



Anonym arkivar og anonyme arkivskuffer på eit udatert bilete utan kjend fotograf. Motivet er derimot klart: eit typisk arkivsystem før den store teknologiske omveltinga starta. Foto: Universitetsbiblioteket i Bergen, Billedsamlingen/Språksamlingene.

Kapittel 8

Teknologi og utvikling av språkressursar

*Gjert Kristoffersen*¹

Dette kapitlet handler om korleis ny teknologi har påverka språkforskinga ved Universitetet i Bergen frå 1946 og fram til i dag. Målet med all forskning er å oppnå ei betre forståing av røynda. Til det treng ein data – observasjonar som kan systematiserast og analyserast ut i ny kunnskap. Jo fleire og meir systematiske observasjonar ein har, jo større er sjansane for at resultatet av forskingsprosessen blir av varig verdi.

¹ Gjert Kristoffersen har skrive dette kapitlet. Han døydde den 29. mai 2021, før det var heilt ferdig. Johan Myking har derfor lagt til avsnittet om stadnamn og gjort nokre andre tilføyingar.

Ei sentral kjelde for språkvitskap og filologi er tekstar, så vel skriftlege som munnlege. Dei eldste skriftlege kjeldene er hogd i stein, i vårt land i form av runer. Med kristendomen kom det latinske alfabetet og handskrift på pergament. Dette ble seinare erstatta med papir, og om lag samstundes gjorde boktrykkekunsten at tekstar kunne mangfaldiggjerast i ein grad som ikkje var mogleg før. Med fotografiet blei det også mogleg å mangfaldiggjere kopiar av ikkje-trykte tekster, som norrøne manuskript og runesteinar. Dette var i grove trekk situasjonen då Universitetet i Bergen blei oppretta i 1946.

Fram til opptaksteknologien blei tatt i bruk for alvor på 1950-talet, var skrift det einaste mediet, også når det galdt å gje att tale. Om ein leitte etter eit bestemt fenomen, t.d. frekvensen av eit visst ord eller ein ordkombinasjon i ein skriftleg tekst, eller eit bestemt lydmonster i eit taleopptak, måtte ein lese seg gjennom heile teksten eller lytte gjennom opptaket ord for ord. Dette la sterke band på kor store mengder data ein kunne analysere, og arbeidet kravde stort tolmod og intens konsentrasjon.

Dette endra seg frå 1960-talet av, då dei første digitaliserte tekstane kom. Eit pionerprosjekt var det såkalla Brown-korpuset frå 1964 – eit elektronisk korpus sett saman av moderne amerikansk-engelske prosatekstar henta frå ulike sjangrar.² Det tilsvarande britisk-engelske korpuset, som blei utvikla på 1970-talet, ber namnet The Lancaster–Oslo–Bergen Corpus. Samarbeidspartnaren i Bergen var likevel ikkje fagmiljø ved universitetet. Det var Norges Almennvitenskapelige forskningsråd (NAVF) sitt EDB-senter for humanistiske fag, som vart plassert i Bergen i 1972. Det samarbeidde med professor Stig Johansson ved UiO og kollegaar i England. Dette samarbeidet førte likevel til at kompetanse i utvikling og utnytting av elektroniske tekstar blei utvikla tidleg i Bergen – ein kompetanse som også tilsette ved UiB skulle få glede og nytte av i åra som følgde.

Tilgang til digitaliserte tekstar revolusjonerte vitskapleg analyse av skriftmateriale i ulike former. Ein tilsvarande tilgang til talemålsopptak var teknisk sett meir utfordrande, og kom ein del år seinare. Digitalisert lyd krev stor lagringskapasitet, noko som lenge var eit problem. Eit anna var at medan orda i skriftlege tekstar relativt lett kan merkast automatisk for kva ordklasse dei

2 McEnery & Hardie 2013: 2.

høyrrer til og kva bøyingsform dei har, er korkje identifikasjon av ord eller vidare grammatisk analyse lett å gjere direkte i eit lydopptak.

Talegjenkjenningsteknologi er enno på langt nær god nok til at den kan gjere ein tilsvarande analyse uavhengig av språk, dialekt og teknisk kvalitet på opptaket. Løysinga har i staden vore manuell transkripsjon av talen, normalt til standardortografi. I tillegg til at transkripsjonen då kan behandlast vidare som all annan skriftleg tekst, kan ein parallellkople transkripsjon og lyd slik at transkripsjonen kan brukast for søking i lyden.

Kva er ein språkressurs?

Språkressursomgrepet rommar meir enn tekst- og talekorpus av den typen eg nettopp har skildra. Sjølve korpus-omgrepet går attende til 1960-talet og gjennombrøtet for korpuslingvistikken. Før det hadde ein arkiv, eller samlingar, til dømes Folkeminnesamlinga som blei grunnlagt ved Bergens Museum i 1921. Innhaldet her var meir heterogent, og fram til bandopptakaren blei teken i bruk på 1950-talet, var alt på papir – feltnotat, ordsetlar, svar på standardiserte spørjelister, kart og meir.

Desse arkiva var i utgangspunktet tenkte som forskingsressursar, dvs. samlingar av forskingsmateriale som samtidige og seinare generasjonar kunne dra fordel av i eiga forskning. Eit spørsmål vi skal stille nedanfor, er i kva grad ikkje berre dei gamle arkiva, men også nyare, digitale ressursar verkeleg blei og blir brukte til vidare forskning. Som vi skal sjå, er det ikkje alltid at dei forventningane ein legg til grunn i utgangspunktet, blir innfridde av ettertida.

Samstundes må vi vere klar over at verdien av desse samlingane ikkje berre kan knytast til seinare forskning. Dei utgjer også ein viktig dokumentasjon av tidlegare språksteg og av språktrekk som i dag kanskje har blitt borte. På same måte som det er viktig å dokumentere språk som er i ferd med å døy, er det viktig å dokumentere dialekttrekk og i verste fall dialektar som står i fare for å forsvinne.³

3 Eit eksempel på det siste er strilemåla som tradisjonelt blei talte i landkommunane rundt Bergen. I dag er bergensmålet mange stader i ferd med å fortrenge desse. Men *Målføresamlinga* inneheld materiale som gjer at kunnskapen om desse dialektane likevel er sikra for ettertida.

Ressursar kan òg ha praktiske føremål ut over forskning og dokumentasjon. Døme er ordsamlingar og tekstkorpus. I tillegg til at dei vil utgjere forskingsmateriale for undersøkingar av mellom anna ordforrådet i språket og endringar i dette, utgjer dei også råmateriale for ordbøker av ymse slag.

Nytten av eit korpus avheng av i alle fall to ting. Det må for det første vere annotert, utstyrt med grammatisk informasjon om kvart ord, som til dømes ordklasse og bøyingsform. Slik informasjon kan ein langt på veg legge inn ved hjelp av såkalla ordklassetagging, ein programvare som med høg grad av presisjon forsyner kvart ord med denne informasjonen med utgangspunkt i leksikografiske basar. Ein må også knyte bakgrunnsinformasjon til elementa i eit korpus, såkalla metadata, som til dømes kvar ein tekst kjem frå, kven som har skrive han, eller kven det er gjort opptak av, kva sjanger teksten representerer, og kor lang han er. Enda meir avansert og utfordrande er det å skrive program som kan analysere *setningane* i korpuset, dvs. analysere dei syntaktisk. Setningsanalyse blir vanlegvis visualisert ved hjelp av såkalla syntaktiske trær, og eit korpus der setningane er analyserte, blir kalla ein trebank.

I tillegg treng ein eit best mogleg søkeverktøy, slik at ein med utgangspunkt i annotering og metadata kan hente ut den informasjonen ein treng frå korpuset.

Teoretisk bakteppe

Forskning blir ikkje berre styrt av observasjonar og kvaliteten og omfanget av desse. Analyse tar normalt utgangspunkt i hypotesar, og kva for hypotesar ein legg til grunn, er i stor grad forankra i den eller dei rådande, overordna teoriene knytte til det fagfeltet ein arbeider innanfor.

Språkvitskapen går attende til tidleg 1800-tall, då oppdaginga av systematiske korrespondansar mellom bestemte språk hadde ført til ei erkjenning av at språka var i slekt med kvarandre og at dei følgeleg kunne først attende til eit felles opphavsspråk meir eller mindre langt tilbake i historia.⁴ Dette la grunnlaget for

4 Ei kort innføring i språkvitskapens historie fram mot slutten av førre hundreåret er Robins 1997.

den historiske språkvitskapen. Denne blei ein viktig støttevitskap for filologien, i Noreg først og fremst studiet av den norrøne mellomalderkulturen. Det vitskaplege studiet av norrønt språk fekk sitt gjennombrøt på 1800-talet, og studiet av folkemåla rundt om i landet blei ei viktig vidareutvikling av dette feltet, av di det var frå dei, og ikkje frå det dansk-norske skriftspråket, ein kunne føre linene direkte tilbake til norrønt. Dette var sentralt då Ivar Aasen midt på 1800-talet utvikla landsmålet med utgangspunkt i ei rekke studiereiser der han hadde samla og systematisert dialektstoff frå ulike delar av landet, særleg Vestlandet.⁵ Medan Aasen var ein viktig forgjengar, var det Johan Storm, professor i romansk og engelsk filologi ved Universitetet i Kristiania, som la grunnen for norsk dialektologi som vitskapsgrein gjennom boka *Norsk Lydskrift med Omrids av Fonetiken* frå 1884.⁶ I sekelet som følgde, kom det ei mengd studiar om norske dialekter bygd på prinsippa til den historiske språkvitskapen og Storms program.

I 1916 blei Ferdinand de Saussures *Cours de linguistique générale* utgitt.⁷ Her blei det primære undersøkingsobjektet i språkvitskapen omdefinert frå korleis språk kan relaterast gjennom historisk samanlikning og rekonstruksjon (diakroni), til korleis språk framstår som eit system bygd på kontrastar på eit gitt tidspunkt, utan omsyn til den historiske utviklinga som ligg bak dette systemet (synkroni). Den nye retninga fekk namnet *strukturalisme*. Eit grunnleggande skilje var dét mellom 'langue' og 'parole', som til norsk kan omsetjast til 'språkssystem' og 'språkbruk'. Systemet er det sentrale, og dette er sosialt konstituert, sidan ingen språkbruker kan seiast å ha tileigna seg det fullt ut. Det er samstundes også dette sosialt konstituerte systemet som gjer språkbruk og kommunikasjon mogleg. For strukturalistane var det *kontrast* mellom einingane i språkssystemet som var sentralt; det var gjennom ulik form ein kunne uttrykke ulik mening. Der den historiske språkvitskapen til dømes var interessert i korleis kvar enkelt språklyd blei uttalt og korleis denne endra seg over tid, var det skilnaden mellom språklydane og korleis dei saman utgjorde eit system av kontrastar som var det sentrale for strukturalismen. Medan interessa for språkhistoriske emne ikkje

5 Sjå t.d. Aasen 1965 [1864].

6 Storm 1884.

7 Saussure 1916.

forsvann, blei det strukturalistiske perspektivet raskt det dominerande innanfor internasjonal språkvitskap, og på 1930-talet var det allmenningvistiske miljøet ved Universitetet i Oslo prega av dei nye ideane. Men til Bergen kom dei, som vi skal sjå, mykje seinare.

Den tredje og siste revolusjonen skjedde mot slutten av 1950-talet. Då publiserte den amerikanske lingvisten Noam Chomsky for første gong sitt generative program i boka *Syntactic Structures*, og eit par år seinare skreiv han ein knusande kritikk i form av ei bokmelding av den amerikanske utgåva av strukturalismen og hennar tette samband med behaviorismen.⁸ Chomsky heldt fast ved det synkrone perspektivet, og på skiljet mellom system og bruk, men omdefinerte det primære forskingsobjektet for språkvitskapen til den individuelle språkevna, ikkje det sosialt konstituerte systemet. Chomskys overordna hypotese var at born si evne til å lære språk kan sjåast som eit abstrakt, medfødt organ som i samspel med innputt frå omverda gjer at alle barn i løpet av kort tid lærer seg å tale det språket eller dei språka som er i bruk rundt dei. Dette skal forklare kvifor alle barn i løpet av nokre få år lærer eit så komplisert system som ein mental grammatikk er, utan trong for målretta undervisning.

Eit viktig poeng i vår samanheng er at Chomsky meinte at innsamla språkdata, t.d. i form av korpus, var meir eller mindre utan verdi. Slike data representerte bruk, og språkbruk var eit resultat ikkje berre av den mentale grammatikken, men også prega av feil som kunne skyldast manglande konsentrasjon, begrensa korttidsminne og andre faktorar. I tillegg var ikkje det at døme på ein bestemt konstruksjon ikkje fanst i eit korpus, eit bevis på at den var ugrammatisk; den kunne berre vere sær seldan brukt. Chomskyanarar såg difor lita nytte i korpus. I staden baserte dei seg på intuisjon, både sin eigen og andre sine gjennom såkalla akseptabilitetstestar, der dei bad andre vurdere om ei gitt setning var akseptabel (= grammatisk) eller ikkje.

Sjølv om Chomskys teori fekk breitt gjennomslag i løpet av 1960-talet, blei han ikkje normalvitskap på same måte som strukturalismen blei det. Både tesen om at det er den mentale språkevna som er det primære forskingsobjektet, og den at forskning basert på

8 Chomsky 1957, 1959. Kritikken hans ramma ikkje i same grad den europeiske tradisjonen.

språkbruksdata ikkje kan gje innsikt, møtte tidleg motstand, og førte til fleire retningar innanfor språkvitskapen og innanfor den generative grammatikken i meir eller mindre direkte opposisjon til Chomsky. Som vi skal sjå, blei språkforskinga ved UiB i liten grad direkte prega av Chomskys idear.

Oversyn

I resten av dette kapittelet skal eg presentere nokre av dei ressursane som er blitt bygde ved Det historisk-filosofiske fakultet, frå 2007 Det humanistiske fakultet. Ressursane kan samlast under fire faglege paraplyar – talemålsgransking, stadnamn, leksikografi og datalingvistikk. Talemålsgranskinga går attende til grunnleggjaren av Bergens Museum, Wilhelm Frimann Koren Christie (1778–1849), som var ein ivrig ordsamlar.⁹ Ho blei tatt opp att då museet i 1918 fekk eit professorat i vestnorsk målgransking, og står framleis svært sterkt. Leksikografien hadde ei kortare blømingstid, frå midten av 1960-talet til slutten av 1990-talet, før den kom tilbake i stor skala mellom 2016 og 2019. Innanfor allmenn språkvitskap har ein sidan midten på 1980-talet bygd opp eit sterkt datalingvistisk fagmiljø som har resultert i ei rekke viktige digitale ressursar, særleg knytte til syntaks og semantikk.

