/2018-06-15-38/gdpr/a7>

22. <https://lovdata.no/lov/2018-06-15-38>, se punkt 32 i innledningen av personvernforordningen

flyktig i eit forum: «Dette gjør at det kan være vanskelig å få tak i de personene man ønsker samtykke fra i etterkant.»

Basert på dette kan det vere rimeleg å diskutere om alternativ 2 i realiteten er tilgjengeleg for forskarar på opne forum i sosiale medium med mange og flyktige deltakarar. Då kan vi samanlikne valet til Hersdal (2018) med Bakkevoll (2019), som seier at han valde å informere informantane i sin studie av det digitale språket på «Nordnorsk debatt» under valkampen 2017: «Personvernombudet for forskning (NSD) har avklart at språkbrukarane på «Nordnorsk debatt» skal informerast [...], men det er ikkje nødvendig med aktivt samtykke for å hente datamaterialet [...]» (Bakkevoll 2019: 32).²³ Det går likevel fram at dei tekniske tilhøva på Facebook gjorde det vanskeleg å følge opp denne todelte konklusjonen (Bakkevoll 2019: 34-35). Rådet frå NSD var nemleg at Nordlys burde publisere informasjonen som ein sjølvstendig post på «Nordnorsk debatt» like før datainnsamlinga starta i 2018, altså mange månadar etter at den siste kommentaren i datamaterialet blei publisert. Til det skriv Bakkevoll (2019: 34) at

[a]lgoritmane på Facebook ville truleg eksponert ei slik melding berre til dei som hadde vore aktive på «Nordnorsk debatt» i dei seinaste dagane før ein slik post. Det ville dermed vore reint tilfeldig om nokon av det aktive språkbrukarane i valkampen 2017 var blant dei som såg informasjonen.

«Nordnorsk debatt» er regionavisa *Nordlys* si Facebook-side vigd debatt, men Bakkevoll (2019: 34) problematiserer òg om Nordlys var den riktige avsendaren av informasjonen til informantane, framleis med algoritmane på Facebook som argument:

Ei viktig endring i algoritmane til Facebook kom i januar 2018. Då blei innhald publisert frå private profilar prioritert opp, mens innhald frå profesjonelle verksemdar blei mindre synleg [...]. Konsekvensen er at informasjonen om forskinga ville ha vore mindre synleg om Nordlys hadde publisert han.

Løysinga blei at Bakkevoll sjølv la att ein kommentar på slutten av alle samtalanane på «Nordnorsk debatt» som er med i datamaterialet, og han seier at

23. Denne vurderinga frå NSD var altså gjort basert på den gamle personopplysingslova (2000) og måten bestemmingane om samtykke blei presiserte då.

«[s]jansen for at Facebook varslar dei om informasjonen blir derfor mykje større» (Bakkevoll 2019: 34).

Sterkare enn «mykje større» kan verken Bakkevoll eller NSD konkludere i spørsmålet, og det står att som ei av største etiske svakheitene med språkstudiar på kommersielle, digitale kanalar som Facebook. Algoritmane er løyndomar, mellom anna brukt for å tene pengar ved sal av annonsar for kommersielle verksemdar (Bucher 2012). Ein kan derfor aldri vite kor mange som ser informasjonen ein deler på plattformar. Basert på det oppmodar Bakkevoll (2019: 37) NESH og NSD til å halde seg med oppdatert kompetanse om algoritmane på Facebook og andre digitale språkbruksarenaer:

[S]lik kompetanse [er] sentral for alle som vil informere og innhente samtykke frå deltakarar i forskning i sosiale medium. Eg må understreke at dei første råda eg fekk, ville halde deltakarane uvitande om kva dei var med på eller reserverte seg mot, nettopp fordi desse råda ikkje var baserte på kunnskap om korleis Facebook deler informasjon.

