

Yonatan Goldshtein er ph.d.-studerende ved Afdeling for Nordiske studier og Oplevelsesøkonomi, Aarhus Universitet
Rasmus Puggaard er ph.d.-studerende ved Leiden University Centre for Linguistics.

Yonatan Goldshtein & Rasmus Puggaard:

Overblik over danske dialektoptagelser

Gennem tiden har forskere fra Aarhus Universitet og Københavns Universitet indsamlet en enorm mængde lydoptagelser af danske dialekttalende. Der er mange grunde til at anvende disse lydoptagelser i studiet af dialekterne. Selv den mest fintmærkende lydskrift kan ikke gengive talen helt præcist, så lyd materialet giver os mulighed for at studere de allermindste forskelle mellem sprogbrugen i forskellige egne. Dette til trods er lyd materialet kun i ringe grad blevet anvendt i forskning.

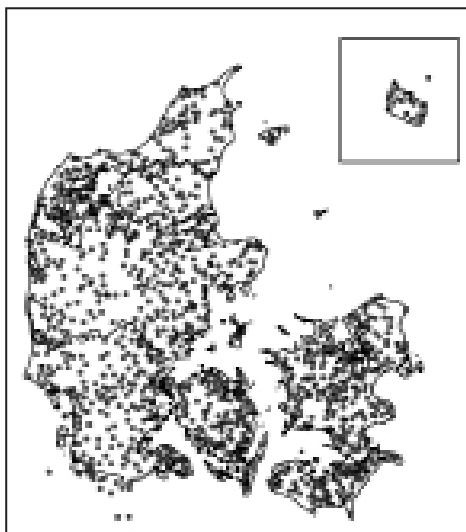
Denne artikel giver ikke et overblik over alle dialektoptagelser, der er lavet af eller findes i samlinger på de to universiteter. Her gives et overblik over det materiale, der er kendt i Peter Skautrup Centret for Jysk Dialektforsknings arkiv. Kernen i dette materiale består af optagelser fra hele landet, der blev indsamlet og optaget under et femårigt projekt, der forløb fra 1971 til 1976, samt et antal optagelser indsamlet så langt tilbage som 1934 og nogle senere indsamlede jyske optagelser – de seneste fra 2006. Med i dette overblik er altså ikke optagelser lavet i forbindelse med senere forskningsprojekter ved Københavns Universitet som del af fx Køgeprojektet (se fx Jørgensen 1998) eller Dialekt i Periferien (se fx Maegaard et al. 2019). Sidst, der blev givet et samlet overblik over Peter Skautrup Centrets lyd materiale er snart 40 år siden (Andersen 1981) – og meget er sket siden!

1. Lydoptagelserne

I sin fulde størrelse består det tilgængelige materiale af 2.383 optagelser optaget på 1.080 forskellige lokationer i Danmark eller områder i

Slesvig hvor en dansk dialekt tales eller taltes. Optagelsernes samlede løbelængde er 2.090 timer eller cirka 87 dage. Punkterne på kort 1 viser de steder hvorfra vi har optagelser.

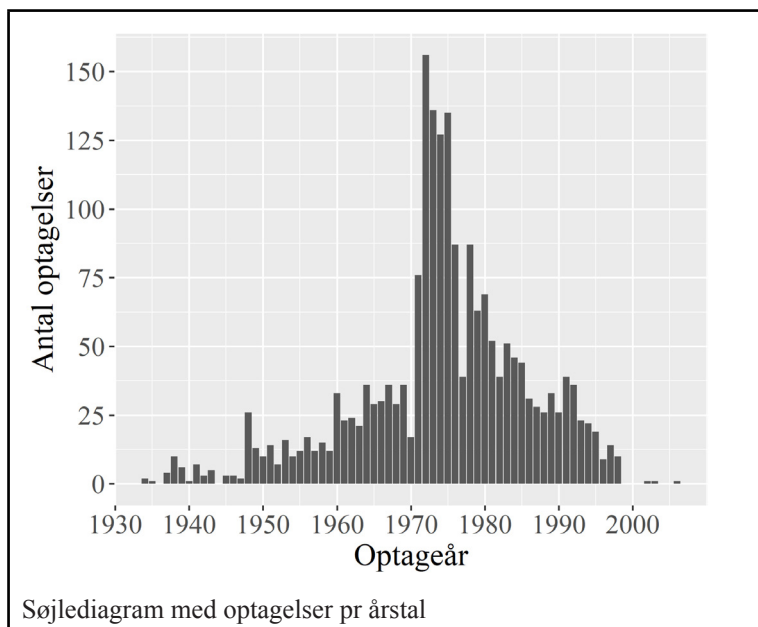
De ældste optagelser i korpusset er fra 1930'erne og er optaget på fonografvalse eller grammofonplade af Statsradiofonien. Disse optagelser indeholder blandt andet nogle samtaler med meget gamle dialekttalende fra Fjorde, Jørl og Hanved. Desværre er lyd kvaliteten på overspilningerne fra fonografvalse sjældent særligt god.