Sidan dei fleste i vitskaplege stillingar ikkje hadde den teknologiske kompetansen som i tillegg til den faglege var naudsynt for bygging av ulike typar korpus, blir dei to viktigaste leverandørane av denne kompetansen trekte inn. *Program for datamaskinell språkbehandling* (PDS) var ei avdeling ved Nordisk institutt i om lag 20 år frå 1967, leia av Kolbjørn Heggstad (1934–). Heggstad var språkvitar og datakyndig, men det var også tilsett programmerarar ved avdelinga. Moglegheitene knytte til den nye teknologien skildra Heggstad slik i 1968:¹⁰

Eigentleg vart desse maskinane kalla reknemaskinar, men no dekkjar ikkje dette namnet lenger. Dei er i dag i stand til å ta mot store mengder språklege tekster, sortere desse,

9 Jordåen 2014: 419–424, Indrebø 1937.

10 Heggstad 1968: 25.

gjere samanlikningar i materialet, systematisere det og lage utskrifter med ein utroleg fart.

Langt viktigare blei likevel *NAVFs EDB-senter for humanistisk forskning*, det nasjonale kompetansesenteret som Rådet for humanistisk forskning hadde oppretta. Her skulle mange fagmiljø, ikkje berre språkvitarane, få hjelp med innføring av digitale metodar og hjelpemiddel. Etter at NAVF trakk seg ut på byrjinga av 1990-talet, blei senteret ein del av Unifob, ein stiftelse UiB hadde oppretta for å ta seg av oppdragsforskning.¹¹ Under skiftande namn – Humanistisk datasenter, HIT-senteret og Aksis – blei senteret her vidareført fram til 2018 då språkgruppa i det som då heitte Uni Computing, blei lagt ned.

Talemålsgransking: Folkeminne og den historiske språkvitskapen

Ei sentral grunngeving for professoratet i vestlandsk målgransking som blei oppretta ved Bergens Museum i 1918, var trongen for innsamling og dokumentasjon av dei vestlandske målføra.¹² Den fremste oppgåva som skulle leggst til stillinga, var «at samle og studere de vestlandske bygdemaal».¹³ Torleiv Hannaas (1874–1929) var einaste søkar. Han var allereie stipendiat ved museet, og vart tilsett som professor frå 1. juli 1918. I tråd med den tidas meir filologiske forståing av innhaldet i målgransking var han ikkje berre opptatt av det reint språkvitskaplege, men også av innhaldet i dei språklege overleveringane, som t.d. segner, stev og gåter. Hannaas var mest opptatt av det siste. I 1921 grunnla han Folkeminnesamlinga ved museet, femten år før Norsk målførearkiv blei grunnlagt ved universitetet i hovudstaden. Grunnstammen dei første åra blei hans eiga samling av folkeminne og språkminne.

Folkeminnesamlinga blei vidareført som eining under Det historisk-filosofiske fakultet då UiB opna i 1948. Først i 1953 blei namnet endra til Institutt for nordisk filologi, som eit resultat av at

11 Om Unifob, sjå Forland 1996: 496 flg. Stiftelsen blei omdanna til aksjeselskap i 2003. Sjå kapittel 11 i bind 1.

12 I arbeidet med denne bolken har eg hatt stor nytte av Sandøy 2007 og 2013.

13 Sitert etter Bondevik 1988.

morsmålsfaget hadde fått fleire tilsette og ansvar for undervisning. Då blei også samlinga splitta i to. Etno-folkloristisk arkiv er i dag knytt til faget kulturvitskap ved Institutt for arkeologi, historie, kultur- og religionsvitskap.¹⁴ Her ligg mykje av det Hannaas samla. Målføresamlinga blei verande ved det som i 1962 endra namnet til Nordisk institutt, som så i 2007 blei ein del av Institutt for lingvistiske, litterære og estetiske studium.¹⁵

Den andre professoren i vestlandsk målføregransking var Gustav Indrebø (1889–1942). Indrebø var svært breitt orientert. Som filolog har han sett varige spor etter seg i norrøn filologi, norsk språkhistorie, namnegransking og ikkje minst målstrid. I 1931 la Indrebø grunnlaget for Stadnamnsamlinga, som eg kjem tilbake til.¹⁶

Den tredje innehavaren av professoratet i vestnorsk målgransking, Olai Skulerud (1881–1963) var den første av dei tre som først og fremst interesserte seg for dialektologiske emne. I åra før han kom til Bergen, hadde han arbeidd som arkivar ved Norsk målførearkiv i Oslo. Skulerud sat i professoratet berre frå 1944 fram til han gjekk av med pensjon i 1951. Men i løpet av denne stutte tida la han grunnlaget for målføredelen av Folkeminnesamlinga. I om lag 250 katalogpostar i samlinga er han oppgjeven som oppskrivar eller intervjuar. I ein fagleg-biografisk artikkel om Skulerud finn vi denne karakteristikken:

[K]an hende det er rett å hevde at han er den mest utrerte samlaren innanfor norsk dialektgeografi. Frå dei tidlege arbeida som prøvande junggrammatikar der eit materiale skal bruka (sic) til å teste hypotesar, gikk (sic) han over til ei materialinnsamling som oftast vantar klåre hypotesar. Å samle og presentere «objektive data» til seinare bruk var målet.¹⁷

Sjølv nytta ikkje Skulerud materialet i eiga forskning i nokon monaleg grad. Det var, som det står i sitatet, samling av data for seinare bruk som var det viktige for han. Mot slutten av Skulerud si tid i

14 <https://www.uib.no/ahkr/22411/etno-folkloristisk-arkiv>, vitja 29.7.2020.

15 I 2020 vedtok universitetsstyret at ansvaret for samlingane skulle overførast til Språksamlingane ved Universitetsbiblioteket, sjå nedanfor.

16 For meir om Indrebø og dei faglege spora etter han, sjå t.d. Bondevik & Nes 1988, eller biografiske opplysningar i snl.no.

17 Sandøy 1988, sjå også Sandøy 2007, som skildrar vestnorsk målgransking fram til 1970-talet.

Bergen fekk UiB midlar til å tilsette hjelp til samlingsarbeidet, og i 1949 blei Håkon Hamre (1914–1972) amanuensis. Han var ein flittig bidragsytar så lenge han var i Bergen, men allereie i 1952 emigrerte han til USA for å bli professor i skandinavistikk ved University of California i Berkeley. Den som først og fremst tok opp arven etter Skulerud, var difor Gunnvor Rundhovde (1918–1987). Ho tok hovudfagseksamen i 1948 med ei avhandling om målet i dåverande Hamre kommune i Nordhordland, der ho sjølv kom frå. Ho var så vitskapleg assistent med løn frå NAVF i åra 1949–1951 for å arbeide med dialektane i Hordaland, særleg i Nordhordland. Etter å ha avslutta cand. philol.-graden i 1953 blei ho tilsett som vikar for Hamre hausten 1954. Først i 1956 sa Hamre opp stillinga si, og då fekk ho fast tilsetting med særleg ansvar for Målføresamlinga. Rundhovde disputerte i 1965 og var den første kvinna ved HF-fakultetet som gjorde det. Ho hadde ansvaret for samlinga fram til ho blei dosent i 1971, ho blei professor i 1983 og har som Skulerud ytt mange bidrag til samlinga.¹⁸

Dei viktigaste delane av samlinga fram til 1970 var oppskrifter av ulike slag på papir, manuskript, ordsetelsamlinga – eit setelarkiv sett inn i kartotekskuffar der kvar setel gjev språkleg informasjon om eit bestemt ord frå ein bestemt dialekt – ei samling utfylte spørjelister baserte på standardiserte lister laga for ulike føremål, og såkalla synopsisar, der resultat frå ulike, nærskyldte dialektar basert på den same spørjelista er sette saman i tabellform. Samlinga av lydopptak var framleis beskjeden.

I kva grad blei innhaldet i samlinga brukt som forskingsmateriale i samsvar med målet til Skulerud om innsamling av objektive data, og bidrog samlinga til ei styrking og i beste fall fornying av forskinga ved Nordisk institutt på 1950- og -60-talet? Svaret på det siste spørsmålet er klart nei. Den papirbaserte samlinga var, slik ho låg føre i 1970, eit ektefødt barn av den historiske målgranskinga grunnlagt på 1800-talet. Å nytte den som grunnlag for strukturalistiske og seinare generative analyser er nok mogleg, men materialet ligg ikkje beinveges til rette for det. Eit sitat frå dåverande førsteamanuensis Jarle Bondevik (1934–2016), der den første setninga viser til strukturalistisk metode, viser kanskje kor rotfest det historiske, før-strukturalistiske perspektivet var:

18 Nes 1996a og b. I tillegg inneheld samlinga bidrag frå mange andre heimelskvinner og -menn.

Det er eit typisk trekk at i fleire monografar vert det sett opp kva fonemsystem som finst i talemålet ved hjelp av kommunikasjonstestar, men forfattarane held ikkje fast på prinsippa og går over til å skriva meir tradisjonelt fonetisk. Ei forklaring på at dette hender, kan vera at i mange avhandlingar vert lyd-systemet skildra diakront, og når det moderne systemet skal jamførast med det gamalnorske, strekk ikkje den fonemiske attgjevinga til, [...].¹⁹

Det eldste eksempelet på dette, som Bondevik kanskje har hatt i tankane, er ei hovudoppgåve frå 1952 om Fana-målet, utgitt som bok i 1957.²⁰ Ho har blitt vist til som eit tidleg eksempel på bruk av strukturalistisk metode i Bergen.²¹ Og det er rett at boka inneheld ei oppstilling av «eit vokal-system med 10 fonem».²² Men forfattaren gjer ikkje greie for korleis han har etablert systemet, og resten av boka er ei reint tradisjonell, historisk framstilling. Noko liknande kan seiast om Rundhovde si doktoravhandling frå 1964, som analyserer målet i gamle Hamre kommune i Nordhordland. Ho bruker omgrepet 'fonem', og set opp vokalsystem for eldre og yngre, men viser ikkje korleis ho har etablert systema.²³ Elles er dette også ei tradisjonell, historisk framstilling av målføret. Først i ein publikasjon som kom i 1978, tar ho skrittet over i strukturalismen fullt ut. Her kombinerer ho stringent strukturalistisk metode med eit historisk perspektiv på ein svært innsiktsgjevande måte.²⁴

Når det gjeld det første spørsmålet, i kva grad arkivet blei brukt i forskning, er det vanskeleg å skaffe seg ei påliteleg oversikt. Det langt frå presise bildet som teiknar seg når ein søker på 'målføresamlinga' og 'målførearkivet' i Nasjonalbibliotekets nettbibliotek, er at Målføresamlinga i Bergen og Norsk målførearkiv ved UiO i liten grad blei brukt av andre forskarar enn dei som sjølve hadde bidratt med materialet. Det er nok symptomatisk at ei artikkelsamling som blei utgitt i 1973, og som på mange måtar kan sjåast på

19 Bondevik 1987.

20 Skre 1957.

21 Venås 2005: 300.

22 Skre 1957: 17.

23 Rundhovde 1964. Litteraturlista inneheld ein referanse til ei av de klassiske internasjonale innføringane i strukturalistisk lydanalyse, Bloch & Trager 1942, men det er vanskeleg å lese ut av teksten korleis ho er nytta.

24 Rundhovde 1978.

som ei oppsummering av den historisk orienterte norske målfø-regranskinga fram til 1970, ikkje inneheld ein einaste referanse til arkivmateriale brukt som data.²⁵ Heller ikkje oversiktsartikkelen til Bondevik nemnt ovanfor, som kom 15 år seinare, nemner arkiva.²⁶

Den viktigaste grunnen til at det blei slik, er nok at det korkje i Bergen eller på landsbasis var mange som forska på talemål på full tid. Og med nokre unntak arbeidde desse med data dei sjølv hadde samla inn. Langt dei fleste forskingsbidraga var hovudfagsoppgåver. Her var det eit krav at studenten sjølv samla inn materialet (som når oppgåva var levert, gjekk inn i samlinga). Det er også grunn til å hevde at skepsisen i nordiskmiljøa til dei strukturalistiske perspektiva på mange måtar avgrensa arkiva sin forskingsrelevans til dokumentasjon og samanstilling av dokumentasjon.

For den historiske målgranskinga var det overordna målet at lydskrifttranskripsjonane skulle vere korrekte. Dersom ein då ikkje interesserte seg for i kva grad det som var transkribert, kunne opne for innblikk i ein abstrakt, underliggande struktur, var den einaste måten ein kunne nytte dette stoffet vidare på, å samanlikne materiale frå ulike dialektar og syntetisere det i såkalla synopsisar. Resultata kunne ein synleggjere på kart, som viste grensa mellom område der ei endring hadde skjedd, og der den ikkje hadde skjedd – såkalla isoglossar.

Til ein viss grad er dette ei karikert beskriving, som dekker over at det frå slutten av 1960-åra kom inn arkivstoff, t.d. i form av hovudoppgåver, som la større vekt på grammatisk struktur og strukturalistiske perspektiv. Men dette endrar ikkje i monaleg grad den konklusjonen at den viktigaste funksjonen til dette materialet i dag er dokumentasjon – dokumentasjon av dialektar og dokumentasjon av språkendingar og deira råderom.

25 Beito & Hoff 1973. Ein av redaktørane, Ingeborg Hoff, hadde sjølv heile sitt yrkesliv ved UiO vore knytt til Norsk målførearkiv og var leiar der frå 1972 til 1981.

26 Bondevik 1987. To unntak er Kjell Venås' to oversiktsverk om verbbygging i norske målføre, Venås 1967, 1974, der også Målføresamlinga i Bergen er nemnt i forordet.

Stadnamn og stadnamngransking

Stadnamn er både overlevert talespråk og folkeminne. Indrebø vart utpeika som styrar av Norsk stadnamnarkiv då Magnus Olsen grunnla arkivet i Oslo i 1921, og han heldt fram i den funksjonen då han vart professor i vestnorske målføre ved Folkemинnesamlinga i Bergen i 1930. I 1939 vart arkivet flytta til Bergen, men etterfølgjaren Hovda tok det i 1944 tilbake til Oslo etter at Indrebø døydd i 1942. Overføringa skjedde i ein kuppliknande operasjon gjennomført «med militær presisjon», i sterk konflikt med interessene til leiinga ved Bergens Museum som såg at arkivet kunne bli eit viktig element i oppbygginga av eit humanistisk fagmiljø ved det nye universitetet som var i emning.²⁷ Arkivet blei overført til UiO i 1978, men kom tilbake til Bergen i 2016 og ligg no i Språksamlingane (sjå nedanfor).

