



UiT Norges arktiske universitet

Fakultet for humaniora, samfunnsvitenskap og lærerutdanning

Nordnorsk dialekt i talegjenkjenningsverktøy

En kvalitativ studie av hvordan Word «Dikter» transkriberer nordnorsk talemål

Maiken Katarina Ågård Salamonsen

Mastergradsoppgave i nordisk språk ved lektorutdanningen trinn 8.-13, NOR-3982, mai 2022

Innholdsfortegnelse

1	Innledning.....	1
1.1	Bakgrunn for valg av tema	3
1.2	Problemstilling og forskningsspørsmål	4
1.3	Disposisjon	5
2	Teoretisk bakgrunn.....	7
2.1	Tidligere forskning	7
2.2	Språkteknologi.....	7
2.2.1	Word «Dikter».....	8
2.3	Nordnorsk talemål	9
2.3.1	Lofotmål	11
2.4	Østnorsk standardmål	15
2.5	Tilpassingsteori.....	16
3	Metode.....	17
3.1	Datagrunnlag	17
3.2	Kvalitativ og kvantitativ metode	18
3.2.1	Elisitering	18
3.3	Informantutvalg og informanter	20
3.3.1	Informanter.....	21
3.4	Forskerens rolle	21
3.5	Etiske problemstillinger.....	22
3.6	Studiens design og gjennomføring	23
3.6.1	Selvbruk «Dikter»	23
3.6.2	Intervju	24
3.6.3	Registrering og kategorisering av data.....	25
4	Empiriske funn	27
4.1	Fonologiske trekk	27

4.2	Substantiv	30
4.3	Verb	32
4.4	Personlig pronomener	36
4.5	Spørreord, form og syntaks	37
4.6	Informantenes tilpasning av eget talemål	38
4.7	Oppsummering funn	39
5	Diskusjon.....	41
5.1	Hvordan transkriberer «Dikter» nordnorske talemålstrekk?	41
5.2	Nordnorsk talemål VS østnorsk standardtalemål	43
5.3	Mulige påvirkningsfaktorer for transkriberingen	45
5.4	Informantenes tilpasning	48
5.5	Forslag til utvikling av talegjenkjenningsverktøy	50
6	Konklusjon	53
6.1	Videre forskning	55
	Referanseliste	57
	Vedlegg 1 NSD	62
	Vedlegg 2 Samtykkeerklæring	64
	Vedlegg 3 Intervjugaid.....	66

Sammendrag

Denne masteroppgaven undersøker hvordan talegjenkjenningsverktøyet Word «Dikter» transkriberer nordnorsk talemål til standardspråklig norsk ortografi. Videre undersøker den hvorvidt informanter tilpasser sitt talemål til dikteringsverktøyet. Studien tar i bruk kvalitativ metode, med utgangspunkt i en hypotese om at talemålsverktøy er for lite tilpasset mangfoldet av norske talemål, med vekt på nordnorsk talemål. Datamaterialet som ligger til grunn, er kvalitative data samlet inn ved hjelp av seks unge informanter med dialekt fra Vestvågøy i Lofoten. Datamaterialet består av transkripsjoner utført av «Dikter», samt intervju mer hver enkelt informant. Empirien er analysert gjennom å sammenligne transkripsjonene fra «Dikter» med lydnære transkripsjoner av det informantene faktisk sa. Resultatet viser at «Dikter» har utfordringer med å transkribere nordnorsk talemål til standardspråklig norsk ortografi. Det er særlig de fonologiske trekkene ved nordnorsk talemål som blir transkribert feilaktig av talegjenkjenningsverktøyet, for eksempel palatalitet, senking og apokope. En hovedårsak til dette er at det østnorske standardtalemålet ligger til grunn for databasene til talegjenkjenningsverktøyet, og at nordnorsk talemål vurderes ut ifra dette. Når det gjelder informantenes tilpasning av eget talemål, viser funn fra denne studien at informantene i liten grad tilpasser talemålet til dikteringssituasjonen. Dikteringssituasjonene hvor informantene snakker til en datamaskin påvirker altså ikke informantenes autentiske talemål.

Forord

Først og fremst, tusen takk til Øystein Vangsnes og Trond Trosterud for god veiledning.

Videre ønsker jeg å takke informantene som har tatt seg tid til å bidra i denne studien.

En tredje takk rettes til mine medstudenter for faglige samtaler på lesehuset og sosiale forstyrrelser utenom. Tusen takk for fem fantastiske år på lektorutdanningen.

Sist, men aller mest, tusen takk til min kjæreste Lone. Uten din støtte og tålmodighet hadde det ikke blitt noen masteroppgave.

Maiken Katarina Ågård Salamonsen

Tromsø, mai 2022

Til bestefar.

1 Innledning

I en praksisperiode som norsklærer på en skole i Nord-Norge gjorde jeg og elevene noen interessante observasjoner. Elevene skulle lage skriftlige manus til presentasjonene sine ved hjelp av talegjenkjenningsverktøyet «Dikter» i Word. Dette er et verktøy som gjør muntlig tale om til skrift, og på den måten forenkler skriveprosessen. I arbeidet snakket elevene inn det de hadde tenkt å skrive, mens «Dikter» transkriberte det de sa. Elevene snakket sakte og tydelig med sitt nordnorske talemål slik at verktøyet enkelt kunne oppfatte det som ble sagt. Til tross for dette inneholdt den skrevne teksten mange feil og mangler. Plutselig begynte en av elevene å snakke østlandsk for å sjekke om verktøyet tok dette lettere. Eleven oppdaget raskt at teksten fikk betraktelig mindre feil på østlandsk enn nordnorsk. Flere av elevene i klassen startet å snakke østlandsk for å sjekke om de fikk samme resultat. Word «Dikter» virket å forstå elevens østnorske dialekt bedre enn deres autentiske nordnorske dialekt. Dette inspirerte meg til å skrive denne oppgaven, hvor målet er *å finne ut hvordan digitale talegjenkjenningsverktøy håndterer nordnorsk talemål.*

Digitalisering blir ofte omtalt som den fjerde industrielle revolusjonen, etter den tredje revolusjonen som ga oss datamaskinen. De siste årene har skjedd en enorm utvikling i den digitale verden, og dette har gitt oss nye oppfinnelser som påvirker måten vi lever våre liv. Vi «lever» på mobiltelefonen, måler joggeturene våre med smartklokken og snakker med venner gjennom trådløse hodetelefoner. Digitaliseringen endrer ikke bare levevanene våre, men fører også med seg nye muligheter. Hjelpemidler som digitale verktøy og ulike applikasjoner gjør hverdagen enklere. Talegjenkjenning er et eksempel på hjelpemidler som har oppstått som et resultat av digitaliseringen. «Taleteknologi i menneske–maskin-kommunikasjon er ein global trend, og den kommande eldrebølga har blant anna sett fart i utviklinga av kunstig intelligens, av mobilteknologi og e-helse» (Eckhoff, 2018:23). Taleteknologi handler altså om kommunikasjon fra menneske til maskin. Gjennom talegjenkjenning kan man be Siri om å «ringe mamma» eller få digitale verktøy til å diktere det man sier muntlig til skrift.

Talegjenkjenning inntar stadig nye domener. Det finnes i biler, i høyttaleren i stua og som applikasjoner på PC og mobil. Ifølge en undersøkelse gjort av amerikanske leger svarte hele 62 prosent at de brukte talegjenkjenning i journalføring på jobb (Healthcare, 2018). Digitale talegjenkjenningsverktøy kan brukes til å diktere tekster i skolesammenheng, transkribere forelesninger i sanntid, til teksting på TV, diktering av pasientjournaler eller for å ringe venner. Talegjenkjenningsverktøy kommer med andre ord til nytte på ulike arenaer slik som skole, jobb og fritid. Folk kan spare mye tid på å bruke talegjenkjenning, og på den måten

frigjøre tid til andre ting. Til tross for at digitale verktøy gjør hverdagen lettere for mange, kan de også føre med seg noen utfordringer. Først og fremst krever bruk av digitale verktøy høy digital kompetanse hos brukeren. Dette kan være en utfordring for de som ikke har digital kompetanse, slik som for eksempel en del eldre mennesker. På en annen side kan teknologien ha feil og mangler som blant annet gjør at verktøyene ikke oppfatter hva folk sier. Hva skjer når teknologien ikke forstår det folk sier?

Det er først og fremst en ulempe å være norsk i denne konteksten. Mange av de teknologiske verktøyene som vi bruker, er utviklet av store, kommersielle aktører som Facebook, Apple, Microsoft, Google og Amazon i Silicon Valley. Dette gjør naturligvis at hovedsatsningen på talegjenkjenningsverktøy foregår på engelsk språk. Vi er avhengig av taleteknologi som forstår norsk talespråk. Norsk språk består av mange forskjellige norske dialekter, og den store variasjonen i dialektene gjør ikke saken noe lettere. Viktigheten av gode verktøy for at de skal være nettopp gode, og at de ikke skaper ekstraarbeid for brukerne av verktøyene. Likevel tyder mye på at språkteknologien ikke er godt nok utviklet for talemålsvariasjonen i Norge. Mange kan oppleve at verktøyene ikke oppfatter det som blir sagt på deres dialekt. Verktøyene som skal fungere som hjelpemidler for folk, blir snarere et hinder. Det kan hevdes at jo mer standardisert talemålet er, desto større sannsynlighet er det for at verktøyet transkriberer ordene riktig.

En grunnleggende innramming for oppgaven handler om norske talemål. Det er høy aksept for bruk av norske dialekter i både private og offentlige sammenhenger. Norge omtales ofte som et dialektparadis på bakgrunn av vår høy aksept for bruk av norske dialekter på alle samfunnsområder (Hårstad et al., 2021:13). Vi snakker den samme dialekten uansett hvem vi snakker med, eller hvilken kontekst vi befinner oss i. Det er utbredt og vanlig å høre et mangfold av norske dialekter fra Stortinget, på TV-sendinger og radio. Dersom folk hadde snakket noe annet enn sin autentiske dialekt, ville dette mest sannsynlig blitt ansett som unaturlig. Men hva skjer når vurderingen skal gjøres av digitale talegjenkjenningsverktøy?

Hele konseptet med talegjenkjenning er at teknologien baserer seg på språket som *ett* felles talemål. En del talegjenkjenningsverktøyene skiller ikke mellom norske dialekter, men behandler norsk som ett språk (for eksempel norsk bokmål). For en rekke andre språksamfunn eksisterer det et standardtalemål, et felles talemål som ofte brukes iblant annet offentlige sammenhenger som på radio og TV. I norsk kontekst opererer vi ikke med noe slikt normert talemål, vi bruker norske dialekter offentlig og privat. I observasjonene jeg gjorde med elevene kan det likevel se ut til at østnorsk har en forhøyet status i det norske dialektlandskapet. At

talegjenkjenningsverktøyene ser ut til å håndtere etterligning av østnorsk dialekt bedre enn autentisk nordnorsk talemål tyder på at østnorsk har en annen stilling.

1.1 Bakgrunn for valg av tema

Språkrådets rapport Språk i Norge – kultur og infrastruktur peker på utfordringer knyttet til språkteknologi og det norske språkmangfoldet. I rapporten kan vi lese følgende:

I Noreg har vi høg toleranse for variasjon i munnleg språkbruk, særleg for skilnader mellom dialekter, og det er eit viktig demokratisk prinsipp at alle har rett til å bruke sitt eige språk og si eiga dialekt. Det eksisterer heller ikkje noko (normert) standardtalemål for norsk, slik det er vanleg for mange andre språk. Ei av dei største språkpolitiske utfordringane i vår tid er derfor å sørge for gode rammevilkår og gode nok data til at det er mogleg å utvikle teknologi som tar høgde både for likestilling av skriftspråka nynorsk og bokmål og har stor talemålstoleranse. (Eckhoff, 2018)

Det er med andre ord klart at gode rammevilkår og data er nødvendig for å utvikle teknologi for likestilt språkmangfold i Norge, både muntlig og skriftlig. Bakgrunnen for valg av tema i denne studien er først og fremst et ønske om å undersøke hvordan digitale talegjenkjenningsverktøy håndterer dialektvariasjon. Ny forskning om hvordan språkteknologien fungerer for talere med ulike norske dialekter, kan derfor gi økt kunnskap om funksjonaliteten til verktøyene. Denne studien undersøker talegjenkjenningsteknologiens brukerfunksjonalitet i et lingvistisk perspektiv.

Målet med prosjektet er å kartlegge mer konkret hvordan et bestemt talegjenkjenningsverktøy fungerer for talere med nordnorsk dialekt, og med dette kunne skissere et bilde av dagens situasjon. Dette kan bidra til økt bevissthet rundt språkteknologiene verktøy og deres fungering. Funnene fra studien kan på den måten bidra til satsninger for å gjøre talegjenkjenningsverktøyene mer tilpasset dialektmangfoldet. En slik bevisstgjøring kan motivere virksomheter til å dele språkdata, slik at utviklingen av språkteknologiske verktøy blir bedret. På den måten kan språkteknologiske verktøy bli mer tilpasset mangfoldet av norske talemål.

Masterprosjektet har også et didaktisk aspekt. Som fremtidig lærer ønsker jeg å benytte språkteknologiske verktøy og talegjenkjenningsteknologi i norskfaget. Derfor er det nyttig å ha kunnskap om hvilke utfordringer som kan oppstå i slik bruk. Jeg ønsker å fremme talegjenkjenning som et verktøy for alle mennesker, uansett dialekt og uansett alder. En utbredt

bruk av talegjenkjenningstøyt i skolesammenheng i dag er å støtte elever med lese- og skrivevansker. Mange av de språkteknologiske hjelpetøytene som Intowords og Lingdys er forbeholdt elever med slike utfordringer. Kraft et al. (2019:104) hevder at dikteringsverktøyt kan bidra til å lette på elevers skriveprosess for elever med lese- og skrivevansker. Med denne oppgaven ønsker jeg å fremme verdien av talegjenkjenningstøyt for alle elever, uavhengig av kognitive evner eller dialekt.

1.2 Problemstilling og forskningsspørsmål

Undersøkelsen tar utgangspunkt i en hypotese om at *digitale talegjenkjenningstøyt ikke er godt nok tilpasset mangfoldet av talemål som finnes i Norge, som for eksempel nordnorsk talemål*. For å besvare problemstilling har jeg formulert følgende forskningsspørsmål:

1. Hvordan transkriberer «Dikter» nordnorske talemålstrekk til standardspråklig norsk ortografi?
2. Hvordan tilpasser nordlendinger talemålet sitt når de bruker talegjenkjenningstøytet?

Som en avgrensing av oppgaven vil denne studien undersøke et bestemt nordnorsk talemål, nemlig lofotmål fra Vestvågøy. Jeg bestemte meg tidlig for å undersøke et dialektområde som sto meg selv nært, nemlig hjemplatsen min Lofoten. Det var et poeng at det valgte nordnorske talemålet skulle tilhøre et ruralt område, ettersom disse dialektområdene tradisjonelt viser mindre grad av dialektutjevning. Med dette ligger det en forventning om å finne flere tradisjonelle talemålstrekk for det aktuelle talemålsområdet. Talemålet på Vestvågøy skiller seg tydelig fra østnorsk, noe som også er et poeng i denne studien. Til tross for at oppgaven undersøker spesifikt lofotmålet, kan studien være relevant for andre nordnorske talemål. Oppgaven vil rette fokus på forholdet mellom det muntlige talemålet og det standardiserte skriftspråket i talegjenkjenningstøytet. Dette betyr at oppgaven ikke vil gå i dybden på hvorvidt informantene snakker autentisk lofotmål eller ikke, utover å belyse talemålstrekkene som samsvarer med nordnorsk talemål. Undersøkelsen vurderer hvordan «Dikter» håndterer og transkriberer de nordnorske talemålstrekkene informantene faktisk har.

I denne studien undersøker jeg et bestemt talegjenkjenningstøyt, «Dikter» i Word. «Dikter» er et verktøyt utviklet av Microsoft, som inngår i flere apper i Office 365 (Microsoft, u.d.). Videre i denne oppgaven vil «Dikter» bare bli omtalt som «Dikter». Dette verktøyet er valgt fordi Word er en programvare som er veldig utbredt både i skole og arbeidsliv. Studien har et tydelig fokus på mediet, hvor *selve* kanalen er forskningsobjektet (Hårstad, 2020:148).

Jeg har valgt dette mediet ettersom mange benytter dette programmet daglig. I tillegg undersøker denne studien hvordan informantene eventuelt tilpasser talemålet sitt når de bruker talegjenkjenningsverktøyet «Dikter». På den måten inkluderes brukerfokuset, som ifølge Hårstad (2020) *handler om å ta utgangspunkt i aktøren*. Denne oppgaven plasserer seg i krysningsfeltet mellom talemålsforskning og språkteknologi. Problemstillingen aktualiserer blant annet spørsmål knyttet til utviklingen av språkteknologiske verktøy.

For å besvare problemstillingen har jeg brukt kvalitativ metode med elisiteringsoppgaver der informantene beskriver to ulike tegneserier for «Dikter». For å sikre at informantene snakker autentisk nordnorsk talemål supplerte jeg med intervju med hver enkelt informant. Grunnlaget for studien min er datamaterialet fra seks informanter fra samme videregående skole i Lofoten. Empirien er analysert gjennom å sammenligne transkripsjonene fra «Dikter» med lydne transkripsjoner av det informantene faktisk sa. Gjennom den analysen har jeg kommet frem til resultater for hvordan talegjenkjenningsverktøyet håndterer nordnorsk talemål.

1.3 Disposisjon

Masteroppgaven består av totalt 6 kapitler. Kapittel 2 gir det teoretiske rammeverket for studien. Delkapittel 2.2 presenterer sentral teori fra språkteknologien. Deretter presenterer jeg trekk ved nordnorsk talemål i 2.3, og spesifikke talemålstrekk for lofotmålet. Teori for østnorsk standardtalemål blir tatt opp i 2.4, før jeg til sist tar for med tilpasningsteori. I kapittel 3 presenterer jeg prosjektets metodiske valg. Her blir oppgavens datagrunnlag presentert og teori om kvalitativ metode, som er sentralt for denne undersøkelsen. Delkapittel 3.3 presenterer informantene og de påfølgende informantkravene. Forskerens rolle og studiens forskningsetiske hensyn blir deretter presentert. I kapittel 4 legger jeg frem studiens funn etter talemålstrekk og grammatiske kategorier, og disse blir diskutert i kapittel 5. I det siste kapitlet oppsummerer jeg og konkluderer funnene og peker på videre forskning.

2 Teoretisk bakgrunn

Det teoretiske rammeverket for denne oppgaven er språkbruksteori, basert på innsikter fra talemålsforskningen. Teori fra språkteknologi vil være et grunnleggende fundament for oppgaven. Oppgaven vil gi en beskrivelse av funksjonen til verktøyet med utgangspunkt i det språkteknologiske aspektet ved talegjenkjenningsverktøy.

2.1 Tidligere forskning

Det finnes svært ingen eller svært lite tidligere forskning knyttet til norske talemål og talegjenkjenningsverktøy. Dette skyldes blant annet at talegjenkjenningsverktøy er en relativt ny oppfinnelse, noe som gjør dette masterprosjektet både motiverende og krevende å arbeide med på samme tid. Et langt mer utforsket fenomen innenfor teknologi og språk, er hvordan språk på internett har endret måten vi samhandler i skrift. Flere undersøkelser viser blant annet en større utbredelse av dialektale trekk på sosiale medier (jf. Evjen 2011, Folkvord 2020). Utviklingen av digitale plattformer har påvirket måten vi uttrykker oss på skriftlig. Forskningen har i større grad undersøkt hvordan personer skriver dialekt i ulike digitale forum. Med dette kan det se ut til at andre områder har fått større plass i forskningen enn spesifikke talemål og språkteknologiens vurdering av disse. Denne undersøkelse ser nærmere på hvordan et språkteknologisk verktøy håndterer nordnorske talemål.

2.2 Språkteknologi

Språkteknologi er all teknologi som rettar seg mot eller tar i bruk menneskeleg språk. Teknologien blir brukt på så ulike område som søk på internett, automatisk omsetting (maskinomsetting), dikteringsverktøy, talesyntese og rettskrivingsprogram, for å nemne nokre få døme (Røyneland, 2018:25). Språkteknologien er med andre ord teknologi som tar utgangspunkt i naturlig, menneskelig språk, og kan brukes på en rekke områder i samfunnet. Formålet med språkteknologi er å gjøre hverdagen enklere. Vi kan blant annet snakke med samtaleroboter på hjemmesidene til bankene, oversette facebook-innlegg fra et språk til et annet, diktere muntlig tale til tekst eller få Google Home til å spille av musikk etter muntlig kommando fra et menneske. I tillegg finnes det verktøy som er utviklet for mennesker med spesifikke utfordringer. Lingdys og Intowords har for eksempel talegjenkjenningsfunksjoner som skal fungere som hjelpemiddel for mennesker med dysleksi. Mennesker med fysiske funksjonsnedsettelse, eldre eller andre særskilte behov kan også ha godt utbytte av slike verktøy.