Etter denne samanlikninga med Bakkevoll (2019) kan ein argumentere for at Hersdal (2018) valde det einaste reelle alternativet i studien frå *Kosegrupper DK*. Om forskaren vel å forlate gullstandarden om det informerte samtykket, ser det nok på betre ut på papiret å seie at ein likevel skal informere deltakarene i prosjektet for å ivareta dei forskningsetiske normene, slik Bakkevoll (2019) gjer. Som vi no har sett kan dette i nokon tilfelle vere så vanskeleg at forskaren er ærlegare om ho eller han, som Hersdal (2018), frå starten seier at det er uråd å dempe ulempene for individet med å informere om prosjektet. Vi skal likevel legge merke til at Bakkevoll (2019) berre går utanom norma om eit informert samtykke. Han går ikkje utanom meldepraksisen til NSD når undersøkinga ikkje har anonyme data (sjå del 3.1) eller den såkalla rådføringsplikta som etter personopplysingslova (2018) kan oppstå når ein behandlar særleg sensitive personopplysingar i forskning utan samtykke frå informantane.²⁴

5. Sekundærinformantar

Fem forskarar (Bröske 2017, Lande 2017, Strand 2018, Folkvord 2020, Morland 2020) i denne studien har altså følgd hovudnorma om at informantane skal gi eit informert samtykke for å bli med i undersøkinga. Bildet blir likevel meir

24. <https://lovdata.no/lov/2018-06-15-38/§9>

komplisert når vi spør *kven* som skal få informasjon, gi sitt samtykke og ha rett til å trekke seg frå studien. Det openberre svaret på spørsmålet er informanten. Til denne diskusjonen treng vi derfor å formulere ein definisjon på eit så grunnleggande omgrep i forkingsetikken. Enkelt sagt er informanten den eller dei som bidrar med datamaterialet som forskaren brukar for å undersøke forkingsspørsmåla sine.

Når vi tar ein slik definisjon med inn i sosiolingvistiske undersøkingar i sosiale medium, blir det ikkje like enkelt. I kjernen av sosiolingvistikken ligg samtalen og den språklege interaksjonen, og forkingsspørsmåla krinsar om korleis individ og grupper brukar språk for å konstruere og forhandle om identitet og tilhørslse. Som hos Morland (2020) i denne studien blir då blant anna teoriar om kodeveksling relevante, eit språkfenomen som Holmes (2013: 35) seier «may be related to a particular participant or addressee.» Folkvord (2020: 12-13) på si side brukar akkomodasjonsteoriar for å forklare den talemålsnære skrivinga i studien sin:

Når ulike former for språklig tilpasning skjer i praksis, kalles det *akkomodasjon*. *Konvergens* er navnet på fenomenet når noen legger seg tettere opp mot samtalepartnerens språkbruk [...] *Divergens* er betegnelsen for å distansere seg språklig fra en samtalepartner.

Desse rispa i overflata av teorimangfaldet innanfor sosiolingvistikken er nok for å forstå at it takes minst two to tango og at alle samtalepartane bidrar med data og må reknast som informantar. Hårstad (2020: 153) presiserer at «[i] ein studie som i større grad handlar om trekk ved interaksjonen [...], kjem ein derimot ikkje utanom å ta med data frå andre deltakarar.» «[D]e som er gjenstand for forkingen» er ei formulering frå den normative litteraturen (NESH 2016: 13), og språkforskarar i sosiale medium bør ha eit klart bilde av kor mange og kven som er gjenstand for forkinga i det enkelte prosjektet.

Om vi for eksempel tenker oss ein sosiolingvistisk studie om konvergens i ein skuleklasse, ville forskaren neppe ha informert og fått samtykke frå halvparten av samtalepartane i klassen. Deumert (2014: 27) presiserer at språkforskinga ikkje kan ha ein annan etisk standard når ho undersøker sosiolingvistiske fenomen på digitale språkbruksarenaar: «If we accept that the actual and the virtual are deeply entangled, then the basic ethical principles for offline research should apply to online research [...]» Når sosiolingvisten hentar data frå den digitale samtalen, bør ho eller han derfor vurdere om alle partane i samtalen bør reknast som informantar.

Dette har Lande (2017) kontroll på sidan han studerer ein gruppechat med ni medlemmar og han sjølv er ein av desse. Lande (2017: 35) har henta inn eit informert samtykke frå alle i gruppa, som dermed òg har fått retten til å trekke seg frå studien. Med unntak av internasjonale, profesjonelle fotballspelarar som Marouane Fellaini, Harry Kane og Julian Draxler, og to utanlandske kampsportutøvarar, er heller ingen utanfor samtalan nemnde med namn.