Det som vart til Stadnamnsamlinga i Bergen, bygde i stor grad på det store innsamlingsarbeidet som Indrebø sette i gang i 1931, dei såkalla skulebarnoppskriftene.²⁸ Norske skuleborn vart i åra fram til 1935 bedne om å skrive opp namn frå heimstaden, så nær uttalen som råd. Det kom inn eit stort materiale, truleg mellom 1 og 1,5 millionar namn.²⁹

I 1960 vart Øystein Frøysadal (1918–1969) tilsett som amanuensis ved det som då hadde vorte til Institutt for nordisk filologi, «med særleg skyldnad til å arbeida med norske stadnamn».³⁰ Frøysadal konsoliderte fagtradisjonen etter Indrebø med djupinnsamling basert på feltarbeid og kartotekkort og namnetolking med sterk vekt på etymologiske perspektiv. Etterfølgjaren Oddvar Nes (1938–2016) vidareførte dette arbeidet, og han utvida arbeidet med å organisere innsamling av stadnamn frå Vestlandet, m.a. gjennom kurs og rettleiing.³¹ Metodisk stod feltarbeid sentralt, etter kvart i kombinasjon med kartfesting på dei nye karttypane gradteigskart og økonomiske kart.

Den siste leiaren av Stadnamnsamlinga var førsteamanuensis Ole-Jørgen Johannessen frå 1989 til 2012. Arbeidet med sys-

27 Ådland 1997: 54, Helleland 1997.

28 Bondevik og Nes 1988: 61.

29 Oversikt over materiale i Stadnamnsamlinga 20.02.2001, i privatarkivet til Ole-Jørgen Johannessen; opplysningar frå Peder Gammeltoft 18. oktober 2021.

30 Nes 1996a: 61.

31 Akselberg 1998: 7.

tematisering og tilgang til stadnamna tok ei digital vending etter år 2000, initiert av Det humanistiske fakultet. Det var bygd opp store samlingar av stadnamn frå Hordaland og andre fylke, men dei var upraktiske og tungvinne å bruke både for forskarar og folk flest. Frå fakultetet kom det krav om at tilgangen for ålmenta måtte aukast gjennom å ta i bruk ny teknologi.³² Stadnamnsamlinga innleidde samarbeid med miljøet for språkteknologi ved Aksis, seinare Uni (sjå ovanfor). Det gav to viktige nyvinningar.³³

Den eldre namnesamlinga frå Hordaland vart innskanna i 2004–2005 i det forskingsrådsfinansierte infrastrukturprosjektet «Etablering av digital nasjonal database for stedsnavn, fortrinnsvis mikrotoponymer (smånavn)», eit samarbeid med stadnamnmiljøet ved UiO.³⁴

Den andre viktige nyvinninga var databasen og nettsida Hordanamn, den òg utvikla i samarbeid med Aksis og Unifob.³⁵ Målet med nettsida var å vise stadnamna kopla til kartreferanse og uttale, og planen var å kanalisere all ny innsamling og innskriving av namn inn mot denne nettsida. Metodisk må særleg koplinga til lydfiler kallast ei sentral teknologisk nyvinning. Rundt 2012 inneheldt databasen 275 000 kartfesta stadnamn. Av dei var 110 000 namn lydfesta, og systemet inneheldt data frå 20 av 33 kommunar i Hordaland.³⁶

Digitaliseringsarbeidet fekk dei første åra små økonomiske bidrag frå institutt og fakultet, som gradvis blei reduserte. Stadnamnsamlinga la derfor i denne tiårsperioden ned mykje arbeid i å hente inn tilskot frå både kommunar og fylkeskommunale arkiv for å finansiere innskriving av namn. Johannessen heldt fram med dette koordineringsarbeidet dei næraste åra etter 2012, sidan stillinga ved Stadnamnsamlinga då var inndregen. Ansvaret for Hordanamn vart overført til Språksamlingane i 2018 då språkgruppa i Uni Research vart nedlagd.

Bak prototypen til Hordanamn låg det tydelege visjonar.³⁷

32 Johannessen 2009.

33 Framstillinga her byggjer dels på samtalar med Ole-Jørgen Johannessen og Øystein Reigem i oktober 2021.

34 Johannessen og Schmidt 2004, Johannessen 2009. Den nasjonale stadnamndatabasen er no ein del av Språksamlingane, sjå nedanfor.

35 Hordanamn.uib.no, sett 16.10.21, siste revisjon ved Øystein Reigem 18.06.2014.

36 Johannessen 2012: 22 flg. Ifølgje Peder Gammeltoft 18. oktober 2021: 25 av 33 kommunar og 192 000 namn.

37 Johannessen 2009. Reigem var tilsett i Uni.

Miljøet som stod bak, var ikkje stort, i tillegg til Johannessen deltok Øystein Reigem med programmeringsarbeid og sekretæren for stadnamntenesta i Språkrådet, Kjell Erik Steinbru, med innskrivingsarbeid på deltid. Nedbygginga både på den namnefaglege og den språkteknologiske sida gjorde altså at arbeidet bremsa opp.

Som vitenskapleg disiplin har stadnamnforskinga dei siste åra opplevd nedgang ved norske universitet, og særleg Kulturdepartementet har uttrykt sterk uro for utviklinga framover.³⁸ I Bergen var 1970- og -80-talet derimot ei glanstid for stadnamngranskinga.³⁹ Det gjekk ikkje noko skarpt skilje mellom samlingsarbeid, forskning og undervisning. Ein stor del av tilfanget i stadnamnsamlinga kom inn til instituttet gjennom hovudoppgåver, og det fanst etter kvart mange lektorar i skuleverket som kunne rettleie elevane i å skrive særoppgåver om stadnamn. Miljøet ved Stadnamnsamlinga la ned mykje arbeid i utoverretta formidling, kurs og rettleiing som stimulerte innsamlingsarbeidet lokalt. Fleire doktorgrader er leverte i tilknytning til miljøet i Bergen. Fleire tilsette med stadnamnkompetanse har vore namnekonseptantar for Vestlandet, og det sterke fagmiljøet gjorde at Språkrådet frå 1990-talet la sekretariatsfunksjonen for den regionale stadnamntenesta hit. Med flyttinga av språksamlingane frå Oslo (sjå nedanfor) er dei nasjonale stadnamnressursane på veg til å bli sentraliserte i Bergen.

Det sosiolingvistiske paradigmeskiftet

Den historiske språkforskinga interesserte seg for utgangspunkt og resultat av språkendingar, men i liten grad for den prosessen som låg mellom og som var prega av variasjon mellom gamle og nye former. Midt på 1960-talet blei dette metodiske grunnlaget for talemålsforskning kasta om, først og fremst av amerikanaren William Labov. Hans teoriar kan sjåast som eit av dei tidlege brota med Chomskys avvising av språkbruksdata, sjølv om Labov sjølv hevda at ideane hans representerte ei utviding og vidareføring av den generative grammatikken.

Labov kombinerte synkrone og diakrone perspektiv. Gjen-

38 Sjø m.a. Prop. 108 L (2019–2020) *Lov om språk (språkløva)*, kap. 9.2.2.

39 Akselberg 1998: 7.

nom nokre nyskapande undersøkingar viste han at variasjonen mellom yngre og eldre former var systematisk, og at den i stor grad samvarierte ikkje berre med talarane sin alder, men òg mellom anna med sosial bakgrunn, etnisitet og kjønn. På grunnlag av dette mønsteret kunne han vise korleis ei endring spreidde seg i språksamfunnet, gjennom generasjonar og samfunnslag. Og mens dialektologien i stor grad hadde arbeidd med små, rurale språksamfunn, blei interessa no retta mot dei mykje meir komplekse språkbruksmønster i større byar.⁴⁰ Rådata blei samla inn i form av lengre innspelningar av eit stort, mest mogleg representativt utval informantar tatt opp i mest mogleg naturlege samtalesituasjonar. Opptaka blei så analyserte kvantitativt med fokus på språktrekk som viste variasjon – såkalla språklege variablar. Boka som har vorte rekna som eit sentralt utgangspunkt for det som raskt fekk namnet sosiolingvistik, *The Social Stratification of English in New York City*, kom i 1966.⁴¹

Det tok ikkje lang tid før Labovs idear, og tilhøvet mellom språk og samfunn meir generelt, vekte interesse også i Noreg. I Bergen organiserte dei unge universitetslektorane Kerstin Nordenstam (1935–2016) og Ole-Jørgen Johannessen (1942–) tidleg på 1970-talet ei lese- og diskusjonsgruppe for interesserte nordiskstudantar. I løpet av dette tiåret vart det så sett i gang store, innbyrdes uavhengige prosjekt i dei tre største byane i landet, *TAUS-prosjektet* i Oslo, som starta i 1970,⁴² *Språksosiologiske forhold i Trondheim bymål* (1975),⁴³ og *TUB-prosjektet* i Bergen, som starta i 1977. TUB var eit akronym for *Talemål hos ungdom i Bergen*. Prosjektet fekk finansiering frå NAVF. Prosjektleiari i planleggings- og feltarbeidsfasane var amanuensis Helge Omdal. Sjølv tilarbeidings- og analysefasen blei leia av dåverande amanuensis Helge Sandøy (1947–).

TUB-prosjektet spelte inn eit stort materiale. 92 ungdomar

40 Her var Norge eit unntak i internasjonal samanheng. Også mange bymål blei undersøkte i løpet av første halvdel av 1900-talet, med ein analyse av bergensmålet som den første – ein analyse som omfatta både dialekten si «mere vulgære og dens mere dannede form», Larsen & Stoltz 1912: 3.

41 Labov 1966. Det oppstod raskt meir kvalitative retningar innan sosiolingvistikken, men i Bergen har det i stor grad vore den kvantitative tradisjonen etter Labov som har dominert.

42 Hanssen, Hoel, Jahr, Rekdal & Wiggen 1978: 1. I tillegg munna prosjektet ut i ei rekke mindre artiklar.

43 Fintoft & Mjaavatn 1980.

fødd i 1962–64, likt fordelte på 8 bydelar, utgjør kjernen. Det var like mange av kvart kjønn, og dei var også fordelte nesten likt på tre sosialgrupper. Alle måtte ha budd i same bydelen heile livet, ha gått på barneskule der og ha mor som også var fødd og oppvaksen i Bergen. Som referansegruppe blei det gjort opptak med 12 eldre bergensarar fødd i Bergen sentrum i 1908–09, oppvaksne der og med foreldre som også var frå sentrum.⁴⁴

Den metodiske kjernen i denne typen undersøkingar er som nemnt den såkalla språklege variabelen, dvs. språktrekk som kan ha meir enn éin lydleg realisasjon utan at tydinga blir endra. Det er desse variantane som samvarierer med ulike bakgrunnsvariablar, som sosial status og ofte kjønn. Variablar i bergensmålet som blei undersøkt, var t.d. vekslinga mellom *ikke* og *ikkje*, *noen* og *nokken*, *jei* og *eg*, og ikkje minst samanfallet av *kje*-lyden og *sje*-lyden, som for første gong i Noreg blei observert i dette materialet hos så gamle språkbrukarar, særleg hos ungdomane frå Fyllingsdalen, Ytre Fana og Bergen sentrum. Sett i forhold til kor vanleg dette fenomenet blei i åra som følgde, var frekvensen likevel svært liten, berre 2,5 prosent av alle førekomstar av dei to lydtypene blei realiserte «feil».⁴⁵

Ein stor del av opptaka blei transkribert, og variantane ein var interessert i å analysere, fekk ulik form. Transkripsjonane blei dataførte, og ved hjelp av programvare utvikla i samarbeid med Bjørn Eide ved Program for datamaskinell språkbehandling ved Nordisk institutt kunne ein søke i transkripsjonane og identifisere alle belegga på ein gitt variabel, samt frekvensen av dei ulike variantane kryssa med dei ulike bakgrunnsvariablane.⁴⁶ Denne transkripsjonen var det første digitale talemålkorpuset i Bergen. Prosjektleiaren Sandøy skildra nytten til korpuset slik, samanlikna med dei tradisjonelle metodane:

I praktisk analysearbeid er det ofte slik at ein etter ei stund kjem i tvil om kategoriane er rette. Det kan komme på tale å omdefinere både variablane og variantkategoriane. Denne revurderinga av hypotesar og variablar skjer når ein sit med

44 Sandøy 1983. Resultata frå prosjektet er rapportert i ein eigen skriftserie, *Talemål i Bergen*, som kom ut med til saman fire hefte frå 1983 til 1988 med Nordisk institutt som utgjevar.

45 Johannessen 1983.

46 Sandøy 1984 og etter personleg opplysning frå Sandøy 2.3.2021.

materialet framfor seg. I slike tilfelle er det ein umisteleg fordel å ha belegga samla i velredigerte filar (eller datautskriftar) – i staden for å ha dei spreidd på dei fleire tusen transkripsjonsark.⁴⁷

På 1970-talet kom også dei første sosiolingvistiske hovudoppgåvene om bymål, også dei baserte på innspelinger av eit relativt stort tal informantar.⁴⁸ Samtidig ble det etter kvart slutt på dei tradisjonelle dialektologioppgåvene baserte på kartlegging av færre og eldre informantar. At målføregranskinga gjennomgjekk eit paradigmeskifte på denne tida, kan vi også lese ut av resultatata frå dei fire feltkursa instituttet arrangerte for studentar mellom 1975 og 1981. På alle fire gjorde ein opptak. Men det ein gjorde opptak av, dokumenterer eit brot. Ifølge katalogen for Målføresamlinga var hovudvekta av opptaka på dei to første kursa, i høvesvis Høyanger i 1975 og Førde i 1976, knytt til dei gamle spørjelistene med røter i den historiske språkvitskapen, sjølv om det også blei gjort opptak av samtalar. Få år seinare, på Stord i 1980 og i Lindås i 1981, var ikkje ordlistene med lenger; no låg hovudvekta på opptak av såkalla sosiolingvistiske intervju, dvs. mest mogleg frie og naturlege samtalar.

Denne dreininga mot sosiolingvistiske metodar førte til at det var opptak på spoleband, kassettar og så digitale medium som fram mot tusenårsskiftet utgjorde den viktigaste tilveksten til Målføresamlinga. I teorien kunne dei ha lidd same skjebne som papirmaterialet, og blitt liggande unytta som dokumentasjon. Men digital teknologi gjorde at opptaka i staden kunne tilgjengeleggjerast for forskning på måtar ein ikkje kunne ha tenkt seg då dei første opptaka blei gjorde.