Til grunn for språkteknologi ligger språkdata, som for eksempel ordbøker, rapportert, korpus, talespråk eller andre store tekstmengder. Det er med andre ord et leksikon bak de digitale talegjenkjenningssystemene. Språkdata inkluderer både skriftspråk og talespråk (Digitaliseringsdirektoratet, u.d). Teknologien er altså matet med språkdata i enten muntlig eller skriftlig form. I norsk sammenheng har Språkbanken vært en sentral aktør for å samle inn og tilgjengeliggjøre språkdata fra ulike fagområder. Språkbanken ble opprettet som det største enkeltstående tiltaket av stortingsmeldningen *Mål og mening. Ein heilskapeleg norsk språkpolitikk* fra 2008. Språkbanken ved Nasjonalbiblioteket tilbyr store datasett med språk, og er viktig for utviklingen av språkteknologi (Nasjonalbiblioteket, u.d). De kommersielle og internasjonale selskapene satser på større språk, slik som engelsk, i utviklingen av språkteknologiske verktøy. Dette fører til at de norskspråklige versjonene ofte blir lansert senere, og er av dårligere kvalitet. Tilgang til et mangfold av språkdata er avgjørende for utviklingen av gode språkteknologiske verktøy blant private og offentlige aktører.

Dikteringsverktøy er en type talegjenkjenning og en underkategori innenfor språkteknologi. Talegjenkjenningsverktøy fungerer gjennom at en språkbruker sier noe til verktøyet (input) og verktøyet velger ord og fraser med utgangspunkt i eksisterende databasene (output). Enkelte dikteringssystemer lærer seg å gjenkjenne stemmen til den enkelte språkbrukeren og kan dermed tilpasse seg brukerens uttale og dialekt i noen grad. Dette gjelder blant annet Apples talegjenkjenningsverktøy, *Siri*. Her kan brukeren sette innstillingene slik at verktøyene gjenkjenner brukerens stemme. Brukeren må svare på en rekke spørsmål, slik at verktøyet får de språkdataene som trengs for å utføre gitte kommandoer. «Dikter» har i skrivende stund ikke stemmegjenkjenning i sine systemer.

For talegjenkjenningverktøy vil naturligvis faktorer som taletempo og diksjon påvirke verktøyenes funksjonalitet. Dette innebærer at verktøyene må få tilstrekkelig med tid til å oppfatte det som blir sagt, for å kunne utføre gitte kommandoer eller diktere. Det er særlig avgjørende for dikteringsverktøy som «Dikter». Microsoft anbefaler brukere å snakke sakte og tydelig når de skal benytte dikteringsverktøyet (Microsoft, u.d.). I dette ligger det en forventning om at taletempo og diksjon kan avgjøre kvaliteten på dikteringen.

2.2.1 Word «Dikter»

Word «Dikter» er et talegjenkjenningsverktøy utviklet av Office 365, først lansert i 2019. Microsoft skriver følgende om verktøyet på sin nettside; «Med Diktering kan du bruke tale til tekst for å redigere innhold i Office med en mikrofon og pålitelig Internett-tilkobling» (Microsoft, u.d.). «Dikter» er et verktøy som dikterer tale til tekst i sanntid. Det fungerer ved at

mennesker snakker til talegjenkjenningverktøyet gjennom PC eller mobiltelefon. «Talegjenkjenning er teknologi som kan oppfatte og tolke menneskelig tale» (Eide & Våge, 2018:11). Teknologien gjør det mulig å styre maskiner, utføre kommandoer eller diktere tekst gjennom menneskelig tale. Dikteringsfunksjonen til Microsoft er gratis og innlemmet i mange av Office programvarer som for eksempel Word, Powerpoint, Onenote og Outlook.

Microsoft er et kommersielt selskap, og de er derfor ikke åpne om hvilken teknologi som ligger bak «Dikter». Vi vet at en del språkteknologi er basert på maskinlæring i nevralt nett. Nevrale nettverk er «datasystem brukt i maskinlæring som skal etterligne menneskers måte å lære noe på» (Eide & Våge, 2018:11). Dette betyr at språkteknologiske verktøy langt på vei kan lære på lignende måte som mennesker gjør. Enkelte dikteringssystemer lærer seg blant annet å kjenne igjen stemmen til den enkelte språkbrukeren, og kan dermed tilpasse seg brukerens uttale og dialekt i noen grad. De systemene Språkrådet arbeider med, er systemer som ikke kan tilpasses enkeltpersoner på denne måten. Det ser ut til at «Dikter» heller ikke kan tilpasses enkeltpersoner i særlig grad. Når dataprogrammer avgjør hvilket ord som skal produseres, gjøres dette basert på en ordliste som interagerer med statistikk over hvor hyppig ordet blir brukt, og hvor sannsynlig det er at det blir brukt i en bestemt kontekst (Davidsen, 2013:766).

For å bruke talegjenkjenningverktøyet i Word, åpner man «Dikter» ved å trykke på Dikter-ikonet. Videre kan man velge hvilket talespråk man snakker, samt aktivere automatisk tegnsetting og filtrering for upassende språk. Deretter starter man diktering ved å klikke på mikrofon-ikonet, og snakker inn det man ønsker at Word skal dikte. «Dikter» skiller ikke mellom de norske målformene, bokmål og nynorsk. «Dikter» kan kun produsere norsk bokmål. Taleren skal i utgangspunktet kunne snakke hvilken som helst dialekt i bruk med verktøyet, og få bokmål som output.

2.3 Nordnorsk talemål

Denne studien undersøker talemål, og forholdet mellom nordnorsk talemål og standardskriftspråket bokmål. For å kunne si noe om hvordan talegjenkjenningverktøyet «Dikter» håndterer nordnorsk dialekt, er det nødvendig å ha tydelig for seg hva som kjennetegner den nordnorske dialekten som undersøkes, nemlig lofotmål. I denne delen vil jeg trekke frem de mest sentrale dialekttrekkene for lofotmålet, med utgangspunkt i beskrivelser hos blant annet Brit Mæhlum & Unn Røyneland (2012), Ernst Håkon Jahr & Olav Skare (1996) og Hallfrid Christiansen (1948). Teoridelen vil starte med å gi et kort oversyn over trekk ved

det nordnorske talemålet, deretter tar oppgaven for seg de trekkene som er særlig gjenkjennelige for lofotmålet.

Det nordnorske dialektområdet utgjør en av de fire hovedområdene innenfor det norske dialektlandskapet. De to nordligste fylkene, Nordland og Troms og Finnmark, utgjør det nordnorske dialektområdet. På bakgrunn av områdets størrelse er det viktig å poengtere at talemålet innad i Nord-Norge er svært varierte. Med andre ord er det noe problematisk å omtale hva som kjennetegner de nordnorske dialektene. Oppgaven inkluderer derfor de trekkene som regnes som mest generelle for det nordnorske dialektområdet. Mæhlum & Røyneland (2012) trekker frem fire trekk som er generelle for dialektene i det nordnorske talemålsområde:

- 1) Fravær av jamvekt
- 2) Apokope
- 3) Med/uten tjukk l
- 4) Høytone

Fravær av jamvekt er det første trekke som kjennetegner det nordnorske dialektområdet. Jamvekt er et språkfenomen fra gammelnorsk som betegner tostavede ord som har likt trykk i begge stavingene (Jahr & Skare, 1996:24). Trøndelag og Østlandet utgjør kjerneområde for jamvektsregelen. Som hovedregel har ikke nordnorsk talemål jamvekt. Nordnorsk omtales som et e-mål (Mæhlum & Røyneland, 2012:106). Hele Troms og Finnmark, samt nordlige deler av Nordland områder der alle infinitiver ender på en slapp e-lyd (for eksempel å vise). For sørlige deler av Vesterålen, hele Lofoten, Salten og nedover i Nordland er infinitivene i ulik grad apokoperte (Mæhlum & Røyneland, 2012:106). Mer spesifikt om dette kommer i neste del som omtaler lofotmålet.

Et annet kjennetegn ved det nordnorske dialektområde er apokope. Apokope vil si bortfall av utlydende vokal eller en trykklett stavelse i slutten av et ord (Jahr & Skare, 1996:26). Statisk apokope er mest utbredt i nordnorske og trønderske dialekter. Kjerneområdet for apokopen er Salten, og graden av apokope minker etter avstand til Salten. Nullmål brukes om det talemålsområdet hvor apokopen er sterkes i område rundt Salten (Mæhlum & Røyneland, 2012:108). Mæhlum & Røyneland (2012) skriver at dette omfatter hele Salten, Lofoten og den sørlige delen av Vesterålen. Nullmåls-område har tradisjonelt apokope av endelsene i infinitiv (for eksempel å lek_) og svake hokjønnsord ubestemt form (ei vekk_). Flere grammatiske kategorier som blant annet tallord og presens og preteritum av svake e-verb kan også bli apokoperte i sentralområdet for apokope.

Det går et geografisk skille mellom statisk og fri apokope. I område nord i Nordland, for mesteparten kysten i Troms og for hele kysten i Finnmark har dialektene statisk apokope. Dette er apokope som knyttes til bestemte grammatiske kategorier og som hovedregel gjelder for alle formene i disse kategoriene (Mæhlum & Røyneland, 2012:107). For det resterende område er det fri apokope, der ord som står i en relativt trykksvak posisjon i en setning, blir apokopert, mens endingsvokalen beholdes i trykksterke former. I delen for lofotmål under vil oppgaven gå nærmere inn på hvordan apokopen kommer til uttrykk i lofotmålet.

Dialektene i det nordnorske dialektområde har varierende grad av tjukk l, som utgjør det tredje kjennetrekket ved nordnorsk. Tjukk l brer seg geografisk fra Østlandet og Trøndelag helt opp til Hamarøy i Nordland (Mæhlum & Røyneland, 2012:108). Tjukk l kommer til uttrykk i ord som soL og daL, for gammelnorsk l. I tillegg har noen dialekter tjukk l av gammelnorsk rð, for eksempel i ordet gård. Lofotmålet har ikke tykk l, men vanlig dental l (Christiansen, 1948:142). Ettersom at det er stor geografisk variasjon i bruken av tjukk l, er det litt utfordrende å bruke dette som en markør for nordnorsk talemål. Et fenomen som derimot omfatter hele det nordnorske dialektområde, er retroflektering. Den retroflekse lateralen regnes å være en utvikling av tjukk-l gjennom en assimilasjonsprosess i en del nordnorske talemål (Mæhlum & Røyneland, 2012:112). Retroflektering er en språklyd hvor taleren artikulerer tungespissen oppover og bakover, slik som i ordet pæ:le (perle). Retroflektering kommer til uttrykk ved r etterfulgt av alveolare konsonanter som n, l, t, d og s (Mæhlum & Røyneland, 2012:48). Fenomenet rammer dialekter i det østnorske, trønderske og nordnorske talemålsområde.

Det fjerde trekket som trekkes frem for det nordnorske dialektområdet er høytone. Høytone handler om at trykket legges i starten av ordet, og deretter faller den mot slutten (Mæhlum & Røyneland, 2012:37). Høytone både i det nordnorske og vestnorske talemålsområdet, mens lavtone er utbredt i det østnorske og trønderske dialektområdet. Lavtone er det motsatte av høytone, og er talemål hvor starten av ordet har en lav tone som stiger mot slutten av ordet. Christiansen trekker også frem høytone som fellestrekk for nordnorske talemål, herunder lofotmålet (Christiansen, 1973[1954]:48). Høytone og annen prosodisk variasjon blir ikke nærmere undersøkt i denne studien, men er verdt å nevne som et trekk ved det nordnorske dialektområde.

2.3.1 Lofotmål

Lofotmålet hører som nevnt til under det nordnorske dialektområde. I denne studien omfatter lofotmålet det talemålet som finnes i Lofoten og den vestlige delen av Vesterålen (Jahr & Skare, 1996:131). Lofoten består av kommunene Værøy, Røst, Moskenes, Flakstad, Vestvågøy,

Vågan og Hadsel. Med andre ord strekker lofotmålet seg over et større geografisk område. Det verdt å merke seg at talemålet i Lofoten varierer en god del etter geografisk plassering. Videre i denne delen vil jeg ta for meg de trekkene som regnes til lofotmålet. Flere av elementene som blir dratt frem er felles for flere nordnorske talemål, men er også sentrale i denne undersøkelsen. Først kommer en del om de fonologiske og morfologiske trekkene ved talemålet. Deretter kommer en del om leksikalske trekk, før syntaks nevnes til sist.

2.3.1.1 Fonologiske og morfologiske trekk ved lofotmålet

Et av de mest sentrale morfologiske trekkene for lofotmålet er mer eller mindre gjennomført apokope (jf. Christiansen 1948, Jahr & Skare 1996). Lofotmålet har statisk apokope, men færre grammatiske kategorier blir rammet av fenomenet sammenlignet med saltenmålet. Apokopen henger sammen med at lofotmålet er et nullmål, og det innebærer apokopen ofte er bundet til hele ordklasser (Jahr & Skare, 1996:14). Dersom ord i en ordklasse blir apokopert, er det stor sannsynlighet for at alle ordene blir det. Christiansen (1948) kaller apokopen knyttet til grammatiske kategorier for en statistisk apokope, hvor alle infinitiver og andre grammatiske kategorier i noen grad blir apokoperte. I lofotmålet er apokopen mest fremtredende i infinitiver av verb og i svake hunkjønnsord i ubestemt form. På lofotmål heter det for eksempel å snakke og å leit i infinitiv, mens det i svake hunkjønnsord ubestemt form tradisjonelt heter ei vekk og ei vis. Det kan finnes lokale forskjeller for mengden apokope innad i Lofoten, men at disse grensene er noe uklare. Dette betyr at synligheten av apokope vil kunne variere etter hvor i Lofoten taleren kommer fra. I tillegg er apokope et trekk som er under press mange steder i Nord-Norge (Jahr & Skare, 1996:108). Derfor er det kan graden av apokope varierer også i lofotmålet.

Et talemålstrekk som henger tett sammen med apokopen, er cirkumfleks. Cirkumflekstonelag «er eit direkte resultat av at tostavingsord blir gjort einstava» (Jahr & Skare, 1996:28). Apokoperte ord som opprinnelig hadde to stavelser, kan få cirkumflekstonelag. For eksempel kan verbet vise bli til vîs i lofotmålet, fordi det blir apokopert og får cirkumflekstonelag. Et typisk bøyingsmønster for verb med apokope er følgende:

infinitiv	presens	preteritum	presens perfektum	imperativ
spis	spis	spist	spist	spis!

Tabell 1: Bøyingsmønster for verb med nordnorsk apokope

Et annet fonologisk fenomen for lofotmål er senking. Senking eller lågning «er når en gammel stutt i, e og y som har «falle ned» eit hakk, skil at ein kan få uttalevariantar som fesk, fæst og løst (subst.)» (Mæhlum & Røynealand, 2012:111). Trange fremre vokaler blir med andre ord uttalt med senket tunge. Ord som *rykk* og *hest* blir til *røkk* og *hæst*. Sekning av vokaler finner vi også i flere nordnorske dialekter, samt i trøndersk og nordøstnorske talemål.

Et tredje fonologisk trekk ved nordnorsk og lofotmålet, er palatalisering av opprinnelige lange alveolarer. Palatalisering er et fonologisk fenomen hvor konsonantsamband blir uttalt med en j-lyd (Mæhlum & Røynealand, 2012:110). Dette gjelder først og fremst for konsonantsambandene nn og ll (Christiansen, 1948:154). For eksempel viser dette seg i ord som 'hund' og 'alle', som blir til /hunn/ (palatal nasal) og /aʎʎe/ (palatal lateral). Den største intensiteten av palatalitet finner vi i Nord-Norge og Trøndelag. Denne oppgaven skiller mellom den historiske utviklingsprosessen palatalisering av alveolarer og selve talemålstrekket palatalitet (jf. Mæhlum & Røynealand 2012:80). Videre i oppgave er det begrepet palatalitet som vil bli brukt om talemålstrekket.

Når det gjelder det morfologiske systemet i lofotmål er det først og fremst relevant å se nærmere på den grammatiske kategorien substantiv. Bøyingsendingsmønsteret for substantiv med endelsen -a det vanligst i nordnorsk bestemt form entall av svake hunkjønnsord, for eksempel *ei jenta* og *ei visa* (Jahr & Skare, 1996:30). Denne endelsen er hovedregelen i bøyingsmønsteret for Troms, Finnmark og Nordland ned til Hamarøy. Det samme bøyingsmønsteret gjelder for sterke hunkjønnsord i bestemt for entall. I områder nord for Helgeland har alltid denne formen endelsen -a, for eksempel *sola* (Jahr & Skare, 1996:32).

Endelsen i flertall av svake hunkjønnsord er et trekk som skiller de nordnorske dialektene fra hverandre, også innad i Lofoten. I mesteparten av Vest- og Aust-Lofoten er det endelsene -e og -en som er vanligst (Jahr & Skare, 1996:34). Bøyingsmønsteret for ordet 'jenta' ville da blitt flere *jente* og alle *jentan* (se tabell 2 under). Dette henger sammen med at lofotmål har bortfall av r i ubestemt flertall i substantiver (Christiansen, 1948:209). Enkelte deler av Vestvågøy og Flakstad har tradisjonelt hatt endelsene -a og -an (for eksempel flere *jenta* og alle *jentan*). Gimsøya skiller seg fra resten av Lofoten fordi området har samme bøyingsmønster som i saltenmålet, med bortfall av endelse i bestemt form flertall og -an i bestemt for flertall (flere *jent_* og alle *jentan*). Jahr & Skare (1996) hevder at bøyingsmønsteret -e og -en som ellers dominerer i Lofoten, vil kunne gjelde for Midt-Lofoten også. Ut ifra denne beskrivelsen kan man si at det tradisjonelle bøyingsmønsteret for hunkjønnssubstantiv på Vestvågøy ser slik ut:

entall		flertall	
ubestemt form	bestemt form	ubestemt form	bestemt form
jenta	jenta	jente	jenten

Tabell 2: Bøyingsmønster for svake hunkjønnssubstantiv i lofotmålet

2.3.1.2 Leksikalske trekk ved lofotmålet

Av sentrale leksikalske enheter i lofotmålet er det verdt å trekke frem pronomen 1. person entall, fordi disse skiller seg fra andre norske talemål. I lofotmålet dominerer e(g) som personlig pronomen i første person entall (Jahr & Skare, 1996:36). Videre står det at æ er mest utbredt for tettstedene Kabelvåg og Svolvær på Austvågøya. Værøy og Røst har e som personlig pronomen i 1. person entall. Personlig pronomen 1. person entall i lofotmålet, skiller seg fra tradisjonelle østnorske dialekter der pronomenet starter på j-. «Dåkk» er den vanligste formen av personlig pronomenen i 2. person flertall i lofotmålet. Dette henger sammen med apokopen, hvor man ser overlappende tendenser mellom området for «dåkk» og apokope. Likevel dominerer formen «dåkker» i det meste av Nord-Norge. Når det gjelder nektingsadverb, er det formen *iççe* den tradisjonelle formen i det nordnorske dialektområdet (Mæhlum & Røyneland, 2012:109). Nord-Troms og Finnmark skiller seg fra resten ved at de bruker formen *ikke*, i tillegg til at denne varianten blir mer vanlig i flere av bymålene i nord.

Et annet leksikalsk trekk som er karakteristisk for nordnorsk er spørreordene. Spørreordene i nordnorsk skiller seg fra andre talemål, som blant annet østnorsk. (Christiansen, 1948:181). De ordene som på norrønt hadde konsonantgruppen hv- har fått ulik fremlyd i ulike norske talemål. I grammatiske ord som spørreordene (hva, hvem, hvor ect.), og i adverb (hvordan), har konsonantgruppen blitt til k- (jf. Vangsnes & Westergaard, 2019). Ordene realiseres typisk som /ka/ (hva, pron.) og /kør/ (hvor, adv.), og /ka/ (hva, pron.) og /korsn/ (hvordan, adv.) i lofotmålet. For nordnorske dialekter er det med andre ord k- i fremlyd som er mest vanlig.

2.3.1.3 Syntaks

Et sentralt syntaktisk trekk ved nordnorsk talemål er at det er vanlig med spørresetninger uten subjekt/verb-inversjon, altså uten V2-regelen. Som hovedregel følger syntaksen i norsk grammatikk, V2-ordstilling. V2-regelen handler om at finite verbalet alltid står på andreplassen i fortellende hovedsetninger (Mæhlum & Røyneland, 2012:114). I nordnorske dialekter er ikke-V2 vanligere enn strukturer med V2 i korte ved korte/enstavede *kv*-ledd (*ka*, *kem*, *kor*, *kordan* og *koffer*) (Vangsnes & Westergaard, 2019:169). Et eksempel på dette er

ordstillingen; *ka du sa*, i stedet for *hva sa du*. Dette trekket er vanlig i norske dialekter med unntak av dialekter på Østlandet og Sørlandet. Syntaksen i strukturer med kv-ledd i nordnorske dialekter skiller seg med andre ord fra østnorsk og det skriftlige standardspråket bokmål.