NESH (2018: 13) brukar omgrepet «primærinformant» og antyder med det at det òg finst «sekundærinformantar». Sameleis brukar Morland (2020: 46) omgrepa a-informant og b-informant, og han gir desse to gruppene det same informerte samtykket og moglegheita til å trekke seg frå studien (Morland 2020: 38-39). Dette løyer Morland ved at den eine parten i samtalan er studenten som han har rekruttert direkte. Denne a-informanten får tilsendt eit informasjonsskriv og eit samtykkeskjema som blir signerte av begge samtalepartane. Det er elles interessant å sjå at Morland (2020: 42-43) reserverer omgrepet tredjepart til dei som eventuelt blir *omtala* i samtalan mellom a- og b-informanten, men som ikkje bidrar med data til studien. Namna til desse tredjepartane er sensurerte i publikasjonen til Morland (2020).

Strand (2018: 12) og Folkvord (2020: 49) har ei tredje løysing, der primærinformantane (NESH 2018: 13) eller a-informantane (Morland 2020: 46) sensurerer namna frå dei andre partane i samtalen. Etter bruken av omgrepa i del 3.1 aidentifiserer altså den eine informanten den andre, og forskaren mottar data frå anonymiserte sekundær- eller b-informantar. Sidan forskaren då ikkje har personopplysningar om sekundær- eller b-informantane, treng ho eller han heller ikkje eit informert samtykke frå denne gruppa av informantar. Dette er i tråd med hovudregelen beskrive i innleiinga av del 4. Eit ope spørsmål vil då vere om forskaren som mottar slike anonymiserte data, framleis må sørge for at òg sekundær- eller b-informantane får informasjon om den behandlinga av personopplysningar som er gjort på vegne av forskaren. Sjølv om samtykke ikkje er påkrevd, bør forskaren vurdere om sekundær- eller b-informanten skal få informasjon om dette via primær- eller a-informanten.

6. Oppsummering og avslutning

Avslutninga av førre avsnitt viser at den som leitar etter fasitsvar i forskingsetikken blir skuffa. Det gjeld òg om ein leitar i denne artikkelen, som samanfatar ein studie av seks nyare bidrag til den norske språkforskinga med data frå sosiale medium.

Studien tyder på at norske språkforskarar er bevisste på skiljet mellom *av-identifiserte* og *anonyme* data og informantar i det praktiske arbeidet, men i formuleringane er omgrepa ofte blanda saman. Det kan i sin tur få konsekvensar for den viktige dialogen med NSD. Denne normgivaren - og ressursen - kan ein halde utanfor berre når forskaren aldri eller på nokon måte kan identifisere enkeltpersonar. Dei aller fleste språkstudiar med data frå sosiale medium behandler personopplysningar og må ha det utgangspunktet i dei forskingsetiske vurderingane.

Særleg Morland (2020) viser korleis ein kan unngå å kombinere personopplysningar i publikasjonen. Slik førebygger han at informanten kan reidentifiserast. Språkforskinga generelt, og kanskje den norske spesielt, har ei særskild forskingsetisk utfordring. Denne utfordringa er at språklege element kan fungere som ei opplysning som gjer at ein kan kjenne ein fysisk person igjen. Norske språkbrukarar har stor valfridom innanfor standardane for dei to skriftspråka, og det norske dialektmangfaldet viser seg òg digitalt. Aleine eller saman kan dei mange standardiserte og ustandardiserte formene bli individuelle markørar i den digitale skriftpraksisen. Ei viktig presisering er at utfordringa òg gjeld for forskingsprosjekt med data frå lukka forum, altså der ein lesar ikkje kan finne identiteten til språkbrukaren ved hjelp av ein søkemotor på internett.

For å møte denne utfordringa kan språkforskararen vurdere mengda av sitat i publikasjonen og slik justere risikoen for at informanten blir identifisert. Med å vurdere innhaldet i sitata kan forskaren redusere skaden om informanten likevel blir identifisert. Her er det sjølvsagt ikkje snakk om å endre på sitata, men å velje dei med mindre sårbart innhald og samtidig illustrerer språkfaglege poeng. Uansett kor mykje forskaren reduserer risikoen for og potensialet for skade av reidentifisering, bør ho eller han vere open for at aidentifiseringa sjeldan er vasstett. Ei slik open haldning viser den eine av forskarane (Bröske 2017) i denne studien, blant anna i samtykkeskjemaet.