Første steg var å digitalisere dei gamle, analoge opptaka. Her var flaskehalsen i bytjinga kostnader knytte til lagringskapasitet, men dette var ikkje lenger noko problem etter tusenårsskiftet. Neste steg var å utvikle ein teknologi der ein først transkriberte opptaka til standard nynorsk eller bokmål og deretter kopla transkripsjon og lydfil på ein måte som gjorde at ein kunne bruke transkripsjonen

47 Sst.: 34 flg.

48 Den første hovudoppgåva ved Nordisk institutt som la Labovs metodar til grunn, var Reidar Sandals analyse av Vågsøy-målet (1976). Deretter følgde fleire oppgåver om bymål: Seim 1977 (Kragero), Kristoffersen 1978 (Arendal), Hangeland 1981 (Kristiansand) og Gabrielsen 1983 (Stavanger).

som søkeverktøy. Dette skjer på den måten at ein skriv inn eit ord eller ein relevant bokstavsekvens i søkevindauget. Programvara returnerer så alle treff i form av ei liste, der ein frå kvart treff kan opne eit utsnitt av lydfila som inneheld treffet saman med samanhengen føre og etter. Slik kan ein raskt og effektivt identifisere og hente ut relevante data. Sentral i den tekniske tilpassinga av teknologien i Bergen frå sist på 1990-talet var Knut Hofland (1949–) og Paul Meurer (1962–) ved HIT-senteret/ Aksis.

I åra som følgde, blei teknologien testa ut m.a. i fleire doktorgradsprosjekt.⁴⁹ I 2008 fekk professor Sandøy finansiering frå Noregs forskingsråd av eit prosjekt kalla *Dialektendringsprosessar*. I 2010 fekk sosiolingvistikkmiljøet endå eit prosjekt finansiert av Forskingsrådet, *Språkutvikling på industristader*. Målet i begge prosjekta var å granske språkendingar i nyare tid ved å samanlikne eldre opptak med nye innspelningar som skulle gjerast i regi av prosjekta. No blei eldre opptak, m.a. dei gamle TUB-opptaka, digitaliserte og kopla til transkripsjonar. Saman med dei nye opptaka – i nokre høve av dei same personane som blei spelt inn på 1970-talet – blei dei samla i eit digitalt talemålskorpus som fekk namnet *Talebanken*.⁵⁰ Dei tekniske løysingane her har det i stor grad vore Meurer som har stått for, medan Sandøy har vore den drivande faglege krafta. Også opptak gjorde i samband med andre prosjekt, ikkje minst knytte til masteroppgåver, blir transkriberte og lagt inn i korpuset, slik at studentane kan nytte teknologien i analysane sine.

Talebanken er med god margin det største talemålskorpuset i Noreg i 2021.⁵¹ I januar 2021 inneheldt banken ca. 7,5 millionar ord frå 48 ulike kommunar, dei aller fleste på Vestlandet. Talet på talarar er snautt 1600 – den eldste fødd i 1864, den yngste i 2005. Omfanget av forskinga knytt til det sosiolingvistiske miljøet og Talebanken har vore betydeleg. Heimesida til prosjekta viser til 20 mastergradar, to doktorgradar og nær 40 vitskaplege og populærvitskaplege

49 Selås 2003, Haugen 2004, Hernes 2006.

50 Sjå <https://clarino.uib.no/korpuskel/corpus-list?collection=Talebanken> og <https://www.uib.no/fg/sosiolingvistikk/90666/talebanken#-rsmelding-2019>, vitja 22.9.2020.

51 Det nest største var det såkalla LIA-korpuset ved UiO med ca. 3,5 millionar ord. Innhaldet her er digitaliserte, eldre opptak opphøveleg gjort på spole- eller kassetband og som var ein del av samlingane ved dei fire eldre universiteta. I dette prosjektet deltok UiB som partnar, og heile 64 % av opptaka som blei tatt inn i korpuset, kom frå Målføresamlinga ved UiB. Sjå elles <http://www.tekstlab.uio.no/LIA/prosjekt.html>, vitja 27.1.2021.

publikasjonar som kan knytast til prosjekta.⁵² I stor grad dreier det seg om forskning på språktilhøve på Vestlandet. Slik sett kan ein seie at målet for professoratet i vestnorsk målgransking, slik det blei formulert i 1918 – «at samle og studere de vestlandske bygdemaal» – blir vidareført, men med oppdaterte metodar og ei utviding av perspektivet til òg å gjelde bymål.

Leksikografi og terminologi

Leksikografi, eller leksikologi, er vitskapen om ordforrådet i eit språk og korleis ein kan framstille dette ved hjelp av ordbøker. I prinsippet skil ein mellom dei to termene, slik at leksikologi er det vitskaplege studiet av korleis ordforrådet er bygd opp, medan leksikografi refererer til det meir konkrete ordboksarbeidet. I praksis blir dei ofte brukte om kvarandre.⁵³

Terminologi refererer til det spesielle ordforrådet som er i bruk innanfor ulike fagfelt og profesjonar. Saman med ulike tekst- og sjangernormer skal den sikre effektiv og presis kommunikasjon innan kvart fagfelt.⁵⁴ Terminologien vil samstundes vere eit vern mot såkalla domenetap til engelsk, i ein situasjon der mange nye termar og omgrep kjem derifrå. Ein vellukka terminologi er eit resultat av samarbeid mellom ekspertar på fagfeltet og filologar. Termene som utgjer terminologien, må ha eit presist, avgrensa innhald, og relasjonane mellom dei må vere klåre og veldefinerte. Ein god terminologi er med andre ord eit system, og ikkje ei meir eller mindre tilfeldig samling definisjonar.

Norsk ordregistrant

Heilt frå starten i 1967 var hovudprosjektet ved Prosjekt for datamaskinell språkbehandling (PDS) *Norsk ordregistrant*, eit datamaskinelt ordarkiv over norsk allmennspråk, bokmål og nynorsk, med opplysningar om ordklasse og eventuelt bøyingsmønster for kvart

52 <https://folk.uib.no/hnohs/DEP/> og <https://folk.uib.no/hnohs/IP/>, vitja 27.1.2021.

53 Sjø til dømes definisjonane i *Det Norske Akademis ordbok* (<https://naob.no>), vitja 28.9.2020.

54 Myking, Dysvik & Hjulstad 2018.

ord. I 1970 inneheldt basen vel 31 000 ord, henta frå vanlege skuleordlister for bokmål og nynorsk.⁵⁵ I 1976 hadde basen vakse til vel 60 000 ord, og planen var å utvide tilfanget til vel 200 000.⁵⁶ Men dette gjekk trått, i 1979 var basen framleis berre på ca. 60.000 ord. Då hadde ein fått tilgang til ein dataleseleg versjon av Tanums rettskrivingsordbok, og ein såg for seg at det kunne lagast ein stor og ein liten versjon av registranten med 300 000 og 60 000 ord.⁵⁷ Men dette skjedde ikkje; i eit udatert notat påført «[1985, ev. før]» med blyant, blir talet på ord i registranten framleis oppgjeve til ca. 60 000.⁵⁸ Arbeidet med og utviklinga av registranten ser ut til å ha stoppa opp.

Ordregistranten var eit samarbeid med Norsk språknemnd, frå 1972 Norsk språkråd (frå 2005 Språkrådet). Tanken var at registranten skulle kunne fungere som utgangspunkt for utgiving av m.a. ordlister til skulebruk. Men då ein tok til å diskutere nye ordlister på midten av 1980-talet, går det fram av eit møtereferat frå Språkrådet at ordregistranten ikkje hadde vore ajourført på fleire år, og at det måtte minst eitt årsverk til for å få gjort dette.⁵⁹ Oppdateringsbehovet må ha vore knytt til rettskrivingsreforma av bokmål frå 1981. Når resultatata av denne ikkje var lagde inn i registranten, var han ubrukbar som grunnlag for offisielle ordlister og ordbøker. Dette årsverket fann ein ikkje midlar til, og resultatet blei at samarbeidet med Norsk språkråd om ordlister tok slutt. Etter dette forsvinn dei fleste spora etter registranten.⁶⁰ Den blei lagt til sides då terminologiprojekta knytte til petroleumsverksemda blei etablerte, og blei aldri tatt fram igjen.⁶¹

Om lag samstundes slutta Heggstad ved UiB i 1984 for å starte

55 Heggstad 1970.

56 Heggstad 1976.

57 Alhaug 1978.

58 UBB-*Ms* 2175, nr. 14.

59 UBB-*Ms* 2175, nr. 16j.

60 Språkrådet samarbeida òg med leksikografimiljøet ved UiO. Dette blei no vidareført og resulterte i utgjevinga av førsteutgåvene av dei to såkalla handordbøkene i 1986, *Bokmålsordboka* og *Nynorskordboka*. Sjå meir nedanfor. Registranten kom òg til nytte som grunnlag for tre større ordbokprosjekt. Auken frå vel 30 000 ord til vel 60 000 på 1970-talet kom hovudsakleg frå ein dataleseleg versjon av den norske delen av Einar Haugens *Norsk-engelsk ordbok* (Alhaug 1978). Denne blei nytta som grunnlag for utgjevingar av tre større ordbøker i samarbeid med Universitetsforlaget. I tillegg til nye utgåver av Haugens bok kom ei norsk-tysk ordbok (Hustad 1979) og ei norsk-fransk ordbok (Grundt 1991), hovudredaktør var professor Lars Otto Grundt ved Romansk institutt.

61 Samtale med Johan Myking 29. januar 2021.

sitt eige firma, Logos A/S, der han ville kommersialisere noko av den kompetansen han hadde opparbeidd seg som leiar av PDS. Etter det blei det usemje om rettane til materialet som var blitt utvikla ved PDS. Ifølge oppslag i Bergens Tidende sommaren 1985 skal universitetstilsette med tilknytning til Logos ha kopiert materiale for firmaet og i tillegg brukt arbeidstid på eit prosjekt som Logos hadde teke over frå universitetet på ein måte som ikkje minst fagmiljøet ved Nordisk institutt fann svært kritikkverdig.

Dette er eit tidleg eksempel på at forskingsdata kunne ha kommersiell verdi, og for universitetet blei dette ei vanskeleg sak. Ifølge assisterande universitetsdirektør Sverre Spildo hadde saka avdekka uklårheiter i regelverket som universitetet ville dra lærdom av. Samstundes var det trass alt «en hovedoppgave for universitetet å spre kunnskap og resultater».⁶² Heggstad på si side hevda det same – universitetet skulle vere «en faglig ressurs for samfunnet». Saka hamna aldri i rettsapparatet, men var eit godt døme på at det trongst nytenking og klårare køyrereglar for korleis interessekonfliktar av denne typen skulle handterast. Desse kom etter kvart på plass, og i dag er kommersialisering og samarbeid med næringslivet eit uttalt mål for universiteta.⁶³ Kommersialisering med utgangspunkt i materialet frå PDS blei det likevel ikkje noko av.

Norsk termbank

Allereie i 1976 drøfta Heggstad tanken om ein fagordregistrant som kunne danne ein heilskap med den allmennspråklege ordregistranten omtalt i førre avsnitt.⁶⁴ Det eigentlege arbeidet hadde enno ikkje begynt, men Heggstad nemner to forstudiar: EDB-terminologi og petroleumsterminologi. I tillegg nemner han det som skal bli framtida, nemleg å lage terminologiar som interaktive databasesystem som kan oppdaterast løpande – i staden for bøker, som raskt blir utdaterte innanfor felt som er i rask utvikling, slik både edb- og petroleumsverksemda var på 1980- og -90-talet.

62 Sjä oppslag i Bergens Tidende 6.7., 10.7. og 19.10.1985.

63 Saksførelegget til kollegiesak 132/1999 gir ei oppsummering av korleis den strategiske tenkinga om dette utvikla seg på 1990-talet.

64 Heggstad 1976.

Då Norsk termbank blei oppretta ved PDS i 1979, var dette ei realisering av Heggstads tankar om ein fagordregistrant. Samarbeidspartnarar var Norsk språkråd, Rådet for teknisk terminologi og Universitetsforlaget. I utgangspunktet var ikkje dette nokon einskapleg base, men «eit samlenamn på terminologiprojekta ved PDS».⁶⁵ Tankar om ei fastare organisering blei det ikkje noko av på grunn av motstand frå departementet.⁶⁶ I 1986 bytta PDS namn til Avdeling for norsk leksikologi, og Norsk termbank blei fagleg underordna denne.

I løpet av det første året blei det utvikla eit felles format som materialet i dei ulike projekta blei konvertert til.⁶⁷ Resultatet blei ei teknisk fagordbok på snautt 750 sider, støypt saman frå ei rekke tidlegare utgitte termordlister.⁶⁸ Boka var ifølge forordet ein samla presentasjon av fagterminologien som til då var skoten inn i termbanken. Det sentrale i prosjektet er likevel den elektroniske basen, som kan ajourførast løpande og gi grunnlag for «oppdaterte og utvidete utgaver av [...] boka». Men det kom aldri fleire utgaver. Grunnane til dette kan vere fleire. Fråvære av tilstrekkelege økonomiske ressursar var nok ein av dei, det at basar av denne typen nokre år seinare kunne leggest ut på internettet, ein annan. Ein tredje er nok at aktiviteten ved Norsk termbank i åra som følgde, meir og meir kom til å dreie seg om oljeterminologi – den utan tvil største suksessen i termbankens historie.

Rådet for teknisk terminologi (RTT) sette allereie i 1975 ned eit utval med ansvar for terminologi knytt til petroleumsverksemda, der UiB var representert.⁶⁹ Det er ikkje oppgitt kven som representerte UiB, men det er vel liten tvil om at det var Heggstad. I førsteutgåva av *Ordbok for petroleumsvirksomhet* frå 1976 er ikkje PDS nemnt, men i forordet til den andre utgåva frå 1980 står det at innhaldet er lagra i databasen Norsk termbank og at ein der har utvikla program og stått for tekstbehandling og redigering av manuskriptet, noko som hadde forenkla arbeidet vesentleg.⁷⁰

65 Hjulstad 1985.

66 Myking 2005.

67 Hjulstad 1980.

68 Hjulstad & Norevik 1984.

69 Rådet for teknisk terminologi var ein medlemsorganisasjon som går tilbake til 1930-talet. Det samarbeidde med Norsk språkråd, og gav ut termlister knytte til ulike fagfelt. Frå slutten av 1970-åra til 1985 vart materialet RTT då radda over, dataført ved PDS. Rådet vart lagt ned i 2001, Myking 2005.