2.4 Østnorsk standardmål

Noe som er spesielt med Norge i lingvistisk sammenheng, er at vi ikke har et offisielt standardtalemål. Standardtalemål er «en varietet som fungerer som norm eller normideal for et større språksamfunn, som regel en nasjonalstat, og som i en del tilfeller er kodifisert.» (Mæhlum, 2009:8). Ut fra denne forståelsen kan man si at standardtalemål kan være en retningsgivende norm for språksamfunnet. Det finnes mange nasjonalstater som opererer med et offisielt standardtalemål. Tyskland har høytysk og britene har Received Pronunciation, for å nevne noen. I stater med standardtalemål er det denne varietetten som brukes i offentlige sammenhenger blant annet på TV, radio eller universitetsforelesninger.

Standardtalemål skiller seg fra standarder for skriftspråk. En skriftspråkstandard er «vedteken og konstruert med eit gitt formål, nemleg å tene som eit fellesspråk. Skriftspråkstanarden har oppnådd autorisasjon av styresmaktene, og det kan ligge ulike politiske eller ideologiske motiv bak utforminga av skriftspråknorma (Akselberg, Mæhlum, Røynealand & Sandøy, 2008:25 – heretter Akselberg et al.). En skriftsspråklig standard er med andre ord en vedtatt standard for skriftspråket i et land, og denne er førende for hvordan det skrives i landet. Norge har hatt bokmål og nynorsk som to likestilte skriftsspråk siden 1885. Mæhlum & Røynealand (2012) skriver at skriftspråkene ofte vokser frem med utgangspunkt i et eller flere talemål, og i et politisk og kulturelt sentrum. Hensyn til talemål har fungert som et av flere normeringsprinsipp for skriftspråk i Norge. Det var viktig å gjøre bokmål forskjellig fra dansk språk, og nynorsk fremmet mangfoldet av norske talemål i større grad. Både bokmål og nynorsk fungerer som rettesnorer for hvordan folk skriver her i landet. Målformene har for eksempel bestemte regler for hvilke ord som er tillatt, hvordan syntaksen skal være og det morfologiske systemet. Vedtatte og standardiserte skriftspråk er et fellesspråk for alle uavhengig av hvilken dialekt man snakker.

Selv om vi ikke har et offisielt standardtalemål i Norge, finnes det likevel grunnlag for å snakke om en uoffisiell standard som påvirker oss og måten vi prater. Mæhlum (2009) argumenterer for at det eksisterer et overregionalt standardtalemål i det norske språksamfunnet (Mæhlum, 2009:13). Standardtalemål kan tillegges en høyere status enn de andre talemålene i et land. Høytysk regnes blant annet for å ha høyere status enn de øvrige dialektene i Tyskland (Askedal, 2009:214). Når et bestemt talemål tildeles status av myndighetene i kraft av å være

standardtalemål, er dette med på å påvirke folk sin dialektbruk. Mæhlum og Røyneland hevder at talemålet i hovedstaden, oslomål, kan fungere som et overregionalt standardtalemål (Mæhlum & Røyneland, 2012:135). Dette henger sammen med både politiske og sosiokulturelle vilkår og blir assosiert med folk som har høy utdanning, status og autoritet. Standardtalemålene får en forhøyet status og makt i språksamfunnene. Man kan snakke om et prestisjehierarki for språkendringer i et land (Mæhlum & Røyneland, 2012:134). Dette henger sammen med Bourdieu (1991) sin påstand om at språkbruk er en del av hverdagslivets økonomisering. Språkbruk utgjør en del av det lingvistiske markedet hvor verdi og kapital råder. Østnorsk talemål med sin status og makt kan altså se ut til å ha en annen verdi enn de øvrige talemålene på det lingvistiske markedet.

Videre hevder Mæhlum & Røyneland (2012) at en slik overregional standard gjerne brukes i geografisk, avgrensede områder. I norsk kontekst representerer hovedstaden sentrumet for denne varieteten. Ettersom vi ikke opererer med en offentlig standard brukes ulike termer blant fagfolk; østnorsk, sørøstnorsk standardtalemål eller talt bokmål. Det er med andre ord et talemål som ligger nært skriftspråket bokmål, og som tillegg inneholder flere sørøstnorske språktrekk. Akkurat hvilke språktrekk som inkluderes er omdiskutert. Videre i denne besvarelsen vil jeg konsekvent bruke termen østnorsk standardtalemål om denne varianten.

2.5 Tilpassingsteori

Norske talemål henger tradisjonelt sammen med geografisk plassering. Det betyr at folk i stor grad snakker det talemålet som er typisk for det området de kommer fra eller bor på. Med økt mobilitet og en mer globalisert verden har det blitt vanligere å endre på talemålet sitt i større eller mindre grad. Noen endrer blant annet når de flytter fra et område til et annet, mens andre justerer enkelte trekk i eget talemål. Innenfor sociolingvistisk talemålsforskning er akkomodasjon et utbredt begrep. Akkomodasjon er talespråklig tilpasning til den man snakker med (Akselberg et al., 2008:160). Akkomodasjon handler med andre ord om å tilpasse sitt eget talemål til personen man snakker med eller en bestemt situasjon. Videre er det vanlig å dele akkomodasjon inn i konvergens og divergens. Konvergens er å tilnærme seg en annen talespråklig varietet, mens divergens markerer avstand til et annet talemål (Akselberg et al., 2008:160). Det er for eksempel vanlig å gjøre større eller mindre justeringer på talespråket sitt dersom man flytter fra en plass til en annen. Bakgrunnen for slike språkvalg kan være mange og sammensatte. Noen tilpasser språket sitt for å gjøre seg forstått, særlig dersom dialekten deres som skiller seg markant fra dialekten til den de snakker til.

3 Metode

I dette kapittelet gjør jeg rede for de metodiske valgene som er gjort i studien. Først kommer en del om datagrunnlaget, hvor jeg begrunner valget om å kombinere talespråk fra intervjuer med informanter og informantenes egen bruk av talegjenkjenningsverktøyet «Dikter». Deretter tar jeg for meg elisitering og skillet mellom kvalitativ og kvantitativ metode. Videre presenterer jeg informantene og forklarer jeg hvordan jeg har gått frem for å velge disse. Deretter trekker jeg frem min egen rolle som forsker og prosjektets forskningsetiske hensyn, før jeg til slutt gjør rede for hvordan datainnsamlingen ble gjennomført og kategorisert i ettertid.

3.1 Datagrunnlag

Datamaterialet i denne undersøkelsen utgjør om lag tolv sider med transkripsjoner fra «Dikter» og Word-transkripsjon av seks intervju, på mellom 5.02 til 10.54 minutter. Både selvbruk av «Dikter» og intervjuene er transkribert manuelt, for å kunne sammenligne det faktiske talespråket med transkripsjonene utført av «Dikter». Datamaterialet ble samlet inn ved hjelp av seks informanter fra Vestvågøy i Lofoten. Informantene er ungdommer mellom 16-18 år, og kjønnsfordelingen er lik.

Hovedvekten av datamaterialet er dikteringer utført av «Dikter», hvor informantene selv har brukt talegjenkjenningsverktøyet og «Dikter» har diktert det de sier. Gjennom informantenes egen bruk av verktøyet kan jeg analysere hvordan verktøyet håndterer nordnorsk talemål. I tillegg til informantenes bruk av verktøyet, har jeg supplert med et kort semistrukturert intervju med hver enkelt informant. Intervjuet brukes med andre ord til å samle inn autentisk språkdata fra informantene. Dette utgjør den andre delen av datamaterialet. Bakgrunnen for at datamaterialet omfatter både bruk av «Dikter» og intervju handler om at dikteringssituasjon kan være med på å påvirke språkbruken til informantene. I intervjusituasjonen kan informantene i større grad få frem sitt autentiske lofotmål.

Et utbredt fenomen innenfor talemålsforskning er at talere endrer talespråket sitt etter hvem de snakker til (jf. Folkevold, 2020). I delen hvor informantene bruker talegjenkjenningsverktøyet, kan det hende at de bevisst eller ubevisst tilpasser eget talemål for at verktøyet lettere skal håndtere talen. Et supplerende intervju vil gi et bredere datagrunnlag for å sikre mer autentisk tale fra informantene. I tillegg gir intervjuet mulighet for å avdekke eventuelle tilpasninger informantene gjør av eget talemål når de bruker «Dikter».

3.2 Kvalitativ og kvantitativ metode

Valg av metode i en studie henger tett sammen med hva man ønsker å finne ut. Innenfor metodelære er kvalitativ og kvantitativ metode sentrale begreper. Ofte er det flytende overganger mellom hva som kan omtales som kvalitativt og kvantitativt. I en del studier vil det være elementer av begge metodeformene, selv om en dominerer. Denne studien hører hovedsakelig inn under kvalitativ metode, fordi den undersøker hvordan «Dikter» håndterer talemålet til kun et fåtall nordnorske språkbrukere fra Lofoten. Studien er interessert i å undersøke hvordan «Dikter» fungerer for disse språkbrukerne. Til tross for at hovedvekten av denne studien sorterer under kvalitativ metode, inneholder den elementer av kvantitativ metode. Dette innebærer blant annet opptellingen av antall nordnorske ord som «Dikter» transkriberte til standardspråklig norsk ortografi.

Kvalitative metoder har som mål å gi «en dypere innsikt og *forståelse* i sammenhengene mellom språk og de konkrete individene som bruker språket» (Akselberg et al., 2008:77). Her er man først og fremst opptatt av å beskrive et emne eller fenomen, ofte ut fra hvordan enkeltindivider fortolker og forstår verden rundt seg. Akselberg et al. (2008) skriver at kvalitative studier kjennetegnes ved at de opererer med ikke-målbare størrelser, inkluderer få informanter, er individbaserte og at de statistisk sett ikke gir generaliserbare resultater. Fordelen med kvalitativ metode er at den kan gi dybdekunnskap om et emne. I tillegg åpner kvalitative metoder for tilpasninger og økt fleksibilitet underveis i undersøkelser.

Kvantitative metoder handler om mengder, altså tall som kan måles og regnes ut (Akselberg et al., 2008:77). Dette markerer et skille mellom kvantitative og kvalitative metoder, der kvantitative metoder er opptatt av målbare størrelser og oversikt. Fordelen med kvantitative metoder er at tall er enkle, adskilte enheter som ofte er lett håndterbare. På en annen side kan kvantitative metoder gå glipp av viktig informasjon som ikke kan tallfestes. Kvalitative metoder er ofte mindre håndgripelige og komplekse, og dette gir forskeren større tolkningsrom. Med større handlingsrom er det opp til forskeren hvordan studien skal utformes for å kunne besvare forskningsspørsmålet.

3.2.1 Elisitering

I datainnsamlingen i denne undersøkelsen representerer elisiteringsoppgaven i bruk av «Dikter» en eksperimentell metode. Datainnsamling av tale kan gjennomføres naturalistisk eller eksperimentelt. «Naturalistisk innsamla data er data som har kome fram naturleg, t.d. gjennom lydopptak av leik, gjennom foreldrerapportering av den lengste setninga barna serier, orda dei bruker [...]» (Selås, 2020:202). Selås (2020) trekker frem at den eksperimentelle måten går ut

på at forskeren rigger til en situasjon hvor barnet sier de ordene eller setningene forskeren ønsker. En fordel med eksperimentell datainnsamling er å skape en situasjon der informantene sier ønskede ord og fraser styrt av forsker.

Elisitering av språkproduksjon er en vanlig metode innenfor lingvistikken. «Elicited imitation (EI) is widely used as a means of studying subjects' language knowledge through eliciting specific forms of their language production bearing on particular aspects of language competence» (Lust & Blume, 2017b:119). Med andre ord er elisiteringsoppgaver utformet på en måte som fremprovoserer språkproduksjon hos subjekter. På den måten kan elisitering gi forskeren tilgang til språkdata som ellers ikke ville kommet frem i det naturlig produserte språket. Lust & Blume (2017a) hevder at språkdata som er elisitert eksperimentelt under kontrollerte forhold kan være mer relevant for å teste hypoteser. Dette markerer et vesentlig skille mellom elisitering og spontan tale. Oppgaver med elisiteringsdesign gir forskeren tilgang til mer relevante data enn spontan tale. I spontan tale vil forskeren i mindre grad ha kontroll over hva informantene sier. Cornips & Poletto (2005) argumenterer for at elisiteringssituasjonen skiller seg totalt fra en hverdagslig samtale. Elisiteringsoppgaver bidrar til å styre språkproduksjonen i en bestemt retning. Dette gir forskeren mulighet til å undersøke spesifikke områder ved språket som for eksempel grammatisk system og ordforråd. Elisitering av talespråk egner seg derfor godt for å fremme autentisk talemål som inneholder de dataene språkforskeren er ute etter.

En form for elisiteringsoppgave som genererer språkproduserte data, er *the pear story*. *The pear story* var opprinnelig en kortfilm laget av Wallace Chafe i 1975 (Chafe, 1980). Den ble vist for mennesker som snakket ulike språk, hvor de fikk i oppgave å forklare hva de hadde sett rett etter filmen var ferdig. Hensikten med å bruke en film som utgangspunkt for språkforskning, handler om å flytte fokuset fra bevisstheten rundt sitt eget språk, til selve narrativet i filmen (Chafe, 1980:18). På den måten kan forskeren få frem talerens autentiske talemål, uten at deltakerne er for fokusert på dette. Mennesker kan bli bevisst sin egen språkbruk når de vet at denne blir undersøkt. Dette kan føre til at de tilpasser språkbruken sin i studier eller i direkte møte med forskeren. En annen fordel med *pear stories* er at det sikrer likhet i fortellingene til de ulike deltakerne, fordi alle tar utgangspunkt i den samme filmen/oppgaven. På den måten kan mange av de samme elementene bli nevnt av språkbrukere med ulike språk eller dialekter. Dette gjør at datamateriale fra *pear stories* er lettere å sammenligne.

3.3 Informantutvalg og informanter

Utvalg av informanter til et forskningsprosjekt henger tett sammen med problemstillingen og datainnsamlingen. (Selås, 2020:201). I forkant av datainnsamlingen bestemte jeg meg for å holde informantutvalget relativt lite. Dette er først og fremst av hensyn til oppgavens omfang, ettersom at det kan være tidkrevende å inkludere svært mange informanter. Før datainnsamlingen er det nødvendig å foreta noen metodiske valg knyttet til informantene. Bugge poengterer at det er «vanlig å stille noen grunnleggende informantkrav for å definere hvilken befolkning du vil undersøke talemålet til» (Bugge, 2020:219). I det følgende vil jeg derfor trekke frem hvilke informantkrav som ligger til grunn i denne studien.

Det primære fellestrekket for informantene som deltar i prosjektet er at de har en nordnorsk dialekt, lofotmål. Talemålet på Vestvågøy i Lofoten skiller seg tydelig fra østnorsk standardtalemål, noe som også er et poeng i denne studien. For å sikre at informantene hadde lofottalemål fastsatte jeg noen kriterier til deres dialekt. Informantene måtte komme fra Vestvågøy, selv identifisere talemålet sitt som lofottalemål og ha bodd på Vestvågøy sammenhengende siden de begynte på skolen. En svakhet for disse kriteriene er at de ikke nødvendigvis sikrer at ungdommer snakker autentisk lofotdialekt. Studien kunne med fordel ha inkludert et krav om at foreldrene også burde ha lofotdialekt. På den måten ville informantenes talemål i større grad kunne ha representert lofotmålet. Likevel vil jeg hevde at kriteriene og informantutvalget er tilstrekkelig for mine formål i undersøkelsen.

Det andre informantkravet i dette prosjektet er knyttet til informantenes alder. Jeg valgte å rekruttere ungdommer som gikk på videregående skole, i aldersgruppen 16-18 år. Valget av unge informanter handler hovedsakelig om at ungdommer ofte er tidlig ute med å ta i bruk moderne teknologi. Innenfor lingvistikken omtales dessuten ofte ungdom som språkendringsagenter. Ungdom som språkendringsagenter betyr at de er tidlig ute med å ta til seg språkendringer og bruke disse i sitt eget språk (jf. Eckert, 1989, 2000). Indirekte betyr dette at unges språk er i hyppigere endring sammenlignet med eldre generasjoner. I tillegg er informantenes alder i tråd med studiens didaktiske aspekt. Som fremtidig norsklærer har jeg et mål om å benytte digitale talegjenkjenningsverktøy som en del av undervisningen. Ungdom er en gruppe som i stor grad benytter seg av digitale verktøy, både i skolehverdagen og på fritiden. Det kan være nyttig å finne ut mer om hvordan talegjenkjenningsverktøy fungerer for denne aldersgruppen.

Kjønn er en variabel som sikrer variasjon i utvalget til denne undersøkelsen. Halvparten av informantene er kvinner, og den andre halvparten menn. Kjønn er en bakgrunnsvariabel som kan påvirke studiens funn, derfor valgte jeg å ha like mange kvinner og menn representert.

For å rekruttere informanter tok jeg kontakt med en lærer på Vestvågøy videregående skole. Jeg kom i kontakt med læreren gjennom felles bekjente, men kjente henne ikke fra før. Læreren fikk informasjon om masterprosjektet, samt hvilke informanter jeg hadde behov for. Denne utvalgsmetoden kalles for strategisk utvalg, fordi forskeren har en strategi for hvilke informanter man velger til studien (Neteland, 2020:54). Videre tok læreren med dette til sine respektive klasser for å finne ut om noen var interessert i å delta. De elevene som var villig til å delta skrev seg opp på en liste og læreren trakk seks kandidater som oppfylte informantkravene. I tillegg ble det oppført to reservereinformanter, dersom noen av informantene skulle trekke seg.

3.3.1 Informanter

Informantene i denne undersøkelsen er seks ungdommer fra samme videregående skole, i aldersgruppen 16–18 år. Kjønnfordelingen mellom informantene er tre jenter og tre gutter. Felles for alle informantene er at de er fra Vestvågøy i Lofoten, identifiserer talemålet sitt som lofotmål og har bodd på Vestvågøy sammenhengende siden de begynte på skolen. Informantene kommer fra ulike steder på Vestvågøy og har alle sin unike språkbakgrunn. I oppgaven blir informantenes talemål omtalt som lofotmål, til tross for at det finnes mindre interindividuelle variasjoner mellom informantene. Med andre ord vil trekkene som typisk for lofotmålet vil bli vektlagt i undersøkelsen. Ingen av informantene hadde brukt «Dikter» tidligere, og bare en av informantene var kjent med funksjonen fra før av.

Studien inkluderer seks informanter fra Vestvågøy i Lofoten mellom 16–18 år som har bodd i området siden de begynte på skolen. Dette er kriterier som kan gjelde for en mye større del av populasjonen på Vestvågøy. Man kan tenke seg at en del resultatene jeg finner i denne undersøkelsen er relevant for større deler av lofotmål, og for andre nordnorske talemål. Talemålstrekk som blant annet palatalitet, apokope og pronomensformen eg/e finner vi igjen i flere nordnorske talemål. På den måten kan resultatene fra denne undersøkelsen være relevante for andre talemål.

3.4 Forskerens rolle

Det kan være en fordel å forske på en varietet man kjenner godt til selv, både for forskeren selv og for informantene. At samtalen mellom informant og forsker foregår på informantens dialekt kan bidra til å unngå mellomliggende skiller mellom den lokale varietet og standardspråket (Cornips & Poletto, 2005:951). Felles talemål kan på den måten skape en nærhet mellom forsker og informant. På en annen side utgjør forskerens talemål er en ekstra variabel som kan

være med å påvirke resultatet (Neteland, 2020:56). I denne sammenhengen betyr det at talemålet til forskeren kan påvirke hvordan informantene snakker under datainnsamlingen. Dialektene i Lofoten bærer preg av svært lokale variasjoner. Et eksempel på dette er personlig pronomen, som jeg uttaler «æ», mens informantene i denne undersøkelsen brukte «eg» eller «e». Derfor forsøkte jeg å skape en trygg og god setting rundt datainnsamlingen, slik at informantene kunne produsere autentisk språkdata uten å bli forstyrret av andre faktorer.