Forskaren må òg vurdere om og korleis informantane skal få informasjon og gi samtykke til å vere med i prosjektet. Fem av dei seks forskingsprosjekta i denne studien har brukt hovudnorma i forskingsetikken om at informantane skal gi eit informert og dokumentert samtykke. Alternativet til denne såkalla gullstandarden for etisk forskning kan vere å orientere informantane om studien for å veie opp for eventuelle risikoar for personvern. Eit tredje alternativ er at informantane verken får informasjon eller gi sitt samtykke. Det finst argument for at teknologien ikkje gjorde det andre av dei tre nemnde alternativa tilgjengeleg for det sjette prosjektet i studien (Hersdal 2018). I slike tilfelle bør

forskaren vurdere om verdien av forskinga er større enn ulempa det er for individet å uvitande bli med i prosjektet.

Eit siste spørsmål er om begge eller alle samtalepartane i sosiolingvistisk forskning på sosiale medium er informantar og ikkje tredjepartar. To av arbeida i studien (Lande 2017, Morland 2020) har følgd ein slik tanke tydeleg og har fått samtykke frå både primær- og sekundærinformantar.

Litteratur

- Bakkevoll, Geir. 2019. «*Mye tøv du skriv*». *Språklege og semiotiske ressursar i politisk CMC*. [Masteroppgåve]. UiT Noregs arktiske universitet, Institutt for språk og kultur. <https://hdl.handle.net/10037/17801>
- Bolander, Brook & Miriam A. Locher. 2014. Doing sociolinguistic research on computer-mediated data: A review of four methodological issues. *Discourse, Context & Media* 3, 14–26. <https://doi.org/10.1016/j.dcm.2013.10.004>
- Boyd, Danah M. 2008. *Taken Out of Context American Teen Sociality in Networked Publics*. [Ph.d-avhandling]. University of California, Berkeley. <http://dx.doi.org/10.2139/ssrn.1344756>
- Brøske, Kari H. 2017. *Tekstskapingsprosesser i sosiale medier. En sosiolingvistisk studie av Instagram*. [Masteroppgåve]. NTNU, Det humanistiske fakultet, institutt for språk og litteratur. <http://hdl.handle.net/11250/2483144>
- Bucher, Taina. 2012. Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society* 14-7, 1164–1180. <https://doi.org/10.1177/1461444812440159>
- Bucholtz, Mary & Kira Hall. 2005. Identity and interaction: a sociocultural linguistic approach. *Discourse Studies* 7-4/5, 585–614. <https://doi.org/10.1177/1461445605054407>
- Deumert, Ana. 2014. *Sociolinguistics and Mobile Communication*. Edinburgh: Edinburgh University Press.
- Eckert, Penelope. 2013. Ethics in linguistic research. I Podesva, Robert J. & Devyani Sharma (red.): *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 11–26. <https://doi.org/10.1017/CBO9781139013734.001>
- Elgesem, Dag. 2015. Consent and information – ethical consideration when conducting research on social media. I Fossheim, Hallvard & Helene Ingjerd (red.): *Internet Research Ethics*. Oslo: Cappelen Damm Akademisk, 14–34. <https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/23/1056-1>