70 Ordbok for petroleumsvirksomhet – RTT 43 1980: 6 flg.

I ein presentasjon av Norsk termbank frå 1980 er samlinga som låg til grunn for ordboka, nemnd i oversynet over innhaldet i banken. Petroleumsfeltet er likevel ikkje nemnt som eit spesielt satsingsområde.⁷¹ Men det må ha vore kort tid etter at det professor Johan Myking (1952–) kallar «sentrale aktørar, ikkje minst i og rundt PDS-miljøet», starta eit «aktivt lobbyarbeid mot oljemiljøet i Stavanger-regionen, det vil seia Oljedirektoratet og dei sentrale oljeselskapa med Statoil i spissen».⁷² Resultatet av arbeidet var at Statoil (Equinor) bestemte seg for å finansiere Terminol-prosjektet, med PDS/Norsk termbank som utøvande faginstans.⁷³ Prosjektet, som skulle vare frå 1983 til 1985, var saman med oppfølgingsprosjektet Gullfaks B «det største terminologiske einskildprosjekt i Noreg nokosinne».⁷⁴ Målet var å utvikle eit norsk arbeidsspråk på Gullfaksfeltet i Nordsjøen og gjere norsk til det offisielle arbeidsspråket der. Det var med andre ord ikkje berre tale om å skape ein norsk terminologi på papiret, det viktigaste var at han skulle gjerast om til praksis:

Den terminologien som dermed vart utvikla, vart planta inn i alle former for tekst på plattformene: skilting, dataskjerm-bilete i prosessstyring, driftsprosedurar og ikkje minst alle former for opplæringsmateriell. Konsekvensen var såleis at det totale arbeidsmiljøet på mange plattformer vart prega av norsk språk.⁷⁵

Det lèt seg i alle fall identifisere tre viktige utgangspunkt for Terminol-prosjektet. Det første er at norske firma, med statseigde Statoil i spissen, blei etablerte og tok opp konkurransen med dei utanlandske pionerane. Samtidig kom norske arbeidstakarar i stadig større grad til å erstatte utanlandske i Nordsjøen. Dette førte til auka merksemd om tryggleiksspørsmål og kommunikasjon.⁷⁶ Skal reglar og prosedyrar følgast, ikkje minst i krisesituasjonar, er det

71 Hjulstad 1980.

72 Myking 2005.

73 Ein forløpar var ei trespråkleg ordbok utgitt av Elf Aquitaine i 1982 (Paruit 1982). I forordet blir PDS og Norsk termbank takka for hjelp «i den generelle oversettelses- og registreringsprosess til norsk for den norske versjonen».

74 Myking 2005, Myking, Dysvik & Hjulstad 2018. Påstanden står seg nok også i dag.

75 Myking & Sæbøe 2000.

76 Sjø t.d. Flydal 1983.

viktig at dei blir forståtte fullt ut. Sjølv om mange i det daglege nok greidde seg med den engelskprega fagsjargongen som dei utanlandske selskapa hadde hatt med seg, blei det sett på som eit problem at engelskkunnskapane blant arbeidarar og funksjonærar var svært vekslande. Og endeleg var fortienesta god fram til oljekrisa i 1986, noko som gjorde at dei store selskapa kunne ta seg råd til å investere i terminologi.⁷⁷ Også kulturpolitiske omsyn er med i bakgrunnen for prosjektet; norske arbeidstakarar måtte kunne bruke sitt eige morsmål også i oljebransjen, og norsk kunne ikkje tape for engelsk på eit så viktig økonomisk felt som petroleumsverksemda var blitt på 1980-talet. Eit visst innslag av språkleg purisme kan også sporast i kjeldene. Til dømes blir blandingsspråket av engelsk og norsk som hadde utvikla seg, kalla «bastardspråk» i sluttrapporten for Terminol-prosjektet.⁷⁸

For PDS kom desse prosjekta på eit ideelt tidspunkt. Det var utvikla programvare for terminologiarbeid som var prøvd ut på mindre prosjekt som Norsk dataordbok.⁷⁹ No blei det mogleg å prøve ut og vidareutvikle denne kompetansen i eit heilt anna omfang. I 1986 sysselsette Norsk termbank 18–19 personar, og i sluttrapporten frå Terminol-prosjektet omfattar lista over medarbeidarar som i lengre eller kortare tid var med, 17 personar.⁸⁰ Resultatet var ein database med 10 000 hovudoppslag.⁸¹ Denne var eit resultat av eit utstrekt samarbeid mellom terminologane ved termbanken og fagfolka i Statoil. Andre selskap som følgde Statoils eksempel, var Mobil, Elf Aquitaine og Esso.⁸²

Toppåret var 1986, men det oljerelaterte terminologiarbeidet varte i fleire år etter dette.⁸³ Eit eksempel er IGU-prosjektet (1990–1997), som utvikla ein norsk gassterminologi for Den internasjonale gassunionens ordbok, *Multilingual Dictionary of the Gas*

77 Pedersen 1994.

78 Roald, Myking & Pedersen 1986.

79 Dette var det første prosjektet der edb blei tatt i bruk i terminologiarbeid. Arbeidet starta i 1972, då Norsk språkråd sette ned ein komité for dataterminologi, og den først utgåva kom i 1976. I forordet skriv komiteen at boka er lagra på eit datamaskinlesbart medium ved PDS, og at ein her har utvikla program og stått for tekstbehandling og redigering. Boka kom i fleire reviderte utgåver fram til 1997. Men PDS fall etter kvart ut som partnar. I fjerde utgåva, frå 1987, er ikkje UiB nemnt.

80 Utne 1987.

81 Roald, Myking & Pedersen 1986.

82 Pedersen 1994, Jonassen 1986.

83 Myking 2005.

Industry.⁸⁴ Men aktiviteten ved termbanken blei på 1990-talet også meir orientert mot verksemder på land, m.a. blei det utarbeidd ei norsk utgåve av WHO's internasjonale sjukdomsklassifikasjon for Sosial- og helsedirektoratet.⁸⁵

Allereie på 1980-talet blei termsamlingane ved Norsk termbank samla i den såkalla NOT-basen, den blei kommersialisert, først av Nordisk institutt som eigar. Det viste seg etter kvart at det baud på rekneskapsmessige problem å integrere instituttøkonomi med kommersiell verksemd, så i 1988 blei Norsk termbank og dei som var tilsette der, overførte til Unifob. Dette skjedde berre få år etter at ein hadde freista å få universitetet med på ei konsolidering og utbygging av fagmiljøet ved instituttet, som fagleg basis for den kommersielle verksemda.⁸⁶ Dette blei det ikkje noko av, utover at PDS i 1986 skifta namn til Avdeling for norsk leksikografi. Då termbanken blei overført til Unifob i 1988, blei avdelinga lagd ned. Ved Unifob blei drifta basert på eksterne oppdragsinntekter. Så seint som i 1998 blei basen brukt i dokumentproduksjonen i oljenæringa. I 1999 kom ein internettversjon.

Kontakten med fagmiljøet ved instituttet blei organisert slik at ein av dei vitskapleg tilsette fekk frådrag i undervisningsplikta mot å bruke ein arbeidsdag i veka på arbeid knytt til termbanken. I 2000 blei ordninga med frådrag for undervisningsplikt avvikla, og då stansa arbeidet med og vidareutviklinga av basen opp.⁸⁷ Sidan ansvaret for basen no låg hos Unifob, såg korkje HF-fakultetet eller Nordisk institutt det som si oppgåve å koste vidareføring og oppdatering. Basen var først og fremst laga som eit praktisk verktøy for næringsliv og offentleg administrasjon, ikkje som ein forskingsressurs. Etterspurnaden frå oljebransjen var tidsavgrensa og forankra i det overordna behovet for å integrere petroleumsverksemda økonomisk, sosialt og kulturelt i det norske samfunnet. Den var med andre ord særeigen, og då dette målet var nådd, fanst det ingen andre samfunnsområde med tilsvarende behov.

84 Ellingsve 1999.

85 WHO 2005.

86 Privatarkiv-OJJ: Notat om terminologiarbeidet ved Nordisk institutt datert 4.1.1985, signert av fem vitskapleg tilsette ved instituttet.

87 Andersen 1997, Myking 2005. Samtale med Johan Myking 29.1.2021. I dag er NOT-basen ein del av Termportalen ved UiB: «This termbase is to be considered a historical resource, and has not been updated for a while». Men den er framleis søkbar: <https://repo.clarino.uib.no/xmlui/handle/11509/75>, vitja 26.10.2020.

Sjølv om Terminol-prosjektet hadde eit praktisk føremål, låg det likevel akademisk fagkompetanse til grunn for arbeidet, og utviklingsarbeidet knytt til petroleumsprosjekta genererte også forskning. I 1995 fekk prosjektet Språkplanlegging, identitet og marknad – Nordsjøen som språkleg-kulturelt laboratorium, med Myking som leiar, finansiering under Noregs forskingsråds program Kultur- og tradisjonsformidlende forskning (KULT). Her stod ein analyse av spenninga mellom kultur/språkplanlegging og marknadskrefter i fokus.⁸⁸ Det blei også forsvart to doktoravhandlingar knytte til terminologimiljøet.⁸⁹ I tillegg har fleire av dei som begynte ved Norsk termbank, markert seg som terminologar seinare, t.d. i RTT så lenge det eksisterte, og i Norges standardiseringsforbund.

Etter tusenårsskiftet låg terminologiarbeid, også på nasjonalt plan, i stor grad nede.⁹⁰ Samstundes blei presset frå engelsk på norsk språk ei viktig sak for Språkrådet. Farane knytte til manglande norskspråkleg terminologi blei diskuterte, og i 2008 kom stortingsmeldinga *Mål og mening – Ein heilskapleg norsk språkpolitikk*.⁹¹ Året etter blei universiteta pålagde å ta ansvaret for å sikre og vidareutvikle norsk fagspråk gjennom eit tillegg til universitets- og høgskulelova. Dei første åra blei dette stort sett ignorert av institusjonane. Først i dei siste åra har verdien av termbasar på nytt kome opp i det offentlege ordskiftet – ein diskusjon som i stor grad er driven fram av Språkrådet og etter kvart statsforvaltninga, ikkje universiteta sjølve.⁹²

UiB tar over ansvaret for nasjonale språksamlingar og ordbøker

I 2016 blei leksikografien vekt til live att ved UiB. I 2014 bestemte UiO seg for å avslutte ordbokssamarbeidet med Språkrådet og seie frå seg ansvaret for dei nasjonale språksamlingane i stadnamn og

88 Myking & Sæbøe 2000 er ein gjennomgang av resultatata frå prosjektet.

89 Andersen 1998 [1991], Myking 2008.

90 Noregs Handelshøgskule og miljøet rundt Magnar Brekke og Ingrid Simonnæs, og seinare Gisle Andersen og Marita Kristiansen, er det einaste vesentlege unntaket eg veit om.

91 St.meld. nr. 35 (2007–2008).

92 Kapittel 9 i meldinga gir ei meir detaljert skildring av denne utviklinga.

målføre. Om ingen andre ville ta over ansvaret, ville ordsamlingane bli pakka ned.⁹³ Dette var ei strategisk avgjerd om heller å bruke ressursane på undervisning og meir spissa språkvitskaplege satsingar enn produksjon av ordbøker for norsk. I fleire av dei andre språkvitskaplege universitetsmiljøa i landet skapte avgjerda uro for at samlingane skulle bli utilgjengelege for forskingsformål. Kulturdepartementet, representert mellom andre av Språkrådet, frykta at det ville stå utan ein vitskapleg partner som kunne vidareutvikle ordbokproduksjonen og sikre dokumentasjonsansvaret for norske stadnamn. I åra 2014 og 2015 blei det derfor arbeidd intensivt med å finne alternative løysingar.

UiB var tidleg ute med å signalisere at ein der var interessert i å overta ansvaret. Initiativet kom frå nordiskmiljøet ved Institutt for lingvistiske, litterære og estetiske studium. Fagmiljøet hadde sjølv ansvaret for vitskaplege samlingar i målføre og stadnamn, og var ottefulle for at utviklinga og forskinga knytt til slike samlingar skulle forvitte nasjonalt. Samstundes voks det fram ein argumentasjon som fekk brei støtte, om at samlingane frå Oslo burde flyttast til eit språkvitskapleg miljø og vere grunnlag for vidare forskning i dei disiplinane som låg til grunn. Ordbokproduksjon i seg sjølv var kanskje ikkje det som stod fremst i tankane i fagmiljøet, men ei nasjonal ordboksgruppe med utgangspunkt i Nynorsk kultursentrum framheva dette som eit nasjonalt viktig kulturpolitisk omsyn. For instituttleiinga i Bergen handla det òg om å ta imot alle dei potensielle vekstimpulsane ein kunne få, i ei tid prega av fagleg nedbygging.

Initiativet frå fagmiljøet blei tatt vidare av Myking som instituttleiar og representant for UiB i den nasjonale ordboksgruppa. Myking fekk støtte frå fakultets- og universitetsleiinga, og det viktigaste arbeidet vart gjort av rektor Dag Rune Olsen i dialog med den politiske leiinga i Kunnskapsdepartementet. I tildelingsbrevet for 2016 frå Kunnskapsdepartementet blei ansvaret for samlingane overført til UiB.⁹⁴ Det lokale ansvaret fekk Språksamlingane, som blei oppretta som avdeling ved Universitetsbiblioteket.⁹⁵

93 Eit godt oversyn over saksgangen ved UiO finst i eit brev frå universitetet til Kulturdepartementet datert 23.2.2015: <https://hf.uio.no/iln/om/organisasjon/styremøter/2013-2016/innkalling-2015/0-sak-3-7-videre-drift-og-forvaltning-av-samlingene.pdf>, henta 17.11.2020.

94 Universitetsstyret, sak 10/2016, Tildelingsbrevet frå Kunnskapsdepartementet, side 8 (18.12.2015).

95 Myking (under utg.) 2022.

Denne saka har lange røter. Samlingsaktivitetane i Oslo var i stor grad bygde opp i einingar som til dels opphavleg var uavhengige, men etter kvart vart institusjonaliserte ved UiO og gjevne nasjonalt ansvar. Mellom desse er Norsk Målførearkiv, Norsk Stadnamnarkiv og Norsk leksikografisk institutt. Frå å vere sjølvstendige einingar var dei frå 2001 innlemma i Institutt for lingvistiske og nordiske studium som forskargrupper. Både Språkrådet og den nasjonale ordbokgruppa foreslo i 2014 og 2015 å danne ein ny, sjølvstendig organisasjon – eit dokumentasjonssenter for norsk språk, eventuelt i sameigarskap mellom universiteta. Den løysinga som vart vald, var likevel i realiteten den historisk etablerte: Ein interessert institusjon med relevant aktivitet får det nasjonale ansvaret med tilhøyrande finansiering.

Samlingane frå Oslo inneheld mykje materiale knytt til namnegransking, målgransking og norrønt, men til no er det først og fremst dei leksikografiske samlingane som har fått eit nytt liv ved UiB. For med desse følgde medansvar (saman med Språkrådet) for dei to handordbøkene, *Bokmålsordboka* og *Nynorskordboka*. Etter overføringa til UiB er det løyvd midlar til full revisjon av dei to. I 2021 var 7,5 årsverk knytte til dette prosjektet, 4,5 faste og 2 mellombelse, pluss delar av pliktarbeidet til ein doktorgradsstipendiat.⁹⁶

Også det faglege ansvaret for storverket *Norsk Ordbok*, utgitt på papir i 12 store bind mellom 1966 og 2016, følgde med over fjellet.⁹⁷ Arbeidet med ordboka starta så tidleg som i 1930. Men allereie på 1920-talet hadde professor Hannaas ved Bergens Museum lagt konkrete planar for ei slik ordbok, med budsjett, tidsplan og kven som skulle vere medarbeidarar under Hannaas sjølv som redaktør.⁹⁸ Men då han døyde i 1929, blei realiseringa av prosjektet flytta til Oslo frå Bergen. Så prosjektet kom på ein måte heim i 2016, og i 2019 løyvde Stortinget midlar til digitalisering og fagleg oppdatering av dei fem første binda, slik at heile verket vil bli søkbart på internettet.⁹⁹ Høgskulen i Volda uttrykte tidleg ei særleg interesse for *Norsk Ordbok*, og har teke medansvar for arbeidet.