I tillegg kan jeg som forsker ha forutinntatte meninger om lofotmål og hva som kjennetegner dialekten. Dette kan påvirke funnene i studien og hvilke elementer jeg trekker frem som relevant. Etersom det eksisterer lite tidligere forskning på dette feltet, kan det være utfordrende å korrigere eventuelle ukorrekte forutinntatte oppfatninger om talemålet. Det er imidlertid noe som er vanskelig å komme unna i kvalitative studier. Som forsker må man tolke datamaterialet med sin egen språkbakgrunn. Total nøytralitet er derfor utfordrende å oppnå i slike studier. Jeg har like fullt prøvd å være bevisst dette og etterstreber nøytralitet i funn og drøftingsdelen.

3.5 Etske problemstillinger

I alt forskningsarbeid er det nødvendig å vurdere studiens forskningsetiske profil. Til tross for at denne studien undersøker et medium, inkluderer det informanter. For mitt prosjekt er det særlig viktig å ivareta personvernet til informantene ettersom de er ungdom mellom 16–18 år. Tidlig bestemte jeg meg for å ta lydopptak av informantene ettersom at studien har fokus på det muntlige talemålet. Derfor meldte jeg inn masterprosjektet til Norsk Samfunnsvitenskapelig Datatjeneste (NSD) den 10. oktober 2021, og fikk det godkjent 14. oktober (se Vedlegg 1).

For å sikre informantenes personvern er det vanlig at de å anonymiseres i publiserte oppgaver. I tillegg bør forskeren ekskludere personopplysninger som kan føre tilbake til en bestemt person (Neteland & Aa, 2020:19). I denne studien er informantene omtalt med koder (for eksempel Informant 1, 2 osv), og mengden personopplysninger er begrenset slik at man ikke skal kunne spore tilbake til personen. Spørsmålene i intervjuet var utformet på en måte som ikke ga private samtaleemner, som for eksempel religiøst ståsted og helseopplysninger. Dersom dette forekom i intervjuene, ble det utelatt i transkripsjonen.

Videre er det viktig at informantene er opplyste om at de deltar i et forskningsprosjekt, gjennom et såkalt frivillig informert samtykke (Neteland & Aa, 2020:20). Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH) poengterer at «[d]et forskningsetiske samtykket skal være frivillig, informert og utvetydig, og det bør være dokumenterbart» (NESH, 2021). Alle informantene mottok et informasjonsskriv om prosjektet

i forkant. Her var det spesifisert hva det å delta i studien vil si for dem og hvilke rettigheter de har. Det samme innholdet ble også gjennomgått muntlig for hver enkelt informant før intervjuet. Til sist i informasjonsskrivet lå en samtykkeerklæring hvor informantene måtte signere for å kunne delta i prosjektet (Se vedlegg 2). Både i informasjonsskrivet og muntlig ble det poengtert at informantene når som helst kunne trekke seg fra studien uten begrunnelse. Alle informantene var over 16 år, og kunne derfor gi samtykke selv.

3.6 Studiens design og gjennomføring

I denne delen vil oppgaven presentere gjennomføringen av datainnsamlingen. Intervjuene og informantenes bruk av «Dikter» ble gjennomført etter hver andre, individuelt med hver enkelt informant. Det ble satt av om lag 30 minutter per informant. All datainnsamlingen ble gjennomført over en dag på informantenes videregående skole. På skolen fikk jeg bruke et grupperom som var egnet til samtale. I forskningssammenheng er det et poeng at informanten er avslappet og uformell, derfor kan det være lurt å gjennomføre for eksempel intervju hjemme hos informanten (Neteland, 2020:57). Skolen representerer en kjent og trygg arena for informantene og kan derfor være et godt utgangspunkt for datainnsamlingen. Videre i denne delen vil jeg gå nærmere inn på det jeg har valgt å kalle selvbruk «Dikter» og intervju.

3.6.1 Selvbruk «Dikter»

I den delen hvor informantene selv skulle bruke «Dikter», startet jeg med å forklare hva som skulle skje, samt gi de en rask opplæring i talegjenkjenningsverktøyet. Informantene fikk utdelt to tegneserier på ark laget for barn som stammer fra språkkartleggingsverktøyet MAIN: Multilingual Assessment Instrument for Narratives (Gagarina et al. 2012). Tegneseriene består av seks bilder, uten tekst, nummerert med tall fra 1–6. Den ene tegneserien inkluderer en katt, en gutt og en ball, mens den andre handler om en gutt, en hund og en ballong. Informantene ble bedt om å studere en tegneserie av gangen i noen minutter, før de deretter ble bedt om å fortelle hva som skjer i tegneserien. Informantene fikk deretter en PC med en åpent Word-fil foran seg, samt beskjed om å starte diktering når de var klar. Informantene startet dikteringen selv og snakket inn det de så på tegneseriene. De fleste informantene hadde fokus på tegneseriearkene mens de pratet, men et par så opp på skjermen noen ganger mens de snakket. Informantene var med andre ord lite opptatt av hva som ble diktert mens de snakket. Når informanten var ferdig, stanset hen dikteringen selv. Til slutt ble Word-dokumentet med transkripsjonen lagret.

Bruk av tegneserier som utgangspunkt for språkproduksjon er inspirert av Chafes (1980) *the pear story*. Formålet med å bruke dem er først og fremst om å gi informantene et

utgangspunkt for produksjon av autentisk, nordnorsk talemål. Dette er sentralt ved elisiteringsoppgaver, der forskeren styrer språkproduksjonen i en bestemt retning ut fra hvilke forsknings spørsmål studien har (Lust & Blume, 2017b:119). Et ønsket resultat av å bruke tegneserier er at informantenes bevissthet rundt eget talemål minskes, og at deres autentiske talemål kommer frem. En fordel med barnetegneserier er at de skal være enkle å forstå. Tegneserier ment for voksne kan være mer komplekse og sammensatte, og informantene kan derfor bruke lang tid på å forstå innholdet. For å unngå at informantene ble for opphengt i selve tegneserietolkningen, valgte jeg derfor enkle barnetegneserier.

Det var et poeng at informantene skulle snakke på en kommunikativ måte, med mest mulig fri tale. Dersom informantene hadde blitt bedt om å diktere, altså snakke inn skrevne setninger, er det fare for at informantene hadde blitt påvirket av den skriftspråklige formen. Når man leser opp setninger eller fraser er det større sannsynlighet for å bli påvirket av den skriftspråklige formen. Bilder som del av en tegneserie har ikke samme påvirkning, og åpner opp for informantenes autentiske talemål. I tillegg fikk ikke informantene beskjed om å eksplisitt følge med på skjermen, mens Word transkriberte. Dersom informantene hadde blitt bedt om å diktere ett og ett bilde mens de fulgte med på skjermen kunne det ha gitt andre resultater. At informantene snakker med fri tale, fremfor å diktere inn setning for setning, utgjør derfor en feilkilde i denne studien. «Dikter» er et dikteringsverktøy som er ment for diktering. Dette medfører at talegjenkjenningsverktøyet ikke alltid greier å oppfatte hva informantene sier i fri tale. En svakhet ved oppgaven kan derfor være at «Dikter» utelater de delene verktøyet ikke oppfatter eller transkriberer det teknologien antar at det oppfatter. Informantenes taletempo kan også påvirke hvorvidt verktøyet oppfatter det informantenes sier.

3.6.2 Intervju

Den andre delen av datainnsamlingen er intervjuene. Intervjuene er strukturert etter en semistrukturert intervjugaid med fjorten spørsmål, fordelt på tre små kategorier (se Vedlegg 3 intervjugaid). Jeg valgte å utforme et semistrukturert intervju, for å skape noen rammer for intervjuet. Intervjugaiden er utformet etter inspirasjon av intervjugaiden i ScanDiaSyn-prosjektet som i sin tur bygger på intervjugaidene i talemålsprosjektene TAUS og NoTa (Johannesen et al., 2009a). Kategoriene i gaiden min er «Generelt», «Skole» og «Hjemlass». Spørsmålene i intervjuet er åpne og generelle, slik at de fleste skal kunne svare på disse. Intervjuene ble gjennomført individuelt med hver enkelt informant. Hvert intervju varte imellom 5.02 til 10.54 minutter.

Intervjuet blir hovedsakelig brukt som en effektiv produksjonskilde for autentisk nordnorsk tale. Vanligvis brukes forskningsintervju for å få frem meninger hos informantene. Forskere som benytter intervjuetoden, legger vekt på meningsinnholdet i det intervjupersonen sier (Brinkmann & Kvale, 2015:42). Denne studien er derimot opptatt av selve talemålet til intervjupersonen, og legger derfor ikke vekt på meningsinnholdet i det som blir sagt.

Fordelen med intervjuformen, er at den åpner for en fri samtale mellom forsker og informant. Strukturen på intervjuet gjør at den er godt egnet til å få frem og samle inn autentisk nordnorsk talemål hos informantene. Fokuset til informantene flyttes til intervju spørsmålene, og ikke *hvordan* de sier ting. På den måten kan informantene snakke mer fritt ut fra ulike spørsmål, samtidig som at dette utgjør noe av datamaterialet i forskingsprosjektet. Poenget med å bruke intervju i tillegg til selvbruk «Dikter» handlet om å få informantene til å produsere språkdata på en mer naturlig måte, gjennom en samtale. På den måten kunne jeg i større grad forsikre meg om at informantene brukte sitt autentiske talemål i dikteringsoppgaven, gjennom å sammenligne hvilke talemålstrekk informantene hadde.

Det er ikke uproblematisk å intervju ungdommer individuelt. For barn og unge kan gruppeintervju oppleves tryggere (Neteland, 2020b:56). En utfordring med slike intervju er imidlertid at datamaterialet kan bli ujevnt, ettersom at noen snakker mye, mens andre snakker lite. I denne studien ville det ha vært særlig utfordrende å gjennomføre gruppeintervjuer, ettersom at det er talemålet som undersøkes. I gruppeintervjuer er det vanlig at folk snakker i munnen på hverandre. Informantene kan også komme til å påvirke hverandre i gruppeintervjuer kontra individuelle intervjuer (Neteland, 2020:56). Dette ville gjort det problematisk å jobbe med materialet i etterkant. Derfor var jeg opptatt av å skape trygge rammer rundt intervjuet, blant annet ved å gjennomføre datainnsamlingen på informantenes skole. Alle spørsmålene tok utgangspunkt i informantenes eget liv og deres tanker om ulike temaer. Summen av dette kan ha vært med på å bidra til at fokuset på den formelle intervjusituasjonen ble minimert.

3.6.3 Registrering og kategorisering av data

I forkant av datainnsamlingen foretok jeg en test av studiedesignet på noen medstudenter. Studentene kom med nyttige tilbakemeldinger på hvordan de opplevde innsamlingen. Med bakgrunn i denne informasjonen gjorde jeg noen justeringer på tegneseriene og spørsmålene i intervjuet.

Det ble tatt lydopptak av intervjuene og sekvensene hvor informantene benyttet talegjenkjenningsverktøyet «Dikter» for å samle inn talemålsmaterialet. Opptakene ble gjort med en ordentlig lydopptaker for å sikre god teknisk kvalitet på lydopptakene ettersom analysen

er avhengig av fonetiske detaljer i det informantene sier. I tillegg til lydopptakene dikterte «Dikter» det informantene sa direkte i Word i sanntid. Det ble ført notater manuelt underveis i intervjuet og umiddelbart etter dikteringen. Lydopptakene fra intervjuene ble også indirekte spilt av for talegjenkjenningsverktøyet «Dikter». Bakgrunnen for dette var å se hvordan «Dikter» håndterer kommunikativ tale og for å avdekke eventuelle intraindividuelle variasjon av talemålet. Dette ga empiri fra både direkte og indirekte bruk av «Dikter», selv om hovedvekten ligger på den direkte bruken av «Dikter».

Lydopptakene ble transkribert manuelt i etterkant av datainnsamlingen i transkripsjonsprogrammet ELAN. Dette var nødvendig for å kunne sammenligne hvordan «Dikter» håndterer det informantene sa med transkripsjonene av de faktisk sa. Alle lydopptak ble lyttet til og transkribert i sin helhet. På denne måten fikk jeg større oversikt over datamaterialet og kunne markere talemålstrekk og interessante funn underveis. I denne undersøkelsen brukte jeg transkripsjonsnøkkelen til Nordisk dialektkorpus (Johannesen et al., 2009b). Transkripsjonen er lydnær slik at de viktigste nyansene i talemålet kommer frem. Siden palatalitet ikke er markert i denne transkripsjonsnøkkelen har jeg valgt å legge inn dette med vanlige bokstaver etter IPA-standard (for eksempel *ballj* og *allje*). Det er vanlig å markere palatalitet med «påhengt» *j*, der dette er formålstjenlig (Hårstad & Opsahl, 2013:58). I denne undersøkelsen er det viktig å markere palatalitet fordi dette regnes som et sentralt trekk ved lofotmålet.

For å kunne analysere og sammenligne empirien valgte jeg å føre dikteringen fra «Dikter» og de manuelle transkripsjonene inn i et Excel-dokument. Dette ga en oversikt over hva informantene faktisk sa på den ene siden, og det «Dikter» dikterte at informantene sa på den andre siden. Dikteringen av talespråket til hver informant har blitt kodet etter de grammatiske kategoriene, substantiv, verb og personlig pronomen. Innenfor hver av disse kategoriene ble ord markert i ulike fargekoder (blå for korrekt, rød for feil) og symbol (x, y, z) for om dikteringen var korrekt eller feil etter det informanten faktisk sa. For at et ord skulle bli kategorisert som korrekt, måtte det være diktert riktig fra lofotmål til norskspråklig standardortografi og ha korrekt bøyning. For eksempel er den korrekte dikteringen av verbet *snakk* (pres.) på lofotmål, *snakke* på bokmål. Til sist startet analysearbeidet hvor jeg undersøkte hvilke type feil som oppsto mellom tale og skrift.

4 Empiriske funn

I dette kapittelet skal jeg presentere funnene fra informantenes bruk av talegjenkjenningsverktøyet «Dikter». Funnene for informantene presenteres samlet siden dette er mest interessant for studiens forskningsspørsmål. Det kom frem mye i dikteringssituasjonene, men jeg har valgt å se undersøke realiseringen av talemålstrekkene som er typisk for lofotmålet. Funnene blir presentert med utgangspunkt i ulike talemålstrekk og grammatiske kategorier: fonologiske trekk, substantiv, verb, pronomen og en oppsamlingskategori for spørreord, form og syntaks. Spørreord og deres syntaks er inkludert fordi dette markerer et skille mellom lofotmål og østnorsk standardtalemål. Funnene vil bli tolket i lys av tradisjonell talemålsteori undervegs. Deretter presenteres informantenes sin tilpasning av eget talemål fra intervju til dikteringssituasjonen. Til sist kommer en oppsummering av de empiriske funnene.

4.1 Fonologiske trekk

Et av hovedfunnene i denne studien er at de fonologiske trekkene ved nordnorsk talemål skaper utfordringer for «Dikter». Utfordringer knyttet til fonologi handler om feil i diktering som oppstår som følge av nordnorske fonologiske trekk. Innenfor fonologiske trekk er det først og fremst palatalitet som er verdt å undersøke. Ordet 'ball' blir brukt av alle informantene og samtlige informantene uttaler dette ordet *ballj*, med palatalitet, når de bruker talegjenkjenningsverktøyet. Palataliteten hos informantene av opprinnelige dentaler er i samsvar med trekk for tradisjonelt nordnorsk talemål (jf. Christiansen 1948, Mæhlum & Røynealand 2012). Talemålsprøver fra Nordisk dialektkorpus viser også at yngre talere i Stamsund på Vestvågøy har forekomst av palatalitet av opprinnelige dentaler, for eksempel i substantivet *kvellan* (Johannesen et al., 2009a). Palatalitet er derfor et talemålstrekk som kjennetegner lofotmålet, og som vi finner hos alle informantene. For inputen *ballj* valgte «Dikter» ord som for eksempel *banen*, *bensin*, *base*, *bare*, *bedre*, *barn*, *balansen* og *akebanen* i output. Tabellen under viser noen eksempler på palatalitet i datamaterialet:

transkripsjon	«Dikter»
a) guttn prøva liksåm å få tak i balljn # men ee han sætt dær åg slit litt	ja gutten prøver liksom å få tak i banen men han setter og der sliter litt
b) på bilde nommer sækks så har ee guttn klart å fått tak i balljn å har den i hainna si mens kattn sett å gommla på fesken	begynner med seks så har gutten klart å få tak i bensin og ha den i hende sin mens katten setter det på fersken.

c) åsså på andre så ser vi en gutt så kjæm då mæ litt fesk å feskestang å en rød ballj	altså det andre så ser du en gutt som kommer og med litt fisk og fiskestang og en rød ball
d) så sku han gå å heint balljn sin så ser e liksom problemstillinga me balljn e i vannje åg ee han e førr langt unnja tell å få tak i dennj	så skulle han gå hen barn sin så ser liksom problemstillinger med vann er vannet og ja nei for langt unna til å ta en
e) han fikk tak i balljn	kontakt akebanen
f) mens hunnj har gådd ætter posn med pølse	mens hun har gått etter posen med pølse
g) så ser hunnj kjøttposn # åsså spis han kjøtte mens han får ballången ut av tree	så ser hun kjøtt posen og så spiser du kjøttdeig men sånn for ballongen ut av treet
h) i bilde fæmm klatra guttn opp førr å henjte ballången	I bildet 5 klatret gutten opp for å hente ballongen
i) å han skvætt vannj ifra feskebøtta	og hans kaster vann i fra fiske bøtta
j) han bynnj å jag ætter dennj	å begynne å jage etter enda (fonologisk feil)

Tabell 3: Eksempler på ord med palatalitet

Som vi kan se over velger altså «Dikter» ord som leksikalsk ligner på *ballj* i transkripsjonene. Flertallet av ordene «Dikter» valgte har b i fremlyd, men semantikken av ordene skiller seg fra den faktiske betydningen. Det samme mønsteret finner vi for substantivet *hund* (mask.), som i de fleste tilfeller ble transkribert til *hun*. Flertallet av informantene uttaler dette ordet *hunnj*, det vil si med palatalitet. I likhet med transkriberingen av *ballj* velger «Dikter» output som ligner på inputen *hunnj* fonologisk, men semantikken skiller de. Palatalitet i substantiv som *ballj* og *hunnj* er med andre ord et nordnorsk talemålstrekk «Dikter» ikke håndterer i særlig grad.

Til forskjell fra palatal uttale av substantiv, ser «Dikter» ut til å håndtere palatalitet av verb. I datamaterialet bruker flere av informantene ord som *henjte* og *vannje*, med palatal uttale. «Dikter» transkriberer disse korrekt til *hente* og *vannet* i de fleste tilfeller, og dette klassifiserer som korrekt transkripsjon (se tabell 3 over). Dersom man ser bort fra den palatale uttalen vi finner på nordnorsk, er verbene helt identiske på nordnorsk og østnorsk standardtalemål. Et unntak finner vi for ordet 'begynner' som «Dikter» transkriberer med feil bøyning fra lofotmål til bokmål. Inputen til informantene er *bynnj* mens output blir *begynne*. Det leksikalske ordet er det samme, men bøyingsformen er feil. Det går et skille mellom palatalitet av ulike grammatiske kategorier, der verktøyet i stor grad mestrer palatalitet av verb, men ikke substantiv.

Et annet talemålstrekk som kan knyttes til fonologiske aspekter ved nordnorsk, er retroflektoring. Retroflektoring er et talemålstrekk verktøyet i de fleste tilfeller greier å håndtere

korrekt. Vi finner eksempler fra datamaterialet på retrofleksjer av rl (*dårli*), rt (*jort*) og rn (*hovedkarakteren*):

transkripsjon	«Dikter»
a) e litt <u>dårli jort</u> åsså # men ijænn han e en ganske #	er litt <u>dårlig gjort</u> også men igjen med ganske
b) e kannj vel sei <u>hovedkarraktern</u> sãm de ser ut	jeg kan vel se <u>hovedkarakteren</u> som ser ut

Tabell 4: Retrofleksjer i datamaterialet

transkripsjon	«Dikter»
a) snik han seg runnjt åsså stæl han <u>fesken</u> ijænn	snike seg rundt og så står han <u>fiskeren</u>
b) så får nu han tak i balljn åsså spis katta opp <u>fesken</u> hannes mens han ikkje ser på	spor noen tak i ballen og så spise katta opp <u>fisken</u> da mens mens jeg har ikke sett på ja
c) ja å guttn har fådd tak i balljn å katta kosa seg me <u>fesken</u>	ja og gutten har fått tak i ballen og katta koser seg med <u>fisk</u>
d) me <u>feskestang</u> # rø ballj å litt <u>fesk</u>	med <u>fiskestang</u> i rød ball og litt <u>fisk</u>
e) så lægg han fra se <u>fesken</u> # bøtta me feskan	så langt fra <u>fersken</u> bedt om e posten

Tabell 4: Eksempler på ord som inkluderer 'fesk-'

Et tredje trekk som kan knyttes til det nordnorske talemålet er senking. *Fesk* viser fenomenet senking av ordet 'fisk', og er et kjennetegn for det nordnorske talemålet (Mæhlum & Røyneland, 2012:111). De nordnorske uttaleformene av ordet 'fisk' viser seg i datamaterialet med ord som *fesk*, *fesken* og *feskestanga*. Disse forekommer mange ganger i datamaterialet og blir brukt av samtlige informanter. I datamaterialet er det totalt 46 ord med stammen *fesk-*. Av disse ble 59 prosent diktert korrekt på bokmål, med *fisk-* som stamme. De resterende 41 prosent blir diktert feil. Det går et skille mellom ubestemt og bestemt form entall av *fisk*. Ubestemt form entall, *fesk*, ble som hovedregel transkribert korrekt til *fisk*, i sju av elleve tilfeller (64 prosent korrekt). De resterende 36 prosentene feilord for fisk er *fersk*, *flesk* og *fest*. Bestemt form entall *fesken* blir transkribert til fisken bare ni av tjue ganger (45 prosent korrekt). Feilordene «Dikter» velger for *fesken* er blant annet *fiskeren*, *fiske*, *festen* og *fersken* (55 prosent). Som vi kan se er det altså større korrekthet i transkripsjonene av den ubestemte formen *fisk*, enn for bestemt form *fisken*. Ordene «Dikter» velger som output for *fisk* og *fesken* minner om originalene fonologisk, men transkripsjonene er feil. Feilen kan henge sammen med ordets bestemthet, noe som blir diskutert nærmere i neste kapittel.