- Enjolras, Bernard, Rune Karlsen, Kari Steen-Johnsen & Dag Wollebæk. 2013. *Liker, liker ikke. #sosiale medier, samfunnsengasjement og offentlighet*. Oslo: Cappelen Damm Akademisk.
- Evjen, Liv R. 2011. *æ sitt med klump i halsen når æ skriv det hær. Dialekt i skriftspråket i debattforum knytt til tre norske nettaviser*. [Masteroppgåve]. Universitetet i Tromsø, Fakultet for humaniora, samfunnsvitenskap og lærerutdanning. <https://hdl.handle.net/10037/4067>
- FEK De nasjonale forskningsetiske komiteene. 2018. *Hvem er vi og hva gjør vi?* <https://www.forskningsetikk.no/om-oss/hvem-er-vi/>.
- Folkvord, Martin S. 2020. *Normer i et ikke-normert skriftspråk – en studie av talemålsnær skriving på Facebook Messenger blant trondheimere*. [Masteroppgåve]. Universitetet i Bergen, Institutt for lingvistiske, litterære og estetiske studium. <http://hdl.handle.net/1956/22890>
- Forskningsetikkloven. 2017. *Lov om organisering av forskningsetisk arbeid (LOV-2017-04-28-23)*. Lovdata. <https://lovdata.no/dokument/NL/lov/2017-04-28-23>
- Fossheim, Hallvard & Helene I. 2015. *Internet Research Ethics*. Oslo: Cappelen Damm Akademisk.
- Hersdal, Helene F. 2018. *«er dritkul kun med et t? mit liv er en løgn». En sosiolingvistisk studie av metadiskurser om språk i en SKAM-nettgruppe*. [Masteroppgåve]. NTNU, Det humanistiske fakultet, Institutt for språk og litteratur. <http://hdl.handle.net/11250/2563620>
- Holmes, Janet. 2013. *An Introduction to Sociolinguistics* 4. utg. Harlow: Pearson Longman.
- Hougaard, Tina Thode & Lasse Balleby. 2019. Medialektale træk i digitale dialoger. *Språk i Norden 2019. Tema: Språket i sociala medier*, 7–23, <https://tidsskrift.dk/sin/article/view/114941>
- Hårstad, Stian. 2020. Språk på nettet - tilnæringsmåtar til digital språkbruk. I Aa, Leiv Inge & Randi Neteland (red.): *Master i norsk. Metodeboka 1*. Oslo: Universitetsforlaget, 140–157.
- Hårstad, Stian. 2021. Digital skriving under den sosiolingvistiske lupen: Har det skjedd en språklig revolusjon? I Jølle, Lennart, Ann Sylvi Larsen, Hildegunn Otnes & Leiv Inge Aa (red.): *Morsmålsfaget som fag og forskningsfelt i Norden*. Oslo: Universitetsforlaget, 21–45.
- Jacobsen, Dag I. 2005. *Hvordan gjennomføre undersøkelser? Innføring i samfunnsvitenskapelig metode*. Kristiansand: Høyskoleforlaget.
- Lande, Thor M. 2017. *My kjedelige grenå i ol ass. Stilistisk variasjon i en gruppechat på Facebook*. [Masteroppgåve]. Universitetet i Tromsø, Fakultet

- for humaniora, samfunnsvitenskap og lærerutdanning, Institutt for språk og kultur. <https://hdl.handle.net/10037/11323>
- Lüders, Marika. 2015. Researching social media: Confidentiality, anonymity and reconstructing online practices. I Fosshem, Hallvard & Helene Ingierd (red.): *Internet Research Ethics*. Oslo: Cappelen Damm Akademisk, 77–97. <https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/23/1064-1>
- Morland, Petter S. 2020. «U have to talk together». *En studie av kodeveksling, emoji og emotikon i Facebook Messenger-samtaler*. [Masteroppgåve]. Universitetet i Bergen, Institutt for lingvistiske, litterære og estetiske studium. <https://hdl.handle.net/1956/22898>
- NESH Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora. 2014. *Etiske retningslinjer for forskning på Internett*.
- NESH Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora. 2016. *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi*. <https://www.forskningsetikk.no/globalassets/dokumenter/4-publikasjoner-som-pdf/forskningsetiske-retningslinjer-for-samfunnsvitenskap-humaniora-juss-og-teologi.pdf>
- NESH Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora. 2018. *Forskningsetisk veileder for internettforskning*. <https://www.forskningsetikk.no/globalassets/dokumenter/4-publikasjoner-som-pdf/forskningsetisk-veileder-for-internettforskning-2018-interaktiv.pdf>
- Nesvold, Anette, Ragnhild Sjurseike & Carl Cato Wadel. 2014. Å bruke nettdebatter som kilde i forskning. *Tidsskrift for samfunnsforskning* 55-1, 106–120. <https://doi.org/10.18261/ISSN1504-291X-2014-01-07>
- Nilsen, Hilde. 2000. Syntaks og språkkontakt i Nordreisa. *Målbryting* 4. <https://doi.org/10.7557/17.4737>
- Personopplysningsloven. 2000. *Lov om behandling av personopplysninger* (LOV-2000-04-14-31). Lovdata. <https://lovdata.no/dokument/LTI/lov/2000-04-14-31>
- Personopplysningsloven. 2018. *Lov om behandling av personopplysninger* (LOV-2018-06-15-38). Lovdata. <https://lovdata.no/lov/2018-06-15-38>
- Rotevatn, Audhild Gregoriusdotter. 2014. *Språk i spagaten. Facebook-språket. Om normert språk og dialekt blant vestlandselevar*. [Masteroppgåve]. Høgskulen i Volda. <http://hdl.handle.net/11250/2357530>
- Røyneland, Unn. 2017. Hva skal til for å høres ut som du hører til? Forestillinger om dialektale identiteter i det senmoderne Norge. I Östman, Jan-Ola, Caroline Sandström, Pamela Gustavsson & Lisa Södergård (red.): *Ideologi,*