96 E-post frå hovudredaktør Margunn Rauset 25.1.2021.

97 Den fulle tittelen er *Norsk Ordbok: ordbok over det norske folkemålet og det nynorske skriftmålet*.

98 Bondevik 1988.

99 <https://www.uib.no/11e/138863/revisjon-og-digitalisering-av-norsk-ordbok-h-no-ah>, vitja 28.1.2021. Prosjektet er eit samarbeid med Høgskulen i Volda, som støtta arbeidet med å få Språksamlingane til Bergen.

I tillegg er det frå 2021 løyvd midlar til å vidareutvikle ein termportal, slik at også terminologiarbeid blir ein del av den fornya satsinga på leksikografi.¹⁰⁰ Slik er UiB i løpet av dei siste åra blitt eit sentrum for ordboksarbeid i Norge, nesten hundre år etter at Hannaas la dei første planane og vel tjue år etter at terminologisatsinga frå 1980-talet ebba ut. Men det som var ein veikskap då, er det også no. I motsetnad til til dømes talemålsgranskinga skildra ovanfor er ikkje ordboksarbeidet (unnateke ei forskarstilling knytt til Termportalen) forankra i faste, akademiske stillingar i form av professorat eller førsteamanuensisstillingar; det same gjeld i endå høgare grad stadnamndisiplinen. Dermed er sambandet med kjerneverksemda ved universitetet usikkert og potensielt ustabil.

Det er likevel ein skilnad: Denne gongen er verksemda finansiert av staten og lagt til universitetet, ikkje til den såkalla randsonen der Unifob høyrde til. Det gir universitetet eit større ansvar for kontinuiteten, og det er ein styrke. Det er tale om eit strategisk val som skil seg frå det valet UiO gjorde: I meldingsdelen til framlegget om språklov er det lagt stor vekt på at UiB er blitt det nasjonale senteret i Noreg for ordboksarbeid, og dette er definert som ein viktig del av staten si satsing på språkleg infrastruktur.¹⁰¹ Den nye satsinga på Termportalen viser at dette har fått konkrete følgjer. Det eksisterer nok òg hjå mange ei forventning om at flyttinga av språksamlingane skal føre til ei revitalisering av stadnamngransking som universitetsdisiplin.

Datalingvistikk og språkteknologi

Dei to utviklingslinene vi har sett på til nå, var begge knytte til Nordisk institutt. Ei tredje er knytt til allmenn språkvitenskap eller lingvistikk.

100 UiB <https://www.uib.no/ub/fagressurser/spesialsamlingene/121707/termportalen>, vitja 28.1.2021.

101 Prop. 108 L (2019–2020) Lov om språk (språklova), kapittel 9.

Etableringa av allmenn språkvitenskap ved UiB

Faget går tilbake til Fonetisk institutt, som blei oppretta i 1953 og som då fekk ansvaret for forprøvene i fonetikk og språkvitenskap. Martin Kloster-Jensen blei tilsett i eit dosentur i fonetikk same året, og etter kvart blei fleire stillingar i fonetikk oppretta for å ta seg av det auka undervisingsbehovet ut over 1960-talet. Allmenn språkvitenskap var først berre tilbydd gjennom førelesingar av professor Bjarne Ulvestad (1922–2004) frå Germanistisk institutt.¹⁰² Han blei tilsett som professor i 1957, same året som Chomsky la fram sin generative grammatikk i boka *Syntactic Structures*. Fram til 1970-talet var Ulvestad den fremste representanten for strukturalismen blant dei språkvitenskapleg tilsette ved fakultetet. Først i 1972 blei det tilsett ein allmennlingvist ved Fonetisk institutt: Marianne Haslev Skånland (1942–). Ho fekk følge av Henry Henne i 1981 og Helge Dyvik (1947–) i 1983.¹⁰³ Samstundes blei talet på fonetikarar redusert, og namnet på instituttet blei i 1982 endra til Institutt for fonetikk og lingvistikk.

I 1972 skulle ein tru at tida var moden for den generative grammatikken også ved UiB. Men som alt nemnt etablerte ikkje Chomskys syn seg som ein ny normalvitenskap på same måten som ein kan seie at strukturalismen gjorde det. Også i Bergen møtte teorien motstand, først ved Skånland, som var ein markert og skarp motstandar på tradisjonell, strukturalistisk grunn.¹⁰⁴ I 1980 disputerte Dyvik med ei avhandling som var eit klart oppgjær med Chomskys mentale realisme. Ifølge Dyvik sette han likskapsteikn mellom det formelle beskrivingsapparatet knytt til teorien og den språklege kompetansen sjølv. For Dyvik var dette eit kategorimistak; ein modell kan vere meir eller mindre formålstenleg og innsiktsgjevande i høve til dei fakta den skal gjere greie for, men han er ikkje dermed ei spegling av forskingsobjektet. Å gjere greie for innhaldet i den menneskelege språkkunnskapen er difor ikkje utan

102 Thue 1996: 542.

103 Sst.: 56f.

104 Sjø t.d. slutten på forordet til [Haslev] Skånland 1985: «Innen fonologien de siste 20 årene har dette særlig vært merkbart i en skoledannelse som kaller seg 'generativ fonologi', og som har spesialisert seg på å generere kjente fenomener og innsikter på særs tungvinte måter og på å slå barnet ut med badevannet. Ubefestede sjeler er hermed advart mot å kaste bort tiden med [å] oppfinne kruttet på nytt.»

vidare det same som å avdekke dei psykologisk-genetiske mekanis-
mane som ligg til grunn for sjølve den menneskelege språkevna.
Dyvik hevda at hypotesane til Chomsky ikkje kan testast i poppersk
forstand, og difor ikkje kan tene som empirisk forankringspunkt
for språkvitskapen, Dyvik tar utgangspunkt i eit alternativt foran-
kringspunkt: «observerbar språkbruk».¹⁰⁵

Datalingvistikk og datastøtta lingvistikk

I 1984 publiserte Dyvik ein artikkel med tittelen «Hva er dataling-
vistikk?». Den kan sjåast som starten på det som ikkje berre skulle
bli hovudinteressa gjennom resten av den akademiske karrieren
hans, men òg ei interesse som kom til å prege den faglege profilen til
instituttet. I artikkelen blir datalingvistikk definert som «å utvikle
programvare for grunnforskningsformål eller for direkte praktisk
anvendelse, som *inkorporerer lingvistisk innsikt*, eller, om man vil,
simulerer språklig kompetanse, på ett eller annet nivå».¹⁰⁶

I artikkelen trekker Dyvik eit skilje mellom datalingvistikk
og datastøtta lingvistikk. Det siste refererer til forskning der bruk av
datateknologi har gjort forskinga meir effektiv på ulike måtar, utan
at bruken har endra dei grunnleggande metodane eller forskings-
spørsmåla. Både Talebanken og Norsk termbank er eksempel på
slik forskning, der digitalisering og databaseteknologi har bidratt
til at analysen av data og etablering av samanhengar i materialet
er sikrere og meir effektiv samtidig som mengda av data som kan
lagrast, strukturert og analyserast, har auka enormt. Her skil ikkje
språkvitskap seg frå andre vitskapar.

Datalingvistikk er òg eit nytt forskingsfelt som oppstod med
datamaskina, der ein simulerer språkleg kompetanse og språkbruk
ved hjelp av datateknologi. Datalingvistikk har vore eit sentralt
forskingsstema ved lingvistikkseksjonen ved UiB heilt frå 1980-talet,
men det har ikkje vore det einaste. Det er likevel dette vi skal kon-
sentrere oss om her.

105 Dyvik 1980: kap. 2. Den som representerte den chomskyanske lingvistikken ved institut-
tet, var Kirsti Koch Christensen (1940–), som etterfølgde Henry Henne i 1991. Christen-
sen blei prorektor i 1996 og var rektor frå 1999 til 2005. Dette la naturlegvis avgrensingar
på den faglege produksjonen hennar.

106 Dyvik 1984.

Røtene

I ei usignert oversikt over verksemda ved PDS frå september 1981 blir det under overskrifta *Norsk tekstbehandling* vist til to prosjekt – eit delvis realisert og eit som kunne vere aktuelt. Det første er eit automatisk system for orddeling, som allereie skal vere på plass. Det andre er eit system for automatisk korrekturlesing, som «[t]ruleg vil bli eit interessant område innanfor datamaskinell språkbehandling».¹⁰⁷ Også datastøtta omsetting mellom bokmål og nynorsk blei det arbeidd med noko seinare på 1980-talet.¹⁰⁸ Men då etterfølgaren til PDS, Avdeling for norsk leksikografi, blei lagd ned i 1988 og Norsk termbank samstundes blei overført til Unifob, var det slutt på det språkteknologiske utviklingsarbeidet ved Nordisk institutt.

Ved NAVFs EDB-senter blei den tidlege interessa også retta mot grunnforskning og korleis ein kunne utvikle grammatikkar som baserte på språkvitskapleg teori kan analysere språklege utsegner på eit gitt språk. Det begynte med eit samarbeid om setningsanalyse med dåverande amanuensis Svein Lie ved UiO på slutten av 1970-talet.¹⁰⁹ Etter dette var Per Kristian Halvorsen frå januar 1982 til sommaren 1983 tilsett som forskar ved senteret. Han arbeidde med å utvikle eit system for automatisk grammatisk analyse med spesiell tillemping på norsk. Med utgangspunkt i ein teori kalla leksikalsk-funksjonell grammatikk (LFG) la han grunnlaget for eit langvarig og fruktbart samarbeid med eit forskingssenter i Palo Alto i California, eigd av Xerox-konsernet.¹¹⁰ LFG er også ein generativ modell, men utan den abstrakte underliggende og ikkje-observerbare djupstrukturen den chomskyanske retninga la til grunn. Den bygger på observerbar språkbruk.

Etter at Halvorsen slutta, blei det datalingvistiske programmet vidareført som eit samarbeid mellom EDB-senteret og Dyvik. Artikkelen hans i meldingsbladet frå senteret i 1984 kan lesast som ei programerklæring. Det blei lagt planar for studietilbod og utvikling av eit program for datalingvistisk forskning. Maskinvare spesielt godt

107 UBB-Ms 2175, nr. 16d, s. 4.

108 Årsmelding 1988: 5 flg.

109 Lie 1980.

110 Årsmelding 1983: 5 flg., Hauge 1983.

eigna for datalingvistisk forskning blei kjøpt inn, og ikkje minst blei samarbeidet med Xerox-senteret vidareført.¹¹¹

Maskinomsetting

Målet om å utvikle automatisk omsetting frå eitt språk til eit anna oppstod om lag like tidleg som datamaskina. Men dei tidlege forsøka på 1950-talet stranda, og på 1960- og -70-talet rådde pessimismen. Det var liten tru på at vegen til adekvate omsettingssystem kunne gå via språkvitskapleg teori og språkvitskaplege beskrivingsapparat.¹¹² I artikkelen frå 1984 viste Dyvik til ny og meir realistisk optimisme. Nyare grammatiske modellar, som t.d. LFG, var eit lovande utgangspunkt for datalingvistisk forskning og dermed også for m.a. å utvikle brukbare omsettingssystem.

Frå dette utgangspunktet har utviklinga av maskinomsetting følgt to hovudvegar, som ikkje kan sjåast som heilt åtskilte, men som snarare utfyller kvarandre. Den eine er den det er tale om her: omsetting basert på språkvitskapleg kunnskap. Her prøver ein å simulere den menneskelege språkevna ved å skrive program som modellerer dei grammatiske reglane vi bruker når vi produserer ytringar, og å bygge leksikon som modellerer ordforrådet vårt. Ei hovudutfordring er at språk er ulike når det gjeld begge modulane – det same innhaldet kan vere koda ulikt både i ordforråd og grammatikk. Eit døme på det første er at det norske ordet *låne* dekker to ord i engelsk: *lend* og *borrow*. Eit døme på ulik grammatikk er kravet på norsk om at det bøygde verbet må stå på andre plass i setninga, noko det ikkje gjer i t.d. engelsk, jamfør norsk *Då køyrd Per til byen* mot engelsk *Then Peter drove to town*.

Dette gjer lingvistisk basert omsettingsforskning til grunnforskning – dess betre eit program verkar, dess meir kunnskap har ein fått fram om kva slags formelle og semantiske relasjonar som finst mellom språka som inngår i omsettingsrelasjonen. Kvaliteten på omsettinga blir eit suksesskriterium, men den språkvitskaplege kunnskapsutvidinga som har ført fram til målet, er like viktig som resultatet.

111 Årsmelding 1984: 11.

112 Dyvik 1990: 2 flg.

Den alternative tilnærminga er statistisk basert omsetting. Denne tar utgangspunkt i store såkalla parallellkorpus for å berekne sannsynet for at ein ordsekvens eller ei frase i eit språk vil svare til ein bestemt ordsekvens eller frase i det andre språket. (Parallellkorpus er eit korpus som inneheld versjonar av same teksten eller dei same tekstane på meir enn eitt språk.) I den reine typen av slike modellar finst ingen grammatikk og ikkje noko leksikon, berre korrespondansar mellom parallellkorpus. Her bidrar med andre ord ikkje programmert kunnskap om grammatiske reglar til å nå fram til målet om ei best mogleg omsetting.¹¹³

Som nemnt er ikkje dette to åtskilte spor. I dag vil dei fleste prosjekta ha innslag av begge, og den reint statistisk baserte metoden er avløyst av det ein kallar nevralt nettverk, dvs. datastrukturar som er inspirerte av måten nervecellene i hjernen er organiserte på. Men om vi held oss til Dyviks definisjon, «å simulere språkleg kompetanse», må den regelbaserte strategien stå sentralt innanfor datalingvistikk forstått på denne måten.

Forskinga ved instituttet kom i åra som følgde etter 1984, i stor grad til å vere prega av datalingvistikk, og eit omsetningsperspektiv var heile tida med i større eller mindre grad. Dette perspektivet vert brukt i det vidare. Det er likevel viktig å peike på at det først og fremst dreier seg om grunnforskning, med potensial for praktisk utnytting. Eit eksempel er det første prosjektet der ein fekk finansiering frå Forskringsrådet, PONS, eit akronym for «Partiell Oversettelse mellom Nærstående Språk», der dei nærstående språka var norsk og svensk. Målet var ikkje å utvikle eit dataprogram som kunne setje om mellom norsk og svensk i stort omfang, men å utvikle eit overordna system som automatisk kunne utnytte likskapen mellom to språk i omsetningsprosessen.¹¹⁴

Omsetningsperspektivet har også prega seinare prosjekt, som prosjektet «From Parallel Corpus to Wordnet», leia av Dyvik (2001–2004), som utvikla ein metode for automatisk avleiing av ordsemantiske skildringar frå omsettingar, og det nasjonale LOGON-prosjektet (2002–2006). Dette siste prosjektet dreidde seg om omsetting frå norsk til engelsk.