4.2 Substantiv

kategori	antall ord totalt	korrekt	feil
Substantiv	697	439 ord	257 ord

Tabell 5: Antall substantiv i datamaterialet

Datamaterialet inneholder totalt 697 substantiver. «Dikter» dikterte 63 prosent av disse korrekt fra nordnorsk talemål til standardspråklig bokmålsortografi, hvilket betyr at 37 prosent av substantivene ikke ble diktert riktig. De substantivene som gjennomgående ble diktert korrekt, var blant annet *bilde*, *ballång* og *såmmerfuggel*. «Dikter» dikterte disse til *bilde*, *ballong* og *sommerfugl* på bokmål. Felles for disse substantivene er at den nordnorske uttalen er fonologisk lik det østnorske standardtalemålet. Det var forventet at de ordene som fonetisk ligner på østnorsk ville bli diktert korrekt i denne studien. Videre er det interessant å se nærmere på hvilke typer feil som oppstår i diksjonen fra lofotmål til bokmål, slik tabellen under viser:

antall dikteringsfeil substantiv	leksikalske	morfologiske
257 ord	184 ord	73 ord

Tabell 6: Oversikt over leksikalske og morfologiske dikteringsfeil av substantiv

Datamaterialet har 257 dikteringsfeil, og disse kan videre kategoriseres som enten leksikalske eller morfologiske som tabellen over viser. 72 prosent av ordene tilhører kategorien for de leksikalske feilene, mens 28 prosent av feilene er knyttet til morfologi. Morfologiske feil i denne studien er alle feil som kan knyttes til bøyingssystemet. I materialet finner vi noen morfologiske feil i dikteringen fra nordnorsk til bokmål. Et eksempel på dette er substantivet *katta* (best., sg., fem.), som i de fleste tilfeller blir diktert til *katter* (ubest., pl., mask.). Tabellen under viser noen eksempler fra datamaterialet:

transkripsjon	«Dikter»
a) åsså mens han då får tak i balljn sin fra innsjøeen så tar katta å et opp fesken hennes	og som en sånn da får tak i ballen sin fra innsjøen så tar katter i et toppesken andres
b) i bilde to har katta håppa mot såmmerfuggeln	i bildet 2 har katter og hopper mot samme fungerende
c) inn mæ en ee gul ballång å en pose mæ pølse	inn med en gul ballong og en pose med pølse
d) men katta sett nu fast i buska å stekk se litt elle nåkka sånt	men katta sittende fast i busker og stek så dette er noe sånt

Tabell 7: Eksempler morfologiske feil av substantiv

Bøyingsystemet til den nordnorske varianten av substantivet *katta* er sterkt hunkjønnord (ei katt – katta – katte – kattn), mens *katter* tilhører hankjønn på bokmål (en katt – katten – katter – kattene). *Katta* på nordnorsk bestemt form singular av ordet *katt*, men «Dikter» velger *katter* (ubes. pl.). Det samme finner vi for ordet *buska* som transkriberes til *busker*. *Buska* har samme bøyingsystem som hunkjønnordet *katta* på nordnorsk, og dette blir også transkribert feil. Til sammenligning blir hankjønnsubstantivet *gutt*n (m., sg., best. form) diktert riktig til *gutt*en i de fleste tilfeller i datamaterialet. Unntaket er et par anledninger hvor «Dikter» velger ordet *gutter*.

Et annet eksempel på morfologiske feil i materialet er hunkjønnordet *pølse* (fem., pl.) på nordnorsk som ble transkribert til *pølse* (fem., sg.) på bokmål. Her ser vi formlikhet mellom entall bestemt form og flertall ubestemt form i input og output. På segmentnivå er det et samsvar mellom det nordnorske ordet som kommer inn og det «Dikter» velger i output, men numerus er ikke korrekt. Med andre ord kan det se ut til at det nordnorske bøyingsystemet for hunkjønnord er utfordrende for «Dikter».

Leksikalske feil utgjør som nevnt 72 prosent av feilene for substantiv og er med dette den majoriteten av kategorien for feil. Leksikalske feil i denne sammenhengen betegner de gangene «Dikter» velger feil ord i output for det informantene sier i input. Noen eksempler fra empirien er de nordnorske substantivene *råtta*, *haue*, *fesken*, *balljn* og *pølse* som blir transkribert til *rota*, *høyde*, *fersken*, *bensin* og *pause* på bokmål:

transkripsjon	«Dikter»
a) men råtta # deinn råtta kryp inn i rota på ett tre	med rota krypinn je ro det bedre
b) så møsta han ballängen åsså slår hunnj haue	som øste av ballongen og så slår hun høyde
c) på bilde nommer sækks så har ee guttn klart å fått- tak i balljn å har den i hanna si mens kattn sett å gommla på fesken	begynner med seks så har gutten klart å få tak i bensin og ha den i hende sin mens katten setter det på fersken .
d) så når står dær å e førnøyd me å ha ballängen sin så sett hunnj dær å kosa seg me pølse	så han står der og er fornøyd med å ha ballongen så sitter hun der og kose seg med pause

Tabell 8: Eksempler leksikalske feil av substantiv

Felles for ordene i den nordnorske inputen og den østnorske outputen, er at de ligner på hverandre. Den leksikalske betydningen av de ulike substantivene stemmer ikke overens. Dette

gjør at semantikken av ordene i nordnorsk input ikke er lik outputen på bokmål. Substantivet *haue* av 'hodet' har en annen semantisk betydning enn *høyde*, til tross for at begge er substantiver og har lik fremlyd *hø-*. Det går et skille mellom de leksikalske feilene av ord som uttales likt på nordnorsk og østnorsk standardtalemål, og uttaleformer som skiller seg fra østnorsk. Ord som *råtta* og *pølse* uttales helt likt på begge talemålene. På den andre siden finner vi ord som har tradisjonell nordnorsk uttale slik som *balljn* (palatalitet), *fesken* (senking) og *haue*. «Dikter» har leksikalske feil av substantiv som er typisk nordnorske ord og av ord som uttales lik på østnorsk og nordnorsk. Dette kan indikere at det er andre faktorer som gir leksikalske feil, enn typisk trekk ved nordnorsk talemål.

4.3 Verb

kategori	antall ord totalt	korrekt	feil
verb	498 ord	318 ord	180 ord

Tabell 9: Antall verb i datamaterialet

Datamaterialet inneholder totalt 498 verb. 64 prosent av disse dikterte «Dikter» korrekt fra nordnorsk talemål til standardspråklig norsk ortografi, bokmål. De resterende 36 prosent av verbene inneholder feil eller mangler i dikteringen mellom tale og skrift.

transkripsjon	«Dikter»
a) på bilde nommer to så håppa katta åver buska førr å kåmm ætter såmmerfuln	er på bildene over tå så hopper katter over budskap for å komme etter sommerfuglen
b) katta prøva å ta såmmerfuggeln	katta prøver å ta samme figuren
c) åsså ser e en kar såm ska på fesketur # så møsta han balljn sin	og så ser jeg en kar som skal på fisketur og så mister han balansen

Tabell 10: Ord med a-endelse i presens på lofotmål

Av de verbene som gjennomgående ble diktert riktig, er blant annet de ordene som har endelsen -a i presens på lofotmål. Eksempler fra datamaterialet er *håppa*, *prøva* og *møsta*, som «Dikter» transkriberer til *hopper*, *prøver* og *mister* (se tabellen over). Informantene sin uttale av verbene med -a i endelse i presens samsvarer med bøyingsmønsteret for verb på lofotmålet som vi finner i talemålsprøver fra unge talere i Stamsund (Johannesen et al., 2009a). Felles for disse ordene er at de har uttale som ligner østnorsk standardspråk, noe jeg kommer tilbake til i drøftingen.

transkripsjon	«Dikter»
---------------	----------

a) i bilde sæks har guttn fådd ned ballången men hunnj har bynnt å spis opp pølsn	i bildet 6 har gutten fått ned ballongen men hun har begynt å spise opp pulsen.
b) så e de sånn deinn hunnj har løst å lek me dennj musa	som er sånn hun har lyst til å leke med mus
c) å katta har klart å kåmm se ut av ee ut av buska å ser bort tell bøtta mæ fesk som guttn har sotte fra seg	og så katter har klart seg å komme seg ut av ut av busker og ser bort til bøtta med flesk som guttene fra seg
d) han bynnj å jag ætter deinn	å begynne å jage etter enda
e) førr å klatre opp å få tak i ballången så sett fast i tree	for å klatre opp få tak i ballongen som sitter fast i treet
f) å guttn har klart å møst ballången sin	og guttene har klart å miste ballongen sin
g) ja så e det en kar såmm kjæm oppi jørne her å han ska ska fesk litt	ja og så er det en kald kommer inn i oppi hjørnet her og han skal skal fiske litt

Tabell 11: Eksempler på apokope transkribert korrekt

Apokope er et talemålstrekk «Dikter» i mange tilfeller transkriberer riktig, ved at «Dikter» legger til suffiksene av ordene. I datamaterialet blir infinitivsformene av verb som for eksempel *spis*, *sett*, *lek*, *møst*, *kåmm* og *jag* (lofotmål) til *spise*, *sitte*, *leke*, *miste*, *komme* og *jage* (bokmål). Uttalen informantene har, samsvarer med apokope av infinitivsformer i det tradisjonelle lofotmålet (jf. Mæhlum & Røyneland, 2012:106). Talemålsprøver fra Stamsund på Vestvågøy viser også at apokope av verb i infinitiv er vanlig, for eksempel *å læs* (Johannesen et al., 2009a). I likhet med ordene med a-endelse over, har også de apokoperte formene en stamme som ellers ligner østnorsk standardtalemål. *Spis* har lik stamme som *spis-e*, dette kan føre til at talegjenkjenningsverktøyet greier å transkribere ordet korrekt. En faktor som ser ut til å være av større betydning for transkripsjonen, er infinitivsmarkøren *å*. Infinitivsmarkøren er en tydelig indikator på at verbet tilhører infinitivskategorien. Dette kan føre til at «Dikter» i større grad greier å transkribere ordene riktig der infinitivsmarkøren er inkludert i frasen.

Eksemplene f) *møst* og g) *fesk* i tabellen over viser verb med apokope hvor ordstammen skiller seg fra østnorsk standardtalemål. Det er interessant at disse transkriberes riktig ettersom at verbene er typiske for nordnorsk talemål, og skiller fonetisk fra østnorsk. *Fesk* ('fiske', inf.) og *møst* ('miste', inf.) er ord som viser fenomenet senking (Mæhlum & Røyneland, 2012:111). Stammen *møst* og *fesk* er ulik stammen for bokmålsordet *mist-e* og *fisk-e*, men dikteringen blir korrekt for disse ordene. Utover den senkede vokalen er stammen for ordet likt på nordnorsk og østnorsk standardtalemål. Det er særlig interessant at verbet *fesk* (av 'fiske') blir transkribert riktig, ettersom dette ordet også kan være et substantiv. I analysen av substantivene blir *fesk* (av 'fisk') transkribert korrekt i de fleste tilfeller (se tabell 4 i kapittel 4.1 Fonologiske trekk). «Dikter» greier med andre ord å skille mellom transkriberingene av *fesk* for verb og substantiv.

antall dikteringsfeil verb	leksikalske	morfologiske	annet
180 ord	98 ord	58 ord	24

Tabell 12: Oversikt over antall leksikalske og morfologiske dikteringsfeil av verb

De 180 verbene «Dikter» ikke dikterte riktig kan, i likhet med substantivene, deles inn i to kategorier, leksikalske og morfologiske feil. Datamaterialet inneholder 98 verb hvor «Dikter» valgte feil ord for det nordnorske, derav leksikalske feil. Av de morfologiske feilene var det totalt 58 verb som ble bøyd feilaktig i dette datamaterialet. De resterende 24 feilene i datamaterialet ble ikke oppfattet av «Dikter», og er dermed ikke inkludert i dikteringen. De utelatte ordene har jeg valgt å ekskludere her fordi det kun omfatter et mindre antall ord som «Dikter» ikke har greid å oppfatte. I disse tilfellene er det trolig andre faktorer som taletempo og informantenes diksjon som avgjør dikteringskvaliteten, og det regnes derfor ikke som interessant videre i denne analysen.

transkripsjon	«Dikter»
a) så slår hunnj hode å kiddn <u>mista</u> ballången oppi tree	så slår den i hodet og krydder <u>mistet</u> ballongen opp i treet
b) guttn får tak i ballången # å bikkja <u>kosa</u> se me pølsn	akutten får tak i ballongen og bikkja <u>kose</u> deg ned pulsen
c) så e de sånn katta gre å <u>kåmm</u> seg ut	ser ut som kataloger <u>kom</u> seg ut
d) <u>snik</u> han seg ruinnt åsså stæl han fesken ijænn	<u>snike</u> seg rundt og så står han fiskeren
e) han <u>bynnj</u> å jag ætter dennj	å <u>begynne</u> å jage etter enda

Tabell 13: Morfologiske feil av verb

Innenfor de morfologiske feilene av verb finner vi en rekke ord som transkriberes feilaktig fra lofotmål til standardspråklig norsk ortografi. Presensformene *mista* og *kosa* på lofotmål transkriberte «Dikter» til *mistet* (pret.) og *kose* (inf.) på bokmål (se a) og b) i tabellen over). I lofotmålet har tradisjonelt svake verb i klasse 1 endelsen -a i presens (Vangsnes & Westergaard, 2019:367). De nordnorske variantene informantene har, er med andre ord i tråd med tradisjonell nordnorsk bøyning av verb med a-endelse i presens. «Dikter» transkriberte de nordnorske presensformene til preteritums- og infinitivsform på bokmål, altså til andre morfologiske kategorier enn inputen.

Den samme tendensen viser seg i andre eksempler fra tabellen over: *kåmm* (inf.), *snik* (pres.) og *bynnj* (pres.). Disse verbene blir transkribert feil av «Dikter» til ordene: *kom* (pret.), *snike* (inf.) og *begynne* (inf.). *Kåmm* er et eksempel på apokope, hvor utlydende vokal i slutten av ordet faller bort (jf. Jahr & Skare, 1996:26). Tabell 11 over viste at «Dikter» i stor grad håndterer verb med apokope, og transkriberer disse til korrekte former. Her ser vi derimot et eksempel som viser det motsatte, nemlig at «Dikter» ikke greier å transkribere ord med apokope til korrekt morfologisk kategori. For de andre presensformene *snik* (av ‘sniker’) og *bynnj* (av ‘begynner’) velger «Dikter» infinitivsformer i output. Felles for alle eksemplene er at «Dikter» velger korrekte ord, men feil morfologisk form for ordene. Verktøyet har med andre ord utfordringer med å gjenkjenne det nordnorske bøyingsystemet for verb.

transkripsjon	«Dikter»
a) mens musa står på andre sia å smil	og musa står på andre siden og smil
b) musa i bilde to løp inn i røttern te trex å hunnj jaga ætter	musa i bilde 2 løp inn i under røttereng til 3 og hun jaget etter
c) så kåmmer guttn inn # ser situasjon # møsta ballången	så kommer gutten inn se situasjoner mester ballongen
d) åg gådd bort te tree får ballången har sotte se fast i tree	og godt bort 3 for ballongene som vi fant 3
e) så hann mista balljn sin å deinn fer utti vannje	så har mista balansen og den den ser ut i vannet

Tabell 14: Leksikalske feil av verb

Tabellen over viser noen av de leksikalske feilene for den grammatiske kategorien verb. I eksempel a) og b) blir den nordnorske inputen *smil* (pres., ‘smile’) og *løp* (pres., ‘løpe’) til *smil* (imp.) og *løp* (imp.) i output. De nordnorske variantene viser apokope av presens som kjennetegner den statiske apokopen i lofotmålet (Christiansen, 1996[1950]:94). Transkripsjonene for disse ordene har to mulige tolkninger for hvordan feilen oppstår. Den første tolkningen kan regne transkripsjonene som feil morfologisk fordi det er variasjon mellom den nordnorske inputen i presens og outputen på bokmål imperativ. Her vil jeg likevel hevde at feilen bør tolkes og kategoriseres som leksikalske feil. Dersom man tolker eksemplene som leksikalske feil, kan det hevdes at «Dikter» ikke greier å skille mellom de grammatiske kategoriene verb og substantiv. I lys av en slik forståelse transkriberes verbene *smil*- og *løp*- (pres.) og *møsta* (pres., ‘miste’), til substantivformene (et) *smil* og *løp* (ubest., sg. / ubest., pl.) og (en) *mester* (ubest., sg.). Fra et fonetisk perspektiv er ordene diktert riktig, men semantikken av ordene stemmer ikke. «Dikter» skiller ikke mellom verb med apokope i presens og substantiv

for ord som ellers er identiske. Med andre ord velger «Dikter» de østnorske formene av ord som er korrelerer på nordnorsk og østnorsk. Unntaket er *møsta*, som ikke ligner på *mester* rent fonologisk på grunn av senkingen av den nordnorske varianten. Det er interessant at «Dikter» også her velger et substantiv i output av den nordnorske inputen. Dette kommer jeg nærmere inn på i drøftingsdelen under.

Andre funn for leksikalske feil av verb er ord som *gådd* (pret., ‘gå’) og *fer* (pres., ‘fare’) på lofotmål. For disse variantene velger «Dikter» ordene *godt* (adj., intetkjønn, sg.) og *ser* (pres.). Semantikken av bokmålsvariantene i output stemmer ikke med den nordnorske inputen. Felles for ordene i input og output er at de ligner på hverandre fonologisk. For eksempel har *gådd* og *godt* lik fremlyd, mens *fer* og *ser* er minimale par. På den annen side er det det fonologiske systemet på nordnorsk som skaper utfordringer for «Dikter», ettersom at det er her problemene oppstår.