- identitet, intervention: Nordisk dialektologi 10*, Nordica Helsingiensia 48. Helsingfors: Helsingfors universitet, 91–106. <http://urn.nb.no/URN:NBN:no-61863>
- Røynealand, Unn & Øystein Vangsnes. 2020. Joina du kino imårgå? Ungdommars dialektskriving på sosiale medium. I Hagen, Kristin, Arnstein Hjelde, Karine Stjernholm & Øystein A. Vangsnes (red.): *Bauta: Janne Bondi Johannessen in memoriam*, Oslo Studies in Language 11-2. 357–392. <https://doi.org/10.5617/osla.8508>
- Segadal, Katrine U. 2015. Possibilities and limitations of Internet research: A legal framework. I Fossheim, Hallvard & Helene Ingierd (red.): *Internet Research Ethics*. Oslo: Cappelen Damm Akademisk, 35–47. <https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/23/1066-1>
- SERISS Synergies for Europe's Research Infrastructures in the Social Sciences. 2019. *Report on legal and ethical framework and strategies related to access, use, re-use, dissemination and preservation of social media data*. https://seriss.eu/wp-content/uploads/2019/11/D6.3-Report-on-legal-and-ethical-framework-and-strategies..._FINAL.pdf.
- Strand, Solveig E. 2018. «Hvis æ skriv bokmål te når av trondheimsvennan mine så får æ kjeft for at æ har blitt så fin». *En sosiolingvistisk studie av språkbruk på Messenger*. [Masteroppgåve]. NTNU, Det humanistiske fakultet, Institutt for språk og litteratur. <http://hdl.handle.net/11250/2563632>
- Stæhr, Andreas. 2015. Sociale medier og hverdagsprog. I Gregersen, Frans & Tore Kristiansen (red.): *Hvad ved vi nu – om danske talesprog?* København: Sprogforandringscentret, 153–164. https://dgcss.hum.ku.dk/online-resources/hvad-ved-vi-nu-om-danske-thesesprog/2015/hvad-ved-vi-nu-alle-filer/Hvad_ved_vi_nu_samlet.pdf
- Vangsnes, Øystein. 2019. Dialekt i sosiale medium – det norske perspektivet. *Språk i Norden 2019. Tema: Språket i sociala medier*, 94–109, <https://tids-skrift.dk/sin/issue/view/8308/831>
- Vikør, Lars. 2007. *Språkplanlegging: Prinsipp og praksis*. Oslo: Novus forlag.
- Vikør, Lars. 2018. Standardspråk og normering. I Bull, Tove (red.): *Norsk språkhistorie III*. Ideologi. Oslo: Novus, 327–419.

Summary

This article presents a review of six recent Norwegian language studies with data from social media, and the topic is the research ethics in such projects. The article first discusses the distinction between anonymous and de-identified data

and informants. When these concepts are used correctly, the linguist can rarely retrieve data from social media without the assessment from the Norwegian Centre for Research Data (NSD). Furthermore, the article explores how to define personal data and argues that linguistic elements can identify the informants. When processing such personally identifiable information in research, the norm of informed consent emerges as the main rule in the Norwegian context. This is just one of the legal bases for processing personal data. In the discussion of these legal bases, the article shows how technology can make it difficult to inform. A final question is how language research in social media defines and treats third parties, and this article argues that these are in fact informants in sociolinguistic research.

Geir Bakkevoll

Ph.d.-stipendiat

Institutt for språk, litteratur, matematikk og tolking

Høgskulen på Vestlandet

Postboks 133

6851 Sogndal

geir.heggas.bakkevoll@hvl.no

Anne-Mette Somby

Fagansvarleg for forskningsetikk

Høgskulen på Vestlandet

Postboks 7030

5020 Bergen

Anne-Mette.Somby@hvl.no