113 Dette er ei svært forenkla framstilling. Sjå t.d. https://snl.no/statistisk_maskinoversettelse, vitja 11.2.2021.

114 Dyvik 1994: 226.

Her kom samarbeidet knytt til Xerox-senteret i California til å bere frukter, det var allereie etablert tidleg på 1980-talet, og rundt tusenårsskiftet blei norsk med i «The Parallel Grammar Project» ved senteret, der målet var å utvikle parallelle grammatikkar for ei rekke språk.¹¹⁵ Grammatikken for norsk blei lagt til grunn for to prosjekt der eit viktig delmål i det siste var å bygge ein såkalla trebank for norsk. Ein trebank er eit korpus der setningane ikkje berre er analyserte på ordnivå, men også med omsyn til syntaks. Dette vil seie at kvar setning i korpuset er blitt tilordna ein analyse av dei større frasene orda i setninga dannar, og kva slags funksjon desse frasene fyller i setninga, som t.d. subjekt, verbal og objekt.¹¹⁶

Det sentrale delmålet i det første av dei to trebankprosjekta, «The Norwegian Treebank Pilot Project» (TREPIL, 2004–2008), var å utvikle ei generell verktøykasse for bygging av trebankar.¹¹⁷ Prosjektleiari var professor Koenraad de Smedt (1954–), som saman med emerita Victoria Rosén (1951–), senioringeniør Paul Meurer og Helge Dyvik og har vore kjernen i datalingvistikkmiljøet. Med utgangspunkt i desse verktøya hadde det neste prosjektet, «The Norwegian Infrastructure for the Exploration of Syntax and Semantics» (INESS 2010–2017), to mål. Under leiing av Rosén var det første å skape ein internasjonal infrastruktur for lagring og bruk av trebankar, for forskning så vel som teknologiutvikling. Denne inneheld i dag meir enn 800 trebankar som representerer ei lang rekke ulike språk.¹¹⁸ Prosjektet hadde ikkje omsettingsteknologi og metode som eit hovudmål, men i INESS-infrastrukturen finn ein ei rekke parallellkorpus, og dei er heilt sentrale for å bygge gode omsettingsprogram baserte på lingvistisk kunnskap.¹¹⁹

Det andre målet var å bygge ein norsk trebank, NorGramBank, ved hjelp av den norske grammatikken utvikla under ParGram-para-

115 Butt, Dyvik, King, Masuichi & Rohrer 2002.

116 Namnet «trebank» kjem av at syntaktiske analysar ofte blir visualiserte i form av såkalla trestrukturar.

117 Rosén, Meurer & De Smedt 2009.

118 <http://clarino.uib.no/iness>, vitja 22.2.2022.

119 INESS er del av Clarino-infrastrukturen, den norske delen av Clarin, ein europeisk infrastruktur for språkressursar (<https://www.clarin.eu>, vitja 3.3.2021). The Clarino Bergen Centre blir leia av Koenraad de Smedt, som også er nasjonal koordinator og den som tok initiativet til norsk deltaking i prosjektet (<https://repo.clarino.uib.no/xmlui/>, vitja 3.3.2021). Clarino har fått finansiering frå NFR, sist i 2021.

plyen. Korpuset inneheldt i 2020 ca. 160 millionar ord.¹²⁰ Tekstane er både på bokmål og nynorsk, og skjønnlitteratur og ulike typar faglitteratur er representerte. For dei fleste setningane i korpuset vil grammatikken føreslå meir enn éin mogleg struktur, ofte svært mange. Det ideelle vil då vere at eit menneske avgjer kva som er det eller dei korrekte av desse forslaga. Men når tekstmengda blir så stor som fleire millionar ord, er ikkje dette praktisk mogleg utan ein urealistisk stor innsats av tid og pengar. Løysinga blei difor at ein mindre del av korpuset, 350 000 ord, «gullstandarden», først blei analysert ved hjelp av grammatikken. Analysane blei så kontrollerte manuelt, og den rette analysen blei valt der grammatikken hadde foreslått fleire moglege. Med utgangspunkt i denne standarden blir resten av det maskinelt analyserte korpuset eintydiggjort på statistisk grunnlag, basert på kor sannsynleg ein analyse er når den blir halden opp mot gullstandarden. Dette vil ikkje gi eit 100 prosent korrekt analysert korpus, men ein kontroll av 500 tilfeldig utvalde setningar viste ein tilfredsstillande grad av samsvar med gullstandarden.¹²¹

Praktisk nytte

Språkteknologiske hjelpemiddel har dei siste åra blitt allemannseige. Vi les kanskje meir på skjerm enn på papir, og vi kan ofte få teksten lesen høgt av maskinen om vi ønsker det. Om teksten er på eit språk vi ikkje forstår, kan vi bruke omsettingsprogram. Når vi skriv, får vi hjelp av korrekturprogram som held styr på ortografi og enkel grammatikk. Etter kvart kan vi også styre m.a. såkalla smarthøgtalarar med stemma. Ikkje alle fungerer like bra, men nye versjonar kjem stadig, med forbetra funksjonalitet.

Har forskinga ved UiB bidratt til denne utviklinga? I beste fall er det eit indirekte samband mellom akademisk forskning og kommersiell, språkteknologisk produktutvikling, trass i relevansen for maskinomsetting. Det kan minst tenkast fire typar samband mellom relevant, akademisk (grunn)forskning og utnytting av resultatata til meir praktiske føremål.

120 <https://clarino.uib.no/iness/page?page-id=iness-descr> og <https://nogramtall.w.uib.no/vitja.11.11.2020>. Ei meir teknisk skildring av banken finn ein i Dyvik mfl. 2016.

121 Sst.

1. Kommersialisering av forskingsresultat gjennom såkalla teknologioverføring, der universitetstilsette engasjerer seg som gründerar, ofte i samarbeid med personar med kompetanse på forretningsdrift.
2. Overgang frå universitetsstilling til å bli arbeidstakar i teknologibedrift.
3. Utdanning av kandidatar på ulike nivå som har den språkvitenskaplege kompetansen utviklingsmiljøa treng.
4. Sal eller deling av forskingsresultat eller -ressursar som kan utnyttast av andre.

Eit kanskje noko marginalt eksempel på det første er Scarrie-prosjektet (1996–1999). Her samarbeidde forskarar frå ei rekke akademiske institusjonar i Norden, frå UiB Rosén og De Smedt, med eit kommersielt firma i Sverige, Word Finder Software, om utvikling av eit korrekturprogram. For norsk var den store ortografiske valfridomen ei stor utfordring. Resultata var gode, men dei blei ikkje kommersialiserte.¹²²

Når det gjeld det andre, fekk den siste fonetikaren ved instituttet, Kolbjørn Slethei, permisjon frå stillinga si for å arbeide for firmaet Nordisk språkteknologi då dette blei oppretta på Voss i 1997, med utvikling av taleteknologi som den viktigaste forretningsideen. Då firmaet gjekk over ende i 2003, vende han attende til universitetet. Det var elles lite forskingskontakt mellom dette firmaet og det datalingvistiske miljøet ved UiB, som ikkje hadde taleteknologi som eit viktig forskingstema. Då firmaet blei oppretta, bidrog UiB med å opprette eit grunnfagstilbod i språkteknologi på Voss, og det blei utdanna kandidatar i faget i Bergen. Til det tredje, utdanning er kanskje det viktigaste bidraget til språkteknologisk produktutvikling, ikkje berre ved UiB, men ved alle høgre utdanningsinstitusjonar som har datalingvistik som fag. Når det gjeld sal eller deling av forskingsresultat, var det mot slutten av PDS si tid tale om å utvikle og kommersialisere eit korrekturprogram.¹²³ Så vidt eg veit, blei ikkje dette noko av, og heller ikkje Heggstads firma Logos A/S blei nokon suksess.

Ein rapport frå 2012 om situasjonen for norsk språkteknologi

¹²² <https://ling.w.uib.no/projects/scarrie/>, vitja 3.3.2021.

¹²³ UBB-Ms 2175, nr. 16p.

peikar på mangelen på basisressursar og -verktøy og på behov for meir grunnforskning og for samordna formidling og utveksling av forskingsresultat.¹²⁴ INESS og NorGramBank er slike ressursar utvikla gjennom grunnforskning, og dei er ope tilgjengelege. Bruk av NorGramBank i revisjonen av Nynorskordboka og Bokmålsordboka og vidareutviklinga av Det Norske Akademis ordbok er eksempel på korleis dei har kome til nytte.¹²⁵ Når ein kan søke på alle førekomstane av eit ord og få dei fram i den setningssamanhengen dei står i, blir det mykje enklare å få oversikt over dei syntaktiske og semantiske eigenskapane ordet har.

I dette kapitlet er det vist korleis den digitale vendinga har bidratt til å fremme og tilgjengeleggjere språkvitskapleg forskning ved UiB. Dei tre fagmiljøa presenterte ovanfor er døme. Fleire fagmiljø kunne ha vore nemnt, mellom anna den store databasen over norrøne manuskript, Menota, og det såkalla ASK-korpuset, eit innlærarkorpus som består av eksamenssvar frå norsk kurs for innvandrarar med ulik språkbakgrunn.¹²⁶ Også digitaliseringa av dei store stadnamnsamlingane kunne ha vore tatt med.¹²⁷ Eit siste døme er dei store talekorpusa med engelsk og spansk tenåringspråk – COLT- og COLA-prosjekta.¹²⁸

Kjelder

UBB-MS 2175. Universitetsbibliotekets spesialsamlingar. *Program for datamaskinell språkbehandling (PDS)*. Interne dokument leverte til samlingane av Ivar Utne i eit faghistorisk dokumentasjonsprosjekt i 2009. Oversynet i dokumentet er på 16 hovudnummer, der det siste, nr. 16, består av 16 kulepunkt. Dokumenta desse viser til, blir refererte til ved hjelp av bokstavkodar i tillegg til nummeret.¹²⁹

(14) Om allmennspråklige ordlister og muligheter for framtidige engasjementer. Prosjekt for datamaskinell språkbehandling, Nordisk institutt, UiB. Udatert, ca. 1985.

(16d) Notat: Verksemda ved PDS, september 1981.

(16j) Referat frå møtet om ordlister og Norsk ordregistrant, 25.4.1984.

(16p) Notat datert 16.9.87 frå Avdeling for norsk leksikologi (tidl. PDS) til Norsk Regnesentral som diskuterer korleis ordregistranten kan brukast til utvikling av eit retteprogram.

124 De Smedt, Lyse, Gjesdal & Losnegaard 2012: 32.

125 Det Norske Akademis ordbok.

126 <https://www.menota.org>, vitja 23.2.2021, Tenfjord, Hagen & Johansen 2009, Ragnhildstveit 2018.

127 <https://www.uib.no/11e/25429/stadnamnsamlinga>, vitja 23.2.2021.

128 Stenström & Jørgensen 2009, Stenström 2014.

129 Garnes & Søndena 2009: 89 flg.

Privatarkiv-OJJ. Interne dokument knytte til PDS og termbanken, særleg til omlegginga midt på 1980-talet, tatt vare på av Ole-Jørgen Johannessen, som då var leiar av fagrådet for PDS.

Litteratur

- Akselberg, G. (red.) (1998). Oddvar Nes. *Ord etter ord. Heiderskrift til Oddvar Nes*. Bergen: Norsk Bokreidingslag L/L. 7–10.
- Alhaug, G. (1978). *Norsk ordregistrant – Status i mars 1978*. Universitetsbibliotekets spesialsamlinger, UBB-Ms 2175, nr. 7. Universitetet i Bergen.
- Andersen, Ø. (1997). NOT-prosjektet: Norsk terminologisk database – Struktur og applikasjoner. I: H. Hjulstad (red.), *Nordterm '97. Guovdageaidnu 24–27. juni 1997* (s. 61–65). Oslo: Nordterm.
- Andersen, Ø. (1998 [1991]). *Lingvistikk og dokumentasjonsbaserte tesauruskonstruksjoner* (Doktoravhandling). Universitetet i Bergen.
- Beito, O.T. & Hoff, I. (red.) (1973). *Frå norsk målføregransking: Utvalde utgreingar 1908–1969*. Oslo: Universitetsforlaget.
- Bloch, B. & Trager, G.L. (1942). *Outline of linguistic analysis*. Baltimore, MD: Linguistic Society of America at the Waverly Press.
- Bokmålsordboka*. Bergen: Språkrådet og Universitetet i Bergen. Hentet fra <http://ordbok.uib.no>
- Bondevik, J. (1987). Gransking av norsk talemål i vår tid. *Eigenproduksjon*, 31, 3–31.
- Bondevik, J. (1988). Torleiv Hannaas. Professor 1/7 1918 – 18/11 1929. I: J. Bondevik (Red.), *Frå smått til stort. Nordistikk i Bergen gjennom 70 år* (s. 7–45). Bergen: Nordisk institutt.
- Bondevik, J. & Nes, O. (1988). Gustav Indrebø. Professor 1/7 1930 – 3/8 1942. I: J. Bondevik (red.), *Frå smått til stort. Nordistikk i Bergen gjennom 70 år* (s. 51–79). Bergen: Nordisk institutt.
- Butt, M., Dyvik, H., King, T.H., Masuichi, H. & Rohrer, C. (2002). The Parallel Grammar Project. I: *COLING-GEE '02: Proceedings of the 2002 workshop on Grammar engineering and evaluation*. Stroudsburg, PA: Association for Computational Linguistics. Hentet fra <https://www.aclweb.org/anthology/volumes/W02-15/>.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1959). A review of B.F. Skinner's Verbal Behavior. *Language*, 35(1), 26–58.
- De Smedt, K., Lyse, G.L., Gjesdal, A.M. & Losnegaard, G.S. (2012). *The Norwegian language in the digital age / Norsk i den digitale tidsalderen. Bokmålsversjonen*. Meta-Net hvitbokserie. Springer. Henta frå <http://www.meta-net.eu/whitepapers/e-book/norwegian-bokmaal.pdf>
- Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur. Henta frå <https://www.naob.no>
- Dyvik, H. (1980). *Grammatikk og empiri. En syntaktisk modell og dens forutsetninger* (Doktoravhandling). Bergen: Universitetet i Bergen.
- Dyvik, H. (1984). Hva er datalingvistikk? *Humanistiske data*, 3, 4–11.
- Dyvik, H. (1990). *The PONS Project: Features of a Translation System*. Institutt for fonetikk og lingvistikk, Universitetet i Bergen.
- Dyvik, H. (1994). Exploiting structural similarities in machine translation. *Computers and the Humanities*, 28(4/5), 225–234.
- Dyvik, H., Meurer, P., Rosén, V., Smedt, K.D., Haugereid, P., Losnegaard, G.S., Lyse, G.L. & Thunes, M. (2016). NorGramBank: A 'Deep' Treebank for Norwegian. I: Calzolari, K. Choukri, Thierry Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (red.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA).
- Ellingsve, E.J. (1999). *IGU-prosjektet 1990–1997*. HIT-senteret, rapportserie 1/1999. Bergen: Forskningsprogram for humanistisk informasjonsteknologi, HIT-senteret.
- Fintoft, K. & Mjaavatn, P.E. (1980). *Språksosiologiske forhold i Trondheim bymål*. Trondheim: Tapir.
- Flydal, E. (1983). *Oljespråk. Språklige lovbrudd og lovlige språkbrudd på sokkelen*. Oslo: Universitetsforlaget.
- Forland, A. (1996). Universitetet i Bergens historie 1946–1996. I: A. Forland & A. Haaland. *Universitetet i Bergens historie*. Bind I (s. 190–546). Universitetet i Bergen.