4.4 Personlig pronomen

Innenfor den grammatiske kategorien pronomen, er personlig pronomen 1. person subjektsform mest interessant ettersom denne formen skiller seg tydelig fra de østnorske variantene med j-lyd, slik som «jæ(i)» eller «je» (Mæhlum & Røynealand, 2012:52). Personlig pronomen forekommer ikke så mange ganger i datamaterialet fra «Dikter», men det finnes noen eksempler. I intervjuene kommer personlig pronomen frem hos samtlige informanter. Informantene bruker *e* og *eg* som personlig pronomen 1. person. Informant 6 bruker konsekvent *eg*, mens informant 1 og 2 veksler mellom formene *eg* og *e*. Pronomenet *e*, blir brukt konsekvent av tre informanter og utgjør derfor flertallet i datamaterialet. Variantene *e* og *eg* er i tråd med personlig pronomen 1. person på nordnorsk (Jahr & Skare, 1996:37). Funn viser at varianten *eg* i samtlige tilfeller blir transkribert til *jeg*. «Dikter» transkriberer altså det nordnorske pronomenet *eg* til korrekt standardspråklig norsk ortografi. Varianten *e* blir i de fleste tilfeller også transkribert korrekt. Unntakene er de gangene *e* ikke blir registrert og derfor ikke kommer med i dikteringen, og dette skjer bare hos informant 4. Personlig pronomen første person entall i nordnorsk talemål er med andre ord et trekk «Dikter» håndterer på en god måte.

transkripsjon	«Dikter»
a) så <u>e</u> ser ei katt å deinn fer bærre å mora seg	så <u>jeg</u> jeg ser ikke at å det er en feil begynt å more seg
b) <u>e</u> kainn vel sei hovedkarakteren såm de ser ut	<u>jeg</u> kan vel se hovedkarakteren som ser ut
c) åkei <u>eg</u> ser ei katt mæ et vattn	det <u>jeg</u> ser en katt med et vatn

Tabell 15: Pronomen fra datamaterialet

4.5 Spørreord, form og syntaks

Spørreord er i likhet med personlig pronomen, et talemålstrekk som skiller nordnorsk fra østnorsk talemål. I datamaterialet hvor informantene bruker «Dikter» forekommer det noen få spørreord. Et eksempel er spørreordet *ka* på lofotmål, som «Dikter» transkriberer korrekt til *hva*. I intervjuene er det flere eksempler på spørreord fordi dette er en toveiskommunikasjon mellom forsker og informant. Her forekommer ord som *ka* og *kordan*, og disse transkriberes til *hva* og *hvordan* av «Dikter». De nordnorske spørreordene med stor uttrykksforskjell fra det østnorske talemålet. Med dette som grunnlag kan man forvente at transkripsjonen av spørreord ikke blir optimal. Derfor er det overraskende at hovedvekten av spørreordene i undersøkelsen blir transkribert korrekt, til tross for talemålsvariasjonene.

transkripsjon	«Dikter»
a) e veit ikkje kordan man sei det sørlænninga	jeg vet ikke hvordan man er søvning da
b) ka du meina me de	hva du mener med det
c) å i bilde sæks så har guttn fått tak i balljn å katta spis opp fesken hannes	og i bildet 6 så har gutten fått tak i ballen og katter spiser opp fisken hadde
d) så reaksjonan hannes bærre kikka inn	så reaksjonene _ bare klikke inn

Tabell 16: Eksempler spørreord, form og syntaks

Et annet interessant funn er at «Dikter» beholder den nordnorske syntaksen for spørresetninger i transkriberingen. I et par av intervjuene spurte informantene blant annet *ka du meina* til enkelte spørsmål (se eksempel b) over). Dette er eksempel på en spørresetning uten intervensjon, som er typisk for nordnorske talemål (Jahr & Skare, 1996:12). «Dikter» transkriberer disse sekvensene til *hva du mener*. Dette betyr at verktøyet transkriberer ord for ord, og tar derfor ikke hensyn til syntaksen. Spørresetningen er et eksempel som skiller seg fra syntaksen i standardspråklig norsk ortografi, *hva mener du*. Talegjenkjenningsverktøyet beholder med andre ord den syntaksen som er vanlig i nordnorsk talemål, der ordstillingen ved spørreord ikke nødvendigvis følger V2-regelen (Vangsnes & Westergaard, 2019:169). Dette fører til at «Dikter» gir syntaksstrukturer som ikke er vanlig i skriftlig standardmål. Med utgangspunkt i V2-reglen i norsk grammatikk blir dikteringen regnet som ukorrekt. På en annen side er det vanlig at språkteknologien jobber med språkdata på denne måten, noe jeg kommer tilbake til i diskusjonsdelen.

Andre observasjoner av syntaks som også er relevant å nevne, er possessiver. Funnene i denne undersøkelsen viser at «Dikter» håndterer possessiver på en utilstrekkelig måte. Et eksempel på dette kan vi se i tabellen i eksempel c) og d). I det første eksempelet ser vi at

possessive *hannes* blir til *hadde* på bokmål. Dette viser at «Dikter» velger et helt annet leksikalsk ord for possessivet. Andre ganger utelukker «Dikter» ord i output for possessiver i input. I eksempel d) i tabellen over er den nordnorske inputen: *så reaksjonan hannes bærre kikka inn*. Mens outputen til «Dikter» er: *så reaksjonene X bare klikke inn*. Her ser vi at «Dikter» unnlater å velge ord i output for det nordnorske possessive *hannes*. Begge eksemplene er i tråd med nordnorsk talemålsteori der possessive pronomen er etterstilt. I tillegg er *hannes* en dialektal variant av *hans* i det nordnorske talemålet. Med andre ord er dette et ord som ikke finnes i østnorsk talemål. «Dikter» håndterer altså konstruksjoner med etterstilte possessiver på en dårlig måte, gjennom at «Dikter» velger andre ord for possessivene eller at transkripsjonen ekskluderer disse ordene.

4.6 Informantenes tilpasning av eget talemål

Gjennom å sammenligne talemålet til informantene i intervjuene med deres talemål når de brukte «Dikter», kan studien finne eventuelle tilpasninger informantene gjør av sitt eget talemål. Et sentralt funn er at informantene i liten grad gjør tilpasninger av eget talemål når de bruker talegjenkjenningsverktøyet «Dikter». Dette gjelder for de typiske nordnorske talemålstrekkene som er omtalt over, slik som for eksempel palatalitet, retroflektering, senking, apokope og spørreord. Funnet er nokså overraskende ettersom det ifølge Akselberg et al. (2008:160) er vanlig å tilpasse talemålet sitt til personer eller situasjoner. Informantene har altså ingen tydelig tegn på at de forandrer talemålet dikteringsoppgaven sitt i. Enkelte informanter gjentok ord eller rettet på seg selv dersom de sa ufullstendige ord eller setninger. Dette regnes ikke som tilpasninger av talemål fordi det er vanlig i det muntlige språket.

Et unntak finner vi hos informant 1 som ikke har palatalitet i enkelte ord i dikteringssituasjonen, men som konsekvent har palatalitet i intervjuet. Dette viser ser blant annet i frasen: *bilde nommer fæm så ee dræg guttn balln opp fra vanne mæ feskestanga*. Ordene *balln* og *vanne* blir uttalt uten palatalitet, noe som markerer et skille fra det resterende datamaterialet. Med bakgrunn i intervjusituasjonen har informanten palatalitet som en del av sitt autentiske talemål, hvor ord med konsonantsambanene *nn* og *ll* får en j-lyd. I dikteringssituasjonen har informantene forekomster både med og uten palatalitet. Dette kan være en intraindividuelle i talemålet, fordi vi finner variasjon hos ett og samme individ (Norrby & Håkansson, 2015:60). På en annen side er det vanskelig å avgjøre om dette er tilfeldig eller en faktisk tilpasning av talemålet. I tillegg er det problematisk å sammenligne forekomster av talemålstrekk i to situasjoner som begge inngår i en formell forskningssituasjon. Dette innebærer at eventuelle tilpasninger av talemålet kan skyldes den formelle situasjonen. Med

andre ord er det grunn til å tro at informantene ikke tilpasser talemålet sitt til dikteringssituasjonen, men snakker sitt autentiske talemål når de bruker «Dikter».

4.7 Oppsummering funn

For å oppsummere funnene ser det ut til at talegjennkjenningsverktøyet «Dikter» ikke fungerer optimalt for talere med nordnorsk talemål. Det oppstår en del feil og mangler i dikteringen fra nordnorsk talemål til norsk standardortografi. Feilene oppstår på tvers av grammatiske kategorier og henger i mange tilfeller sammen med det fonologiske systemet i nordnorsk talemål. På bakgrunn av funnene over kan man lage en oversikt over hvordan «Dikter» håndterer ulike nordnorske talemålstrekk, samt hvordan informanter tilpasser seg dikteringssituasjonen. Dette blir naturligvis en generalisering av trekkene, men det vil være viktig grunnlag for den videre drøftingen. Oppsummeringen er som følger:

Palatalitet er et fonologisk trekk som «Dikter» ikke håndterer på en tilstrekkelig måte. Dette gjelder særlig for palatalitet av den grammatiske kategorien substantiv, i ord som *hunnj* og *ballj*. For kategorien verb er det flere forekomster av korrekte transkripsjoner av ord med palatalitet.

Retroflektering er et talemålstrekk «Dikter» håndter godt. Dette gjelder blant annet for konsonantsammenfallet *rl* (*dårli*), *rt* (*jort*) og *rn* (*hovedkarakteren*).

«Dikter» sin håndtering av talemålstrekket senking er todelt. Det er større sannsynlighet for korrekt transkripsjon av substantiv i ubestemt form, som *fisk*, sammenlignet med den bestemte formen *fisken*. For verbformene *fesk* og *møsta* blir transkripsjonene som regel transkribert korrekt.

Apokope er et trekk som «Dikter» håndterer både korrekt og feil. Håndteringen fungerer godt der infinitivsmarkøren er inkludert, og mindre bra der denne ikke er med.

Personlig pronomen 1. person entall er et trekk «Dikter» håndterer på en god måte. Begge variantene, *eg* og *e*, transkriberes riktig.

Spørreord, som *ka* og *kordan* blir transkribert korrekt, og utgjør derfor en kategori «Dikter» håndterer godt.

Syntaksen blir ikke behandlet av «Dikter», hvilket betyr at ordene i samme rekkefølge som de sies. Dette innebærer at ordstillinger som viker fra standardsspråklig norsk ortografi, for eksempel spørresetninger uten intervensjon, blir transkribert slik.

Possessiver er et annet trekk som «Dikter» ikke håndterer på en tilstrekkelig måte. Innenfor denne kategorien velges enten helt andre ord som erstatning eller så utelater «Dikter» possessivene.

Når det gjelder tilpasninger av talemål til dikteringssituasjonen er det lite som tyder på at informantene endrer på lofotmålet sitt. Talemålstrekkene som er nevnt over forekommer hos alle informantene i større eller mindre grad, og i både intervjusituasjonen og når de bruker «Dikter».

5 Diskusjon

I denne delen vil jeg drøfte funnene i lys av relevant teori. Drøftingsdelen vil besvare studiens forskningsspørsmål; 1) Hvordan transkriberer «Dikter» nordnorske talemålstrekk til standardspråklig norsk ortografi og 2) Hvordan tilpasser nordlendinger talemålet sitt når de bruker talegjenkjenningsverktøy? For det første spørsmålet vil diskusjonen legge vekt på årsaker til at det oppstår feil mellom den nordnorske inputen og outputen på bokmål.

5.1 Hvordan transkriberer «Dikter» nordnorske talemålstrekk?

Som det kommer frem i delkapittel 4.7 fungerer ikke talegjenkjenningsverktøyet «Dikter» optimalt for talere med lofotmål. Det er særlig det fonologiske systemet i nordnorsk skaper utfordringer for talegjenkjenningsverktøyet. De feilaktige transkripsjonene er knyttet til både det leksikalske og til det morfologiske systemet. Palatalitet, syntaks og possessiver er talemålstrekk som ofte får feil output på bokmål. For senking og apokope er det variasjon i korrekte og feilaktige transkripsjoner av «Dikter». De talemålstrekkene som «Dikter» håndter godt, og transkriberer riktig er blant annet retroflekttering, personlig pronomen 1. person entall og spørreord.

Felles for de talemålstrekkene «Dikter» transkriberer feil, er at disse trekkene er typiske for nordnorsk talemål og skiller seg fra det østnorske standardtalemålet. Dette gjelder primært palatalitet, possessiver og syntaks. I tillegg er det variert kvalitet på transkripsjonene ved forekomst av senking og apokope. Østnorsk standardtalemål har ingen av disse talemålstrekkene (jf. Mæhlum & Røyneland, 2012:59). Ord med nordnorske talemålstrekk gir både leksikalske og morfologiske feil, som vist i delkapittel 4.2 og 4.3. Det fonologiske systemet i nordnorsk fører til en del leksikalske feil. Typisk for de leksikalske feilene er at «Dikter» velger output som ligner på den nordnorske inputen, for eksempel at ordene har lik fremlyd eller utgjør minimale par.

En undersøkelse gjennomført av Davidsen (2013) viser at det norske talegjenkjenningsverktøyet Maxmanus dikterer en rekke legnavn feil. Et par eksempler fra undersøkelsen er; Dr. Gjorde En Rekke (dr. Jorunn Brekke) og Dr. Lorentz Er Vanlig Under (dr. Lorentz Erland Linde). Studien til Davidsen (2013) gir ingen informasjon om talemålene til legene som brukte talegjenkjenningsverktøyet. I denne studien har vi sett at transkripsjonene til «Dikter» får en del leksikalske feil innenfor flere grammatiske kategorier. De nordnorske variantene av *ballj* og *hunnj* med palatalitet blir ofte transkribert feil av «Dikter». Palatalitet er et trekk som ikke eksisterer i østnorsk standardtalemål (jf. Mæhlum & Røyneland, 2012:59).

Dette tyder på at «Dikter» ikke greier å gjenkjenne palatalitet i talespråket. Det samme viser seg for inputen, *gådd* ('gått', verb, pret.) og *haue* ('hodet', sub.), som blir til *godt* og *høyde* på bokmål. I likhet med palatalitet er dette uttaleformer som er typisk for det nordnorske talemålet. Endelsen -dd markerer et skille fra den østnorske formen -tt, i ord som *gådd* og *nådd*. Kvaliteten på transkripsjonene ser med andre ord ut til å henge sammen med den fonologiske avstanden mellom lofotmål og østnorsk standardtalemål. Funnene i denne undersøkelsen skiller seg fra Davidsen (2013) ved at «Dikter» i større grad velger ord som ligner på inputen. Dette kan henge sammen med at inputen i Davidsen (2013) var egennavn, mens inputen i denne undersøkelsen er vanlige ord. Til tross for dette, ser vi likheter i output og input i begge studiene ved at verktøyene velger output som ligger nært inputen.

Apokope blir i en del tilfeller transkribert med feil endelse av «Dikter», og gir derfor en del morfologiske feil. Eksempler på denne er ord som *kåmm* (inf.), *snik* (pres.) og *bynnj* (pres.) (se delkapittel 4.3 Verb). «Dikter» velger feil bøyingsform av disse ordene i output. Talemålstrekket apokope i input kan være årsak til at «Dikter» ikke gjenkjenner det morfologiske systemet i lofotmålet, og følgelig transkriberer disse formene feilaktig.

Andre ganger gir språktrekket apokope leksikalske feil. Verb med apokope blir til substantiv i transkripsjonen, (å) *løp* blir (et) *løp* (se tabell 14 i delkapittel 4.3 Verb over). Ordene viser formlikhet mellom verb med apokope og substantiv. Outputen til «Dikter» indikerer at verktøyet kun har språkdata for substantivet *løp*, og av den grunn ikke greier å transkribere ordet som verb på bokmål. «Når den ønskede analysen ikke finnes, må grammatikken og/eller leksikonet oppdateres» (Rosén, 2014:20). Språktrekket apokope ser med andre ord ut til å mangle i databasene som ligger til grunn for «Dikter». Apokope finnes i nordnorske og trønderske talemål (Mæhlum & Røynealand, 2012:42). Språktrekket er med andre ord fraværende i østnorske dialekter.

«Dikter» ser også ut til å ha problemer med å gjenkjenne det nordnorske bøyingsystemet. I datamaterialet mitt blir for eksempel ordet *katta* (best. sg., fem.) diktert til *katter* (ubest. pl., mask.) i de fleste tilfeller. *Katta* (best. sg., fem.) er i tråd med bøyingsystemet for lofotmålet med bortfall av r i substantiv ubestemt flertall (Jahr & Skare, 1996:12). Dikteringen viser at «Dikter» ikke håndterer bestemt form entall av den nordnorske inputen, ettersom outputen kommer i flertallsform. En mulig forklaring på dette er at «Dikter» tolker -a-endelsen i *katta* som en maskulin flertallsendelse (jf. *hæsta*), som passer med at *katt* er maskulin på bokmål. På en annen side kan andre talemål dominere over nordnorsk i dikteringen. Bøyingsystemet for *katt* på nordnorsk er ei *katt* – *katta* – fleir *katte* – alle *kattn*. Dersom vi sammenligner bøyingsmønsteret med trøndersk talemål bøyes ordet *katt* – *katta* – fler *katta* –

allj *kattan*. Både bestemt form entall og ubestemt for flertall på trøndersk av ordet katt, er *katta*. For nordnorsk talemål er det bare bestemt form singular som blir *katta*. «Dikter» velger flertallsformen ubestemt form *katter*, for den nordnorske inputen *katta*. Det kan være en indikasjon på at talegjenkjenningsverktøyet i større grad kjenner det morfologiske systemet for trøndersk talemål, og at dette ser ut til å dominere over nordnorsk.

5.2 Nordnorsk talemål VS østnorsk standardtalemål

Feilene i dikteringen ser med andre ord ut til å oppstå der det er differanse mellom det nordnorske talemålet og østnorsk standardtalemål. «Dikter» velger bokmålsord som fonologisk ligner på de nordnorske formene. Funnene viser at talegjenkjenningsverktøyet derimot har lite utfordring med ord som på segmentnivå er like på nordnorsk og østnorsk standardtalemål. Språktrekk som retroflektering blir i de fleste tilfeller diktert riktig fra nordnorsk talemål til bokmål. Retroflektering finner vi i ord som *dårli* ('dårlig'), *jort* ('gjort') og *hovedkarakteren* ('hovedkarakteren') i datamaterialet. Retroflektering er et trekk som også finnes i østnorske talemål (Mæhlum & Røyneland, 2012:47). Dette kan føre til at «Dikter» gjenkjenner språktrekket og kan gi korrekte transkripsjoner.

Antakelsen forsterkes ved å se på ord som uttales likt på lofotmål og østnorsk standardtalemål. For eksempel blir *ballång* gjennomgående transkribert korrekt til *ballong*. Dette ordet bærer ikke preg av nordnorske språktrekk, men styrker tanken om at likhet mellom nordnorsk talemål og østnorsk standardtalemål gir større sannsynlighet for korrekt transkripsjon. Det samme gjelder for verb med -a-endeelse i presens, som *håppa* og *prøva*. Morfologien i disse formene skiller seg fra østnorsk standardtalemål, men «Dikter» greier likevel å transkribere disse formene. Stammen i ordene *håpp-* og *prøv-* er dessuten identiske for nordnorsk talemål og østnorsk standardtalemål. Dette er ord som rent fonetisk ligger nært østnorsk standardtalemål. Informantene i denne studien uttaler ordet *håppa* i presens, mens den østnorske standardtalemål uttalelsen av samme ord er *håpper*. Fremlyden på både lofotmål og østnorsk standardtalemål er lik, *håpp-*. Avledningen -a og -et greier talegjenkjenningsverktøyet å skille mellom i dikteringen.

På bakgrunn av disse funnene er det mulig å hevde at desto likere den nordnorske inputen er det østnorske standardtalemål, jo større sannsynlighet er det for at ordet blir transkribert riktig. Indirekte betyr dette at verktøyet håndterer østnorsk standardtalemål bedre enn nordnorske talemålstrekk fordi normen ligger på det østnorske talemålet. I korrespondanse med Per Erik Solberg i Språkbanken kommer det frem at Språkbanken har gjennomført interne prosjekter hvor de finner ut det samme som denne studien viser, nemlig at

talegjenkjenningssystemer for norsk håndterer noen dialekter bedre enn andre. «Gjenkjenningen av østlandsk er ofte ganske god, mens det ofte blir mer feil med dialekter som er annerledes enn østlandsk» (Solberg, 2022, personlig korrespondanse). Det kan tyde på at talegjenkjenningverktøy som «Dikter» håndterer østnorsk standardtalemål bedre enn nordnorsk talemål.

Det østnorske standardtalemålet fungerer altså som en rettesnor for hvorvidt «Dikter» transkriberer ulike nordnorske talemålstrekk. Østnorsk standardtalemål får med dette en form for opphøyet posisjon over de andre talemålene. Posisjonen til østnorsk i talegjenkjenningsverktøy er i tråd med Mæhlums forståelse av standardtalemål. Mæhlum (2009) hevder at «Et suksessfullt standardiseringsforløp impliserer at én varietet blir «valgt» framfor alle andre varieteter – les dialekter, og denne varietetten tilhører i utgangspunktet grupper og personer med status og prestisje i samfunnet» (Mæhlum, 2009:9). Empiri fra denne studien viser at statusen til østnorsk standardtalemål gjelder like fullt innenfor utviklingen av språkteknologiske verktøy. Den østnorske standarden blir med andre ord normgivende for talegjenkjenningsverktøy og deres håndtering av ulike dialekter.

«Dikter» operer kun med bokmål som mulig output for talegjenkjenning. Dette betyr at man ikke kan få nynorsk diktering med verktøyet. Videre er det mye av outputen i funnene som peker i retning av at moderat eller konservativt bokmål ligger til grunn for «Dikter». Ord som for eksempel *kasta* med nordnorsk uttale, blir til *kastet* på bokmål. Den nordnorske talemålsvarianten *kasta* er identisk med formen på radikalt bokmål, hvor -a-endelse dominerer (jf. Hårstad et al., 2021:125). Det samme ser vi for substantiv hvor hunkjønn kommer ut med suffikset suffikset -en og ikke -a, slik som at *buska* i nordnorsk input blir til *busken* (se tabell 7 i delkapittel 4.2). Hårstad et al. (2021) skriver at det er konservativt bokmål som dominerer i bokmålstekster, med bøyingsendelsene -en og -et i substantiv og -et i svake verb (Hårstad et al., 2021:17). Bokmålsnormen er vid og tillater former som assosieres med konservativt bokmål, så vel som former som assosieres med radikalt bokmål. Til tross for dette, velger «Dikter» den konservative bokmålsvarianten, i ord som for eksempel *kastet* og *busken*. Normen i verktøyet er altså konservativt bokmål, selv om radikalt bokmål ligger nærmere nordnorsk dialekt. «Dikter» holder seg til en snever norm i output og tilpasser seg altså ikke til input ved å velge for eksempel a-former der dette blir sagt. Dette henger naturlig sammen med at konservativt bokmål er den formen som dominerer og har en høyere status i språksamfunnet ellers, slik som Hårstad (2021) trekker frem. Dominansen av konservativt bokmål i bokmålstekster samsvarer med funnene i denne undersøkelsen, hvor det ser ut til at talegjenkjenningsverktøyene i stor grad baser seg på tekstmateriale som følger den konservative

standarden. Her er det en parallell til andre språkteknologiske verktøy, som blant annet Google Translate. Selv om det verktøyet kan ta både nynorsk og bokmål som input ved oversettelse til for eksempel engelsk, vil alle oversettelser til norsk komme på bokmål. Disse oversettelsene blir også korrigert til konservativt bokmål, slik som «Dikter». Språkteknologiske verktøy godtar altså språklig variasjon i input, men har ingen variasjon i egen output.

For å forstå hva som regulerer outputen til «Dikter» må vi se nærmere på bakgrunnen for talegjenkjenningsverktøyene. Talegjenkjenningsverktøy behandler tale om til tekst med utgangspunkt i en database som består av en mengde ord. Eide & Våge (2018) hevder at «Skal språkteknologiske produkter fungere tilfredsstillende på norsk, trenger de store mengder gode norske språkdata». Språkteknologiske verktøy er med andre ord avhengig av språkdata for å kunne utføre operasjoner som tale-til-skrift. Talegjenkjenningssystemer bygger på en maskinlæringsmodell som er trent på store mengder transkriberte taledata (Digitaliseringsdirektoratet, u.d). Hvilke ord og talemålstrekk talegjenkjenningsverktøyene registrerer henger altså sammen med hvilke språkdata som er tilgjengelige for verktøyene. Per Erik Solberg i Språkbanken, trekker han frem tilgangen på transkriberte taledata er en hovedfaktor til at feilene oppstår (Solberg, 2022, personlig korrespondanse). Dette samsvarer med funnene fra undersøkelsen min, hvor ord og språktrekk som er fonologisk like østnorsk standardtalemål som regel blir transkribert riktig. Dette gjelder blant annet retroflektering, som er felles for både det nordnorske og østnorske standardtalemålet. Med utgangspunkt i empirien fra denne studien kan det hevdes at «Dikter» i stor grad er basert på østnorske talespråkdata. For de nordnorske talemålstrekkene finnes det langt færre eksempler på språkdata enn for østnorske språkdata, noe som gir utgangspunkt for feil transkribering, særlig leksikalsk feil. Som vist i kapittel 4 gjelder dette særlig for palatalitet, senking, apokope og possessiver. Disse kategoriene finnes det trolig begrensede språkdata for ettersom at disse skiller seg fra østnorsk standardtalemål. Et eksempel er senking i ordet *fesk* ('fisk'), hvor funnene viser variasjon mellom bestemt og ubestemt form entall. Fersken blir en "bedre" kandidat for *fesken* enn *fisken* i talegjenkjenningsverktøyet, mens for *fesk* er det ikke noen annen feilkandidat som utkonkurrerer *fisk* (se kapittel 4.1 Fonologiske trekk). Mangelen på variasjon i norske dialekter i datasettene gjør at språkteknologien er bedre utviklet for østnorsk enn nordnorsk.

5.3 Mulige påvirkningsfaktorer for transkriberingen

Til tross for at østnorsk standardtalemål ser ut til å fungere som en rettesnor for «Dikter», greier verktøyet å transkribere enkelte ord som er typiske for nordnorsk talemål. Høyfrekvente ord regnes som den første faktoren som kan gi korrekt transkripsjon, selv om den nordnorske

inputen skiller seg fra østnorsk. Eksempler på dette er de personlig pronomene *e* og *eg*, verbet *e* ('være', pres.) og spørreordene *ka* og *koffør*. Felles for disse kategoriene er at de er høyfrekvente. Dette betyr at de forekommer hyppig og at det sannsynligvis finnes en større mengde språkdata for disse formene. Derfor er det rimelig å anta at dette forklarer hvorfor nettopp disse formene transkriberes korrekt, til tross for at de er fonologisk svært forskjellig fra østnorsk standardtalemål. Som vi har sett over, er talegjenkjenningsverktøy matet med en viss mengde språkdata som ligger til grunn for dikteringen. Flere ord og språkdatagrunnlag gir større sannsynlighet for korrekt diktering.

En annen faktor ved de høyfrekvente ordene i datamaterialet, er at de også forekommer i flere andre norske dialekter. *Eg* og *e* som personlig pronomen 1. person subjektform finner vi også i trøndersk og vestnorske talemål (jf. Mæhlum & Røyneland, 2012:182). Det samme gjelder for verbformen *e* ('være') og spørreordene *ka* og *kordan* som eksisterer i flere andre talemål utover nordnorsk. Dette betyr at datagrunnlaget for ulike former er felles for flere norske dialekter, og følgelig at språkdatabasene til «Dikter» trolig er større når grunnlaget kommer fra ulike talemål. Likhet i talemålstrekk for nordnorsk og andre talemål gir «Dikter» større datagrunnlag for å utføre korrekte dikteringer. I en intern av Språkbanken av Google talegjenkjenningsverktøy gjennomført av Språkbanken, finner de at verktøyene håndterer nordnorsk bedre enn trøndersk og vestlandsk (Solberg, 2022, personlig korrespondanse). En svakhet for denne undersøkelsen er at den bare undersøker hvordan «Dikter» håndterer nordnorsk talemål. Det betyr at undersøkelsen ikke har grunnlag for å si noe om «Dikter» sin funksjonalitet på nordnorsk sammenlignet med andre talemål. Uavhengig av dette er det større sannsynlighet for korrekte transkripsjoner for språktrekk som er felles på flere talemål.

For det tredje faktor som ser ut til å påvirke hvorvidt transkripsjonene blir korrekte er konteksten inputen kommer i. Kontekst i denne sammenhengen handler om hvilken sekvens den nordnorske inputen kommer i. For at talegjenkjenningverktøyet skal regne en setning som en sekvens må informanten si setningen i ett uten pause. Dersom informanten deler opp setningen med pauser, vil ikke nødvendigvis «Dikter» se setningen som helhet selv om det er det. Dersom ubestemt artikkel (en, ei eller et) inngår i frasen informanten sier, er det større sannsynlighet for at ordet transkriberes riktig, uavhengig av nordnorske talemålstrekk. Et eksempel på dette fenomenet er når informantene sier *en hunnj* med palatalitet og «Dikter» transkriberer dette korrekt til *en hund*, slik som i denne sekvensen: *ja de e et døtt tre # ei mus å en hunnj å et hus i bakgruinn*. «Dikter» output: *det er 3 er mus og en hund og hus i bakgrunnen*. Til tross for at ikke alle artiklene er inkludert i transkripsjonen, blir *hund* med palatalitet transkribert korrekt.

Det samme gjelder for input med apokope dersom ordrekken består av infinitivsmarkøren å. Som vist i tabell 11 i delkapittel 4.3 ser en høy forekomst av apokoperte former med infinitivsmarkøren som blir transkribert korrekt. Kontekstbaserte merker som ubestemt artikler, infinitivsmarkøren, determinativer (den, hans, en osv.) og preposisjoner (i, på, over, under osv.) kan altså påvirke kvaliteten av transkripsjonen til «Dikter». Senking er et annet språktrekk som blir korrekt når det inngår i en kontekst. Tabell 3 i delkapittel 4.1 viser at *feskebøtta* uttalt med senking blir korrekt til *fiske bøtta*, selv om ordet er delt oppi output. Den korrekte transkripsjonen er forskjellig fra de feilaktige variantene av *fesk/en* (se tabell 8 i delkapittel 4.2 Substantiv). Det kan se ut til at konteksten senkingen presenteres i, *feske**u**øtta*, påvirker transkripsjonen.

Davidson (2013) argumenterer for at språkteknologien velger ord som inngår i en logisk sammenheng. Eksemplene over viser hvordan kontekst spiller inn på transkripsjonen. I ordrekker er tendensen av mesteparten eller alt blir transkribert riktig på bokmål. Når informantene snakker i avgrensede sekvenser og inkluderer ubestemte artikler foran substantiv, infinitivsmarkør og andre indikatorer som rammer inn språktrekkene, er det større sannsynlighet for at «Dikter» håndterer nordnorske talemålstrekk. Teknologien bak «Dikter» er talemålsdatabaser som er matet med språkdata (Digitaliseringsdirektoratet, u.d). Språkdataene består av både enkeltord og fraser som talegjenkjenningsverktøyet gjenkjenner. På den måten fremstår «Dikter» som et kontekstbasert verktøy, ut fra de etablerte databasene for språkmaterialet. «Dikter» transkriberer enkeltord med utgangspunkt i hvilken ordsekvens eller kontekst inputen blir presentert i.

For syntaks som avviker fra standardspråklig norsk ortografi, ser derimot ikke «Dikter» ut til å ta hensyn til konteksten. Funnene viser at «Dikter» ikke endrer på syntaks fra nordnorsk til bokmål (se tabell 16 i delkapittel 4.5 over). Der nordnorsk avviker fra standardspråklig syntaks, blir dette stående i output. *Ka du meina* blir for som nevnt til *hva du mener*, i stedet for normert ordstilling, *hva mener du*. Dette tyder på at talegjenkjenningsverktøyet i liten grad håndterer syntaks i transkriberingen. På en annen side kan det henge sammen med hvordan språkteknologiske verktøy er bygd opp. Talegjenkjenningsverktøy er matet med språkdata og jobber systematisk med input og output (Digitaliseringsdirektoratet, u.d.). Det språkbrukere dikterer «inn», kommer derfor «ut» i samme rekkefølge som det er sagt. «Dikter» behandler altså ikke syntaksen i det språkbrukerne sier uavhengig av hvilket talemål de har. Språkbrukere med østnorsk standardtalemål vil trolig få lik «output» i ordstillingen dersom de sa *hva du mener*, fordi verktøyet behandler ord for ord i dikteringen. Forskjellen handler først og fremst

om at nordnorske talere kan ha en syntaks som skiller seg fra østnorsk standardtalemål iblant annet spørresetninger. Dette gir feil syntaks etter standardspråklig norsk ortografi.

Bourdieu (1991:76) argumenterer for at språk har ulik verdi på det lingvistiske markedet. I dette ligger det at noen språk og varieteter har større verdi enn andre. Østnorsk standardtalemål har en høyere verdi på det språkteknologiske markedet. Funnene tyder på at nordnorsk talemål blir vurdert etter østnorsk standardtalemål i «Dikter». Der det oppstår feil i transkripsjonene, både innenfor de morfologiske og leksikalske feil, kan disse i stor grad knyttes til fonologiske trekk ved nordnorsk talemål. Dette betyr at verktøyet ikke har optimal brukervennlighet for språkbrukere med nordnorsk talemål. Ifølge Mæhlum (2009) ligger denne forhøyede statusen på det østnorske talemålet. Språkdataene som ligger til grunn for talegjenkjenningsverktøy domineres av østnorsk standardtalemål og varieteter som ligger nært dette. Nordnorsk talemål blir derfor vurdert med utgangspunkt i østnorsk, i stedet for å bli vurdert på talemålet egne premisser. Østnorsk standardtalemål utgjør med andre ord en sterk og dominerende norm i det norske språksamfunnet. Defineringsmakten for talemål i «Dikter» ligger altså på det østnorske standardtalemålet, fordi norsk talemål blir entydig med østnorsk standardspråk.

Med utgangspunkt i diskusjonen så langt er det mulig å hevde at det knytter seg et demokratisk problem til de språkteknologiske verktøyene. Empirien i denne undersøkelsen viser at talemålet til språkbrukere påvirker funksjonaliteten til «Dikter». Det østnorske standardtalemålet har hegemoniet på språkteknologien. Dette kan ekskludere store deler av den norske befolkning som snakker andre talemål enn den østnorske standardvarietetten i bruk av språkteknologiske hjelpemidler. Språkteknologiske verktøy, som «Dikter», kan gi muligheter og være en forutsetning for demokratisk deltakelse i samfunnet. Det kan være særlig nyttig for grupper som trenger støtte i hverdagen. For eksempel at blinde kan få lest opp tekst, eller at elever med lese- og skrivevansker kan diktere inn det de ønsker å skrive. Språkteknologiske verktøy kan være med å sikre disse gruppene samfunnsdeltakelse. I tillegg kan språkteknologiske verktøy være en støtte og hjelpe for alle i samfunnet. Derfor må «Dikter» og andre språkteknologiske verktøy fungere for alle folk, uavhengig av hvilket talemål de snakker.

5.4 Informantenes tilpasning

Opptakene gjort i denne studien viser at informantene med få unntak snakker autentisk lofotmål. Dette innebærer at informantene ikke tilpasser dialekten sin i merkbar grad mellom bruken av «Dikter» og intervjuet, men at de beholder lokale talemålstrekk. I den grad de gjør tilpasninger handler det i hovedsak om å gjenta ord eller si ordene tydeligere.

Mennesker har en tendens til å akkomodere sitt talemål til andre personer, grupper eller situasjoner (Akselberg et al., 2008:160). Det informantene eventuelt kan ha tilpasset talemålet sitt til, er forskeren i intervjuet. Men forskeren i denne studien og informantene snakker omtrent det samme talemålet, noe som altså danner et godt utgangspunkt for autentisk tale hos informantene. På en annen side gjelder ikke akkomodasjon bare tilpasninger av talemål i møte med andre mennesker og situasjoner. Kunnskap om akkomodasjon kan også overføres til språkbrukeres møte med digitale talegjenkjenningsverktøy. Det var forventet at den unaturlige konteksten – å snakke til en maskin – kunne få frem et mer formelt og standardisert språk. Informantene tilpasser ikke talemålet sitt til hverken forskeren eller datamaskinen. Det kan hevdes at det å snakke til en datamaskin kan bidra til å hindre de tradisjonelle mekanismene for akkomodasjon.

Det faktum at informantene beholder sin autentiske tale i ulike situasjoner kan henge sammen med deres identitet. Identitet betyr «den samme» og kan både handle om individets «sammehet» med seg selv og likhet med andre individer i et fellesskap (Akselberg et al., 2008:108). Denne undersøkelsen har undersøkt intraindividuell variasjon, altså språklig variasjon innad i individet (Norrby & Håkansson, 2015:60). Når informantene beholder sitt talemål er dette en form for divergering, ettersom at de beholder eksisterende språklige forskjeller og med dette markerer en distanse til samtalepartner og talegjenkjenningsverktøy. Informantene er valgt på bakgrunn av at de selv påstår at de snakker lofotmål. Felles for alle er at de kommer fra Lofoten og har bodd der de siste ti årene av sitt liv. Dette kan markere en sterk tilknytning til plassen og den lokale varieteten, lofotmål. En slik forståelse er i tråd med Akselberg (2008) sin tolkning av språkvalg og sosial identitet. Informantene tillegger seg selv en sosial identitet ved å kategorisere eget talemål som lofotmål. Informantenes identitet som lofotværing og brukere av lofottalemål kan være av betydning for at de beholder sitt autentiske talemål også i formelle situasjoner.

Det er et paradoks at de det digitale talegjenkjenningverktøyet virker å være regulert etter østnorsk standardtalemål, mens de menneskelige språkbrukerne i liten grad endrer sitt talemål til dikteringsverktøyet. Dette skiller seg fra generelle forventninger om at språkendringer og tilpasninger blir styrt av et overregionalt standardtalemål. Det innebærer at språkendringer på individ- og samfunnsnivå blir påvirket av et østnorsk standardtalemål (Mæhlum & Røyneland, 2012:134). Mæhlum & Røyneland (2012) viser til generelle språkendringer og utviklinger over lengre tid, altså diakron variasjon. Undersøkelsen min sammenligner bare talemålet brukt i to ulike kontekster synkront: intervju og selvbruk av

«Dikter». Talemålsendringer for Vestvågøy over tid er ikke en del av denne studien. Det er likevel interessant at informantene ikke tilpasser talemålet sitt til dikteringssituasjonen.

Ofte regnes regionale sentrum og det overregionale sentrumet Oslo som hovedsete for talemålspåvirkning i dag (Mæhlum & Røyneland, 2012:134). Vestvågøy er et ruralt område som ligger langt unna hovedstaden Oslo. Nærmeste storby er Bodø, som ligger over 100 km unna i luftavstand (Jahr & Skare, 1996:49). Språkbrukere i rurale språkområder kan derfor sies å være mindre utsatt for direkte språklig påvirkning enn de i urbane strøk. Informantene i denne undersøkelsen ser ut til å snakke autentisk lofottalemål, med utgangspunkt i de tradisjonelle talemålstrekkene. I tillegg handler tilpasning og bruk mye om vaner og kultur. Disse elevene hadde ikke brukt «Dikter» tidligere og hadde liten kjennskap til talegjenkjenningsverktøy generelt. Å bruke slike verktøy handler i stor grad om vane. Dersom elevene visste at kvaliteten ville blitt bedre om de tilpasset talemålet ville det vært interessant å sett om de ville tilpasset talemålet sitt.

5.5 Forslag til utvikling av talegjenkjenningsverktøy

For at talere med ulike norske dialekter skal kunne ta i bruk digitale talegjenkjenningsverktøy, må disse være tilpasset det mangfoldet av talemål som finnes. Eide & Våge (2018) skriver at det er en utfordring for utviklere av talegjenkjenningsverktøy å få tak i nok norske språkdata. Mangelen på norske språkdata begrenser potensialet som ligger i språkteknologien. En sentral utfordring for det norske språksamfunnet er at vi har stor dialektal variasjon. Transkriberte taledata som utviklere av talegjenkjenningsverktøy behøver, er kostbart å lage. Ettersom at norskspråk har stor dialektal variasjon, må det svært mye talespråksdata til for å få like god talegjenkjenning for alle dialekter. Norsk har et lite lingvistisk marked sammenlignet med for eksempel engelsk. Ofte vil engelsk bli prioritert i en større sammenheng i utviklingen av språkteknologiske verktøy. Derfor er det nærliggende å tro at kommersielle aktører som Microsoft, prioriterer å ha god talegjenkjenning norsk talemål som ser ut til å ligge nært det østnorske standardtalemålet.

I rapporten Språk i Norge – kultur og infrastruktur oppfordrer Språkrådet institusjoner til å dele språkdata som kreves for at språkteknologiske verktøy skal få de databasene som er nødvendige (2018). Utviklere av språkteknologien har behov for språkdata i både skrift fra tekster og transkripsjon av muntlig talespråk. Talegjenkjenningssystemer bygger på en maskinlæringsmodell som er trent på store mengder transkriberte taledata (Digitaliseringsdirektoratet, u.d.). Gjennom tilgang til varierte språkdata kan teknologien komme til nytte på de to norske skriftspråkene og for språkbrukere med ulike norske dialekter.

I tillegg må de to fagfeltene, talemålsforskning og språkteknologi, samarbeide for å få frem brukervennlige løsninger som er tilpasset mangfoldet av nordnorske talemål. Dette er en forutsetning for utviklingen av gode og inkluderende talegjenkjenningsverktøy som tar hensyn til de ulike norske dialektene.

Talegjenkjenningsverktøy som tar høyde for de dialektale variasjonene i norsk, sikrer norsk som samfunnsbærende språk også i den digitale fremtiden. En større satsning på norske talemål i utviklingen av talegjenkjenningsverktøy er avgjørende for brukerfunksjonaliteten til hver enkelt språkbruker, uavhengig av alder og talemål. Hovedformålet med slike verktøy er at de skal gjøre hverdagen vår lettere og sikre mulighet til å ta del i samfunnet. Verktøy som ikke håndterer dialektale variasjoner, vil gi feil og mangler som utfordrer brukerfunksjonaliteten. Dette kan føre til at språkbrukere lar være å benytte språkteknologiske verktøy.

6 Konklusjon

Denne studien har undersøkt hvordan talegjenkjenningsverktøyet «Dikter» håndterer nordnorsk dialekt, og hvordan nordnorske språkbrukere tilpasser talemålet sitt i møte med slike verktøy. Datamaterialet i undersøkelsen består av transkripsjoner utført av «Dikter», og dette er samlet inn ved hjelp av seks ungdommer fra Vestvågøy i Lofoten. Det skriftlige datamaterialet har blitt supplert med intervjuer for å sikre autentisk nordnorsk språkdata fra informantene. Empirien er analysert i lys av tradisjonell talemålsforskning og teori fra språkteknologi. I tillegg blir sosiolingvistisk teori om språk og identitet brukt for å drøfte informantenes tilpasninger når de bruker talegjenkjenningsverktøy.