- Gabrielsen, F. (1983). *Eg eller je? Ei sosiolingvistisk gransking av yngre mål i Stavanger* (Hovudoppgåve i nordisk språkvitskap). Universitetet i Bergen.
- Garnes, K. & Søndena, O. (2009). *Prosjektrapport. Faghistorisk dokumentasjonsprosjekt*. Universitetet i Bergen.
- Grundt, L.O. (1991). *Stor norsk-fransk ordbok*. Oslo: Universitetsforlaget.
- Hangeland, S. (1981). *Sosiolingvistiske forhold i Kristiansand bymål* (Hovudoppgåve i nordisk språkvitskap). Universitetet i Bergen.
- Hanssen, E., Hoel, T., Jahr, E.H., Rekdal, O. & Wiggen, G. (1978). *Oslomål. Prosjektbeskrivelse og syntaktisk analyse av oslomål med henblikk på sosiale skilnader*. Oslo: Novus.
- Hauge, J.H. (1983). Automatisk analyse av norsk. Intervju med Per-Kristian Halvorsen. *Humanistiske data*, 1, 51–56.
- Haugen, E. (1965). *Norsk engelsk ordbok*. Oslo: Universitetsforlaget.
- Haugen, R. (2004). *Språk og språkbaldningar hjå ungdomar i Sogndal* (Doktoravhandling). Universitetet i Bergen.
- Heggstad, K. (1968). Perspektiv for norsk humanistisk forskning med moderne databehandling. *Forskningsnytt*, 13(2), 25.
- Heggstad, K. (1970). *Prosjekt for datamaskinell språkbehandling. Eit oversyn*. Nordisk institutt, Universitetet i Bergen.
- Heggstad, K. (1976). *Datamaskinell språkbehandling: PDS 1967–1976*. Nordisk institutt, Universitetet i Bergen.
- Helleland, B. (1997, 12. august). Stadnamna og Bergen. Kronikk, *Bergens Tidende*, s. 22.
- Hernes, R. (2006). *Talemål i endring? Ein longitudinell studie av talemålsutvikling og språkleg røyndomsoppfatning hjå ungdomar i Os* (Doktoravhandling). Bergen: Universitetet i Bergen.
- Hjulstad, H. (1980). Norsk termbank. *Humanistiske data*, 1, 40–44.
- Hjulstad, H. (1985). Datamaskina i terminologiarbeidet. I: B. Norevik (red.), *Datamaskinen og språket* (s. 57–62). Oslo: Universitetsforlaget.
- Hjulstad, H. & Norevik, B. (1984). *Norsk teknisk fagordbok*. Oslo: Universitetsforlaget.
- Hustad, T. (1979). *Stor tysk-norsk ordbok*. Oslo: Universitetsforlaget.
- Indrebo, G. (1937). *W. F. K. Christie: Norsk Dialect-Lexicon og nokre folkeminne og brev*. Bergens Museums Årbok. Historisk-antikvarisk rekke 1. Bergen: Bergens Museum.
- Johannessen, O.-J. (2009). Stedsnavnene tilbake til folket. På veg mot det snakkende navnekartet utlagt på internett. A. Aune & G. Valderhaug (red.), *Undervegs: festskrift til Egil Nysæter* (s. 130–136). Oslo: ABM-Media.
- Johannessen, O.-J. (2012). Digitale stedsnavnbaser ved Universitetet i Bergen. I: B. Helleland (red.), *Stadnamn og stadnamninsamling med bakgrunn i eit pilotprosjekt: Ein rapport frå fagnmiljøet* (s. 20–23). Oslo/Bergen: Samarbeidsnemnda for namnegransking.
- Johannessen, O.-J. & Schmidt, T. (2004). NFR-prosjekt om dataføring av stedsnavnsamlinger. *Norsk namnelag: Nytt om namn*, 40, 23–24.
- Johannessen, S.H. (1983). Om 'sjendisar' og 'chipsreiarar'. Bruken av sje-lyd og kje-lyd i bergensmålet. *Talemål i Bergen*, 1, 5–28.
- Jonassen, K.J. (1986). *Esso/Odin-prosjektet: Sluttrapport*. Norsk termbank, Avdeling for norsk leksikologi, Universitetet i Bergen.
- Jordåen, R. (2014). *Wilhelm F. K. Christie. Presidenten*. Bergen: Fagbokforlaget.
- Kristoffersen, G. (1978). «Mi bræga' så i Æ'ndal»: *En sosio-fonologisk undersøkelse av sekvenser med vokal+r i talemålet til skolebarn i Arendal* (Hovudoppgåve i nordisk språkvitskap). Universitetet i Bergen.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Larsen, A.B. & Stoltz, G. (1912). *Bergen bymål*. Kristiania: Bymålslaget, i kommisjon hos H. Aschehoug & Co.
- Lie, S. (1980). *Automatisk syntaktisk analyse. Del 1, Grammatikken*. Bergen: NAVFs EDB-senter for humanistisk forskning.
- McEnery, T. & Hardie, A. (2013). The history of corpus linguistics. I: K. Allan (red.), *The Oxford handbook of the history of linguistics* (s. 727–746). Oxford: Oxford University Press.
- Myking, J. (2005). Terminologi i Noreg: Historisk oversyn. I: J. Hoel (red.), *Hvem tar ansvaret for fagterminologien?* (s. 2–15). Oslo: Språkrådet.
- Myking, J. (2008). *Motivasjon som termdanningsprinsipp. Ein teoretisk diskusjon på grunnlag av norsk oljeterminologi* (Doktoravhandling). Vasa: Vasa universitet.

- Myking, J. (under utg., 2022). Frå «språksamlingane» til Språksamlingane – Prosessen og vurderingane. *Maal og Minne*.
- Myking, J. & Sæbøe, R. (2000). Språkplanlegging, identitet og marknad – Nordsjøen som språkleg-kulturelt laboratorium. I: H.W. Andersen, S. Lie & M. Melhus (red.), *KULT – i kulturforskningens tegn* (s. 180–196). Oslo: Pax.
- Myking, J., Dysvik, S. & Hjulstad, H. (2018). Fagspråksarbeid. I: B. Mæhlum (red.), *Norsk språkhistorie II. Praksis* (s. 365–415). Oslo: Novus.
- Nes, O. (1996a). Øystein Frøysadal. I: J. Bondevik, W. Dahl & S.A. Aarnes (red.), *Fagprofilar ved Nordisk institutt, Universitetet i Bergen 1946–1996* (s. 61–65). Bergen: Nordisk institutt, UiB.
- Nes, O. (1996b). Gunnvor Runhovde. I: J. Bondevik, W. Dahl & S.A. Aarnes (red.), *Fagprofilar ved Nordisk institutt, Universitetet i Bergen 1946–1996* (s. 124–131). Nordisk institutt, Universitetet i Bergen.
- Norsk ordbok: ordbok over det norske folkemålet og det nynorske skriftmålet*. Oslo: Det Norske Samlaget. Digital utgåve i-å: <http://no2014.uib.no>.
- Nynorskordboka*. Språkrådet og Universitetet i Bergen. <http://ordbok.uib.no>.
- Ordbok for petroleumsvirksomhet – RTT 43*. (1980). Oslo: Universitetsforlaget.
- Paruit, B.H. (1982). *Illustrated glossary of process equipment. Glossaire illustré des équipements de procédé. Illustrert ordbok for prosessutstyr*. Stavanger: Elf Aquitaine Norge.
- Pedersen, P.-B. (1994). Norsk oljespråk i eit tiårs perspektiv. I: J. Myking, H. Sandøy & I. Utne (red.), *Helsing til Lars Vassenden på 70-årsdagen* (s. 161–170). Nordisk institutt, Universitetet i Bergen.
- Ragnhildstveit, S. (2018). ASK – et elektronisk innlærerkorpus til bruk i andrespråkforskning. I: A.-K.H. Gujord & G.T. Randen (red.), *Norsk som andrespråk – Perspektiver på læring og utvikling* (s. 422–448). Oslo: Cappelen Damm Akademisk.
- Roald, J., Myking, J. & Pedersen, P.-B. (1986). *Terminol. Sluttrapport fra et terminologisk prosjekt*. Norsk termbank, Avdeling for norsk leksikologi, Universitetet i Bergen.
- Robins, R.H. (1997). *A short history of linguistics* (4. utg.). London: Longman.
- Rosén, V., Meurer, P. & De Smedt, K. (2009). LFG Parsebanker: A toolkit for building and searching a treebank as a parsed corpus. I: F. Van Eynde, A. Frank, K. De Smedt & G. Van Noord (red.), *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7)* (s. 127–133). Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap).
- Rundhovde, G. (1964). *Målet i Hamre. Ljodvokster og stutt utsyn over formverk i eit Nordbordlandsmål*. Oslo: Universitetsforlaget.
- Rundhovde, G. (1978). Ein studie i historisk-strukturell målgeografi. I: I. Hoff (red.), *På leit etter ord. Heiderskrift til Inger Frøysset frå medarbeidarar og studentar* (s. 299–334). Oslo: Universitetsforlaget.
- Sandal, R. (1976). *Talesyntaks hos utvalde informantar i Vågsøy* (2. band). (Hovudoppgåve i nordisk språkvitskap). Universitetet i Bergen.
- Sandøy, H. (1983). Kort orientering om prosjektet «Talemål hos ungdom i Bergen». *Talemål i Bergen*, 1, 1–4.
- Sandøy, H. (1984). Edb og talemålsforskning. *Humanistiske data*, 3, 29–37.
- Sandøy, H. (1988). Olai Skulerud – Vitskapsmannen. I: J. Bondevik (red.), *Frå smått til stort. Nordistikk i Bergen gjennom 70 år* (s. 80–112). Nordisk institutt, Universitetet i Bergen.
- Sandøy, H. (2007). Vestlandsk målgransking fram til 1970-talet. I: K.A. Bergsvik, O.-J. Johannessen, K.G. Losnedahl, H.K. Sveen & H.-J. Ågotnes (red.), *Humaniora i nasjonen. Nasjonen i humaniora. Det historisk-filosofiske fakultets bidrag til hundreårsmarkeringen Norge 1905–2005* (s. 23–29). Universitetet i Bergen.
- Sandøy, H. (2013). Perspektiv på norsk talemålsforskning siste 100 år. I: L. Elmevik & E.H. Jahr (red.), *Talemålsforskning i Norden dei siste 100 år: eit symposium i Åseral i 2011 i høve 100-årsjubileet for Didrik Arup Seips hovedfagsavhandling om åsdølmålet*. Uppsala: Kungl. Gustav Adolfs akademien för svensk folkkultur.
- de Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.
- Scott, P., Brusdal, O.B., Heggstad, K., Lium, G. & Løland, S. (red.) (1976). *Norsk dataordbok*. Oslo: Universitetsforlaget.
- Seim, T. (1977). *Talemålsvarianter i Kragerø: en studie i forholdet mellom dialekt og normalmål i et østnorsk bygdesamfunn* (Hovudoppgåve i nordisk språkvitskap). Universitetet i Bergen.
- Selås, M. (2003). *Dialektal variasjon og endring aust på Agder: beskrive og forklart ved to språkelege variablar* (doktoravhandling). Universitetet i Bergen.
- Skre, I. (1957). *Fana-målet. Ljodlæra med stutt omrit av formlæra*. Institutt for nordisk filologi, Universitetet i Bergen.

- Skånland, M.H. (1985). *Lorang eller den forbudte fonologi*. Institutt for fonetikk og lingvistikk, Universitetet i Bergen.
- Stenström, A.-B. (2014). *Teenage talk. From general characteristics to the use of pragmatic markers in a contrastive perspective*. London: Palgrave Macmillan.
- Stenström, A.-B. & Jørgensen, A.M. (red.) (2009). *Youngspeak in a multilingual perspective*. Amsterdam/Philadelphia: John Benjamins.
- Storm, J. (1884). Norsk Lydskrift med Omrids av Fonetiken. *Norvegia. Tidsskrift for det norske folks maal og minder*, 1, 1–179.
- Tenfjord, K., Hagen, J.E. & Johansen, H. (2009). Norsk andrespråkkorpus (ASK) – Design og metodiske forutsetninger. *NOA norsk som andrespråk*, 25(s. 52–81).
- Thue, F.W. (1996). Det humanistiske fagfeltets historie. I: N. Roll-Hansen, J. Goksøyr, L.M. Irgens, K.B. Helle, T.I. Bertelsen, R.K. Lie, A. Nernæs, K.A. Selvig, F.W. Thue, S. Bagge & K.H. Teigen *Universitetet i Bergens historie*. Bind II (s. 434–579). Universitetet i Bergen.
- Utne, I. (1987). Grunnlagsmateriale for informasjonssøking ved Norsk termbank. *Humanistiske data*, 1, 39–53.
- Venås, K. (1967). *Sterke verb i norske målføre*. Oslo: Universitetsforlaget.
- Venås, K. (1974). *Linne verb i norske målføre*. Oslo: Universitetsforlaget.
- Venås, K. (2005). *Sigurd Kolsrud*. Oslo: Novus.
- WHO (2005). ICD-10. *Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer*. Oslo: Sosial- og helsedirektoratet.
- Ådland, E. (1997). Ein krig i krigen. Striden om Norsk Stadnamnarkiv i Bergen. Bergen: *Årbok for Bergen museum* (s. 54–60).
- Årsmelding* (1983, 1984, 1988). Bergen: NAVFs EDB-senter for humanistisk forskning.
- Aasen, I. (1965 [1864]). *Norsk grammatik*. Oslo: Universitetsforlaget.