Resultatene i undersøkelsen viser at «Dikter» ikke håndterer nordnorsk talemål på en tilfredsstillende måte. Transkripsjonene inneholder en rekke feil og mangler som særlig er knyttet til det fonologiske systemet for nordnorsk talemål. «Dikter» har vanskeligheter med å transkribere ord med palatalitet, apokope og senking. Talegjenkjenningsverktøyet har vanskeligheter med å kjenne igjen det morfologiske systemet til nordnorsk. Når det gjelder syntaks som skiller seg fra normert V2-ordstilling, slik som i nordnorske spørresetninger, beholdes syntaksen akkurat slik informantene sier det. Etterstilte possessiver i nordnorsk talemål blir enten transkribert feilaktig eller utelatt. Et språktrekk som gjennomgående blir transkribert korrekt, er retroflektering. For spørreord med k-fremlyd og de personlig pronomener eg og e, er transkripsjonene stort sett alltid korrekte. Ord og språktrekk (retroflektering) som er like mellom nordnorsk talemål og østnorsk blir som hovedregel transkribert korrekt. Vi kan imidlertid fastslå at faktorer som hvilken kontekst den nordnorske inputen kommer i spiller inn på kvaliteten av transkripsjonene. Infinitivsmarkøren å og ubestemt artikkel foran substantiv kan for eksempel hjelpe verktøyet til å gjenkjenne ordene. Det samme gjelder for høyfrekvente ord, slik som personlig pronomener 1. person entall og spørreord. Til tross for at disse kategoriene skiller seg fra østnorsk standardtalemål blir de i stor grad transkribert korrekt. Andre faktorer som taletempo og informantenes diksjon ved bruk av «Dikter» kan ha påvirket resultatet i denne undersøkelsen. Funnene bekrefter derfor langt på vei studiens hypotese; om at talemålsverktøy er for lite tilpasset mangfoldet av norske talemål, med vekt på nordnorsk talemål.

Ut fra resultatet i denne undersøkelsen tyder mye på at østnorsk standardtalemål dominerer i talegjenkjenningsverktøyet. Dette betyr at den nordnorske inputen blir vurdert med utgangspunkt i østnorsk. Databasene som ligger til grunn for «Dikter» er matet med østnorske språkdata, og mangler derfor en del nordnorske talemålstrekk. På den måten får verktøyet dårligere fungering for enkelte talemål, slik som nordnorsk. Dette strider mot oppfatningen av

Norge som et dialektparadis, hvor alle dialekter er akseptert og brukes i både formelle og uformelle sammenhenger. Det østnorske standardtalemålet får implisitt en forhøyet status og makt over de andre talemålene i Norge.

Når det gjelder informantenes tilpasning av eget talemål, er det lite i denne studien som tyder på at informantene tilpasser talemålet sitt i bruk av talegjenkjenningsverktøy. Ingen av informantene har noen konsekvente endringer av talemål mellom intervjuet og selvbruk av «Dikter». Med andre ord har ikke dikteringssituasjonen direkte påvirkning for informantens språkvalg. De snakker autentisk lofotmål uavhengig av hvilken situasjon de befinner seg i. Informantene virker å være mindre opptatt av hvordan talegjenkjenningsverktøyet håndterer deres talemål. Til tross for at «Dikter» ser ut til å være bedre egnet for østnorsk standardtalemål, påvirker denne varieteten informantenes talemål i liten grad.

Det er ingen tvil om at digitaliseringen og språkteknologien har ført med seg oppfinnelser som gjør vår hverdag enklere. Dersom talegjenkjenningsverktøy ikke fungerer optimalt kan det føre til at folk velger disse bort eller tilpasser språket når de skal bruke talegjenkjenningsverktøy. På den måten knytter det seg et demokratisk problem til talegjenkjenningsverktøy som «Dikter». Derfor er det viktig at det blir satset på utviklingen av språkteknologiske verktøy som er tilpasset ulike dialekter.

Datamaterialet i denne undersøkelsen er ikke tilstrekkelig til å si noe generelt om «Dikter» sin håndtering av nordnorsk talemål. På en annen side kan funnene generere hypoteser som kan undersøkes i større kvantitative forskningsprosjekter med representative utvalg. Først og fremst har denne studien bare undersøkt et spesifikt talegjenkjenningsverktøy, «Dikter». Det betyr at andre verktøy kan ha annen håndtering av nordnorske talemål. For det andre kan studien peke på noen tendenser som kan være gyldig for andre nordnorske dialekter og andre talemål i talegjenkjenningsverktøy. Selv om informantene er fra Vestvågøy, er flere av talemålstrekkene også gyldig i andre nordnorske talemål. For det tredje er nærliggende å tro at talegjenkjenningsverktøyenes håndtering av andre talemålstrekk heller ikke er uproblematisk.

Fra et didaktisk perspektiv er det relevant å vite noe om talegjenkjenningsverktøy sin håndtering av nordnorske talemål. Dette kan gi læreren nyttig informasjon om hvordan verktøyene fungerer for elever med ulike talemål. På den måten kan lærere tilrettelegge for god bruk av språkteknologiske verktøy i skolesammenheng. Talegjenkjenningsverktøy med god fungering kan blant annet forenkle skriveprosesser for elever. Dette kan gi støtte for alle elever i skolen, uavhengig av hvilket talemål de snakker. I tillegg kan feilaktige transkripsjoner utført av talegjenkjenningsverktøy brukes som læring om den norske dialektsituasjonen. Som

fremtidig lærer ser jeg frem til å teste flere talegjenkjenningsverktøy sammen med elever med forskjellige norske talemål.

6.1 Videre forskning

Som nevnt innledningsvis er det mye upløyd mark i krysningen mellom språkteknologi og talemålsforskning. I denne studien har jeg undersøkt hvordan talegjenkjenningsverktøyet «Dikter» håndterer et nordnorsk talemål. Undersøkelsene i denne studien peker på noen tendenser som kan være gjeldende for andre talemålene og andre språkteknologiske verktøy. Det er grunn til å tro at man vil finne noe av det samme dersom man undersøker hvordan «Dikter» håndterer andre talemål. Derfor vil være interessant å undersøke hvordan talegjenkjenningsverktøyet håndterer andre norske talemål, for å sammenligne funksjonaliteten til verktøyet med ulike talemål. Særlig interessant vil det være å undersøke østnorsk, som ut fra denne studien ser ut til å være normgivende for språkteknologiske verktøy. En mulig studie kan sammenligne funksjonaliteten til «Dikter» på nordnorsk og østnorsk talemål. I tillegg kan videre forskning se på flere talegjenkjenningsverktøy for å undersøke hvorvidt tendensene i denne studien kan sies å være aktuelle.

Referanseliste

- Askedal, J. O. (2009). Tysk språknormering og standarduttale i historisk perspektiv. *Norsk Lingvistisk Tidsskrift*, 27, s. 199 – 217.
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge: Polity Press.
- Brinkmann, S. & Kvale, S. (2015). *Det kvalitative forskningsintervju* (3. utg). Oslo: Gyldendal akademisk.
- Bugge, E. (2020). Sosiolingvistisk undersøkelse av talemålsvariasjon. I L. Aa & R. Neteland (Red.), *Mater i norsk. Metodeboka 2* (s. 216 - 231). Universitetsforlaget: Oslo.
- Chafe, W. (1980). *The pear stories. Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. New Jersey: Ablex Publishing Corporation.
- Christiansen, H. (1948). *Norske dialekter 3: De viktigste målmerker og deres råderom*. Oslo: Tanum.
- Christiansen, H. (1996[1950]). Geografisk oversyn over den nordnorske apokope. I E. H. Jahr & O. Skare (red.). *Nordnorske dialekter* (s. 93 – 99). Oslo: Universitetsforlaget.
- Cornips, L. & Poletto, C. (2005). On Standardising syntactic elicitation techniques (part 1). *Lingua*, 115, s. 939 – 957.
- Daidsen, C. (2013). Legenavn og talegjenkjenning. *Tidsskriftet den norske legeforening*, 113, s. 766.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. & Walters, J. (2012). Multilingual Assessment Instrument for Narratives (MAIN). *ZAS Papers in Linguistics* 56. Berlin: Zentrum für Allgemeine Sprachwissenschaft.
- NESH Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (2021). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora*. Hentet fra NESH <https://www.forskningsetikk.no/retningslinjer/hum-sam/forskningsetiske-retningslinjer-for-samfunnsvitenskap-og-humaniora/> lastet ned 10.01.2022

- Digitaliseringsdirektoratet (u.d). Språkdata – korleis kan vi hauste og dele? Hentet fra Digdir <https://www.digdir.no/datadeling/sprakdata-korleis-kan-vi-hauste-og-dele/2367> Lastet ned 12.03.22
- Eckert, P. (1989). *Jocks and Burnouts. Sosial Categories and Identity in the High School*. New York: Teachers College Press.
- Eckert, P. (2000). *Linguistic Variation as Social Practice. The Linguistic Construction of Identity in Belten High*. Oxford: Blackwell.
- Eckhoff, K. K. (2018). *Språk i Norge – kultur og infrastruktur*. Oslo: Språkrådet. Hentet fra Språkrådet https://www.sprakradet.no/globalassets/diverse/sprak-i-norge_web.pdf lastet ned 30.09.21
- Eide, K. & Våge, O. (2018). Sanker språkstemmer. *Språknytt*, 4, s. 10 – 12.
- Evjen, L. R. (2011). *æ sitt med klump i halsen når æ skriv det hær : dialekt i skriftspråket i debattforum knytt til tre norske nettaviser*. Tromsø: UiT Norges arktiske universitet.
- Folkvord, M. S. (2020). *Normer i et ikke-normert skriftspråk – en studie av talemålsnær skrivning på Facebook Messenger blant trondheimere*. Bergen: Universitetet i Bergen.
- Healthcare (2018). 62 Percent of Docs Use Speech Recognition, But Cite Concerns About Accuracy. Hentet fra <https://www.hcinnovationgroup.com/clinical-it/news/13030744/survey-62-percent-of-docs-use-speech-recognition-but-cite-concerns-about-accuracy> lastet ned 10. januar 2022
- Hårstad, S. (2020). Språk på nettet – tilnæringsmåtar til digital språkbruk. I L. Aa & R. Neteland (Red.), *Mater i norsk. Metodeboka 1* (s. 140 – 157). Oslo: Universitetsforlaget.
- Hårstad, S. & Opsahl, T. (2013). *Språk i byen. Utviklingslinjer i urbane språkmiljøer i Norge*. Oslo: Fagbokforlaget.
- Hårstad, S., Mæhlum, B. & Van Ommeren, R. (2021). *Blikk for språk. Sosiokulturelle perspektiver på norsk språkvirkelighet*. Oslo: Cappelen Damm akademisk.
- Jahr, E. H. & Skare, O. (1996). *Nordnorske dialekter*. Oslo: Novus forlag.

- Johannessen J. B., Priestley J., Hagen K., Åfarli T. A. & Vangsnes, Ø. A. (2009a). The Nordic Dialect Corpus - an Advanced Research Tool. I K. Jokinen & B. Eckhard (Red.): *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009. NEALT Proceedings Series*, Volume 4. Hentet fra Nordisk dialektkorpus <http://tekstlab.uio.no/nota/scandiasyn/>
- Johannsen J. B., Priestley J., Hagen K., Åfarli T. A. & Vangsnes, Ø. A. (2009b). *Transkripsjonsretteiing for ScanDiaSyn*. Hentet fra Nordisk dialektkorpus <http://www.tekstlab.uio.no/scandiasyn/Transkripsjonsretteiing%20for%20ScanDiaSyn.pdf> lastet ned 02.09.21
- Kraft, S., Thurfjell, F., Rack, J. & Wengelin, Å. (2019). Lexikala analyser av muntlig, tangentbordsskriven och dikterad text producerad av barn med stavningssvårigheter. *Nordic journal of literacy research*, Vol.5 (3), s. 102 – 122.
- Lust, B. and Blume, M. (2017a). *Experimental Tasks for Generating Language Production Data*. Berlin/Munchen/Boston: De Gruyter. Hentet fra <https://doi.org/10.1515/9783110415339-007> lastet ned 15.02.2022
- Lust, B. and Blume, M. (2017b). *Introduction to Experimental Methods: Design and Analysis*. Berlin/Munchen/Boston: De Gruyter. Hentet fra <https://doi.org/10.1515/9783110415339-006> lastet ned 15.02.2022
- Microsoft. (u.d.). Diktere dokumentene dine i word. Hentet fra Microsoft [418f-b8ab-db7ce0d11d3c](#) lastet ned 12.12.21
- Mæhlum, B. & Røynealand, U. (2012). *Det norske dialektlandskapet. Innføring i studiet av dialekter*. Cappelen Damm akademisk: Oslo.
- Mæhlum, B. (2009). Standardtalemål? Naturligvis! *Norsk Lingvistisk Tidsskrift*, 27, s. 7 – 26.
- Mæhlum, B., Akselberg, G., Røynealand, U. & Sandøy, H. (2008). *Språkmøte. Innføring i sosiolingvistikk*. Oslo: Cappelen Damm akademisk.
- Nasjonalbiblioteket (u.d.). Språkbanken. Hentet fra <https://www.nb.no/sprakbanken/> Lastet ned 20.01.22

- Neteland, R. & Aa, L. I. (2020). Å designe og gjennomføre eit forskingsprosjekt. I L. Aa & R. Neteland (Red.), *Mater i norsk. Metodeboka 2* (s. 12 – 24). Universitetsforlaget: Oslo.
- Neteland, R. (2020). Kvalitative intervju i norskfaglige oppgaver. I L. Aa & R. Neteland (Red.), *Mater i norsk. Metodeboka 2* (s. 50 – 67). Universitetsforlaget: Oslo.
- Norrby, C. & Håkansson, G. (2015). *Introduktion til sociolingvistik*. Lund: Studentlitteratur.
- Rosén, V. (2014): Språkteknologiens behov for leksikalsk informasjon. I *Nordiske studier i leksikografi*, nr 12, s. 13 – 41.
- Selås, M. (2020). Å undersøke munnleg språkutvikling i skulealder. I R. Neteland & L. I. Aa (Red.), *Master i norsk. Metodeboka 2* (s. 198 – 215). Oslo: Universitetsforlaget.
- Vangsnes, Ø. & Westergaard, M. (2019). *Ka LIA fortæll? Eit gjensyn med kv-spørsmål i norske dialektar* (s. 155 – 178). Oslo: Novus Forlag.

Vedlegg 1 NSD

[Meldeskjema](#) / [Nordnorsk dialekt i talegjenkjenningsverktøy](#) / [Vurdering](#)

Vurdering

☰ 29.10.2021 ▾

🖨 Skriv ut

Referansenummer

537698

Prosjekttittel

Nordnorsk dialekt i talegjenkjenningsverktøy

Behandlingsansvarlig institusjon

UiT Norges Arktiske Universitet / Fakultet for humaniora, samfunnsvitenskap og lærerutdanning / Institutt for språk og kultur

Prosjektperiode

11.10.2021 - 16.06.2022

[Meldeskjema](#) [↗](#)

Dato

29.10.2021

Type

Standard

Kommentar

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 29.10.2021. Behandlingen kan starte.

DEL PROSJEKTET MED PROSJEKTANSVARLIG

Det er obligatorisk for studenter å dele meldeskjemaet med prosjektansvarlig (veileder). Det gjøres ved å trykke på "Del prosjekt" i meldeskjemaet. Om prosjektansvarlig ikke svarer på invitasjonen innen en uke må han/hun inviteres på nytt.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 16.6.2021.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen

formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål

dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet

lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18) og dataportabilitet (art. 20).

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og eventuelt rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: <https://www.nsd.no/personverntjenester/fyll-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema> Du må vente på svar fra NSD før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Kontaktperson hos NSD: Lisa Lie Bjordal

Lykke til med prosjektet!

Vedlegg 2 Samtykkeerklæring

Forespørsel om deltakelse i masteroppgaveprosjektet «nordnorsk talemål i talegjenkjenningsverktøy»

I forbindelse med mitt masterprosjekt i nordisk språk ved UiT Norges arktiske universitet gjennomfører jeg en undersøkelse om hvordan digitale talegjenkjenningsverktøy håndterer nordnorske dialekter. Formålet med prosjektet er å avdekke hvilke utfordringer som dukker opp når personer med nordnorsk dialekt skal bruke talegjenkjenningsverktøyet Office «Dikter».

Hvorfor får du spørsmål om å delta?

Jeg vil gjerne rekruttere deg som informant til prosjektet ettersom at du er ungdom med nordnorsk dialekt. Det er totalt seks informanter i prosjektet.

Hva innebærer det for deg å delta?

Hvis du velger å delta i dette prosjektet, innebærer det at du må delta i et kort intervju og selv bruke talegjenkjenningsverktøyet «Dikter». Det kreves ingen forkunnskaper for å bruke verktøyet, og nødvendig opplæring vil bli gitt. Totalt vil det ta ca 30 minutter å delta i prosjektet, men jeg anbefaler at du setter av minst 45 minutter slik at vi har god tid. Jeg tar lydopptak og notater fra intervjuet. I tillegg vil verktøyet «Dikter» gjøre om tale til tekst, og denne informasjonen blir registrert elektronisk.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern

Opplysningene om deg vil kun bli brukt til formålene jeg har fortalt om i dette skrevet. Lydopptakene og transkripsjonene fra «Dikter» vil bli behandlet konfidensielt og i samsvar med personvernregelverket. Det er kun veilederne mine og jeg som vil ha tilgang til lydopptakene mens prosjektet pågår. Formidlingen av resultatene fra prosjektet vil være tilgjengelig i form av en masteroppgave. All informasjon om deltakerne i prosjektet vil bli anonymisert i oppgaven. Det vil si at du vil bli omtalt som «informant» eller med et fiktivt navn. Informasjon som det er nødvendig å oppgi, vil bli gjort mer generell. Opplysninger om alder og bosted vil for eksempel bli oppgitt som generelle kategorier (elev på vg1, vg2, vg3) og region/dialektområde. Du vil med andre ord ikke kunne identifiseres i oppgaven.

Hva skjer med opplysningene dine når prosjektet avsluttes?

Opplysningene anonymiseres når prosjektet avsluttes/oppgaven er godkjent, noe som etter planen er juni 2022. Når prosjektet avsluttes, vil lydopptakene og øvrige personopplysninger bli slettet.

Hvem er ansvarlig for prosjektet?

Ansvarlig for prosjektet er min hovedveileder Øystein Vangsnes, professor i nordisk språkvitenskap ved UiT. Biveileder er professor Trond Trosterud ved Senter for samisk språkteknologi, UiT. Dersom du har spørsmål til prosjektet, kan du ta kontakt med veilederne (oystein.vangsnes@uit.no / trond.trosterud@uit.no) eller undertegnede.

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 53 21 15 00.

Dersom du ønsker å delta i undersøkelsen, kan du signere den vedlagte samtykkeerklæringen og returnere den tilbake til din lærer X så snart som mulig.

Med vennlig hilsen,
Maiken Katarina Salamonsen
Lektorutdanning 8-13 i nordisk
UiT Norges arktiske Universitet
msa185@uit.no

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «nordnorsk talemål i talegjenkjenningsverktøy», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- Å delta i intervju med lydopptak av meg
- Å bruke talegjenkjenningsverktøyet «Dikter»
- At opplysninger fra intervjuet og transkripsjonen fra «Dikter» kan bli brukt i oppgaven
- At mine opplysninger behandles frem til prosjektet er avsluttet

Sted og dato

Signatur

Vedlegg 3 Intervjugaid

Semi-strukturert intervjugaid

Generelt:

1. Hvor er du født og oppvokst?
2. Hvordan var det å vokse opp i/på (navn på sted)?
3. Hva er det første du kan huske fra barndommen?
4. Hva liker du å gjøre på fritiden?

Skole:

5. Hvilket fag liker du best, og hvorfor?
6. Av alt det du lærer på skolen, hva tror du at du vil få mest bruk for senere i livet?
7. Hva pleier du å gjøre i friminuttene?
8. Bruker skolen digitale verktøy, i så fall hvilke?
9. Dersom skolen bruker digitale verktøy, får dere opplæring i disse?
10. Hva synes du om å bruke digitale verktøy som en del av undervisningen?

Hjemlass:

11. Hvordan vil du karakterisere språket der du vokste opp?
12. Hva er det som ikke er så bra ved det stedet der du bor?
13. Om du skulle forsøke å lokke noen til det stedet der du bor, hva ville du si da?
14. Avslutningsvis, er det noe i intervjuet du ønsker å oppklare eller legge til?