Kort 1. Alle optagelser

Indtil 1970 indsamlede dialektforskere ved det daværende Institut for Dansk Dialektforskning ved Københavns Universitet og Institut for Jysk Sprog- og Kulturforskning ved Aarhus Universitet (nu Afdeling for Dialektforskning og Peter Skautrup Centret for Jysk Dialektforskning) optagelser af dialekttalende forskellige steder i landet. Kasettebåndets fremkomst gjorde indsamling af materiale mere tilgængeligt end tidligere.

I 1971 fik forskere fra de to institutter en femårig bevilling til indsamling af dialektoptagelser i hele landet. I perioden 1971 til 1976 tog indsamlingen derfor virkelig fart. Der blev i den periode indsamlet 525 optagelser (løbelængde ca. 370 timer, eller 15 dage). I indsamlingen var det primære fokus at indsamle materiale omkring de kendte dialektgrænser. Der var mindre fokus på sprogligt mere ensartede områder. Den geografiske dækning er dog stadig helt fænomenal. Indsamlingen af data bliver løbende mindre frekvent og de yngste optagelser i korpusset er fra 2006. Flere af de senere optagelser er optaget i digitalt format.



I slut-1990'erne blev lydoptagelserne overspillet på CD for at fremtidssikre materialet. De jyske optagelser er senere blevet udtrukket fra CD'erne som mp3-filer. Disse er i dag tilgængelige på Peter Skautrup Centrets computere.

Det Kongelige Bibliotek har stået for digitaliseringen af de optagelser, der blev lavet i løbet af den femårige bevilling 1971-1976. I dette projekt blev der foretaget i alt 524 optagelser fra 402 lokationer (løbelængde 366 timer). Punkterne på kort 2 viser hvor disse optagelser blev foretaget.

Som man kan se på kort 2 blev der foretaget optagelser fra hele landet til projektet. Dog er Midtjylland relativt dårligt repræsenteret. Lydkvaliteten på bibliotekets digitaliseringer er langt højere end på Peter Skautrup Centrets egne digitaliseringer, både pga. filformatet og fordi lyden er blevet restaureret direkte fra de oprindelige bånd fremfor CD. Der er potentielt gået data tabt i overspilningen fra bånd til CD, og desuden går en del data tabt, når lyd overspilles til mp3-formatet



Kort 2. Optagelser i bevillingsperioden 1971-1976

frem for andre formater. Lydkvaliteten på mp3-filerne er skaleret ned til en bitrate på 320 kilobytes i sekundet (kb/s), dvs. lidt under en fjerdedel af den oprindelige CD-kvalitet, hvorimod lydkvaliteten på bibliotekets restaurationer ikke er beskåret og har en bitrate på 1.411 kb/s. Forskellen er måske ikke let at høre med det blotte øre, men kan gøre en væsentlig forskel for optagelsernes anvendelighed til akustisk analyse.

Desværre er disse optagelser kun tilgængelige via bibliotekets hjemmeside¹, som kan være svær at navigere. Lydmaterialet kan høres på hjemmesiden, men vi har ikke kunnet få adgang til hele lydmaterialet på andre måder, hvilket giver nogle udfordringer i analysearbejdet.

Materialet, der findes på Peter Skautrup Centret har gennem tiden været organiseret på en række forskellige måder. De optagelser, der oprindeligt har været på centret som bånd, har fået tildelt et båndnummer. Båndnummeret bruges bl.a. til at finde de udskrifter, der findes af dele af materialet. Da båndene blev overspillet på CD, blev flere bånd ofte overspillet på samme CD. Hver CD er udstyret med et CD-nummer. Der virker ikke til at være megen systematik i hvilke bånd, der ligger på samme CD. Mp3-filerne er organiseret i mapper på baggrund af CD-numrene. I mapperne ligger der mellem en og fem filer, der blot er automatisk navngivet efter hvilket spornummer optagelsen havde på CD'en. Særligt blandt de ældste optagelser findes der tit mange optagelser på samme bånd, og derfor også på samme fil, og organiseringen kan være temmelig kaotisk.

¹ <https://dansklyd.statsbiblioteket.dk/samling/dialektsamlingen/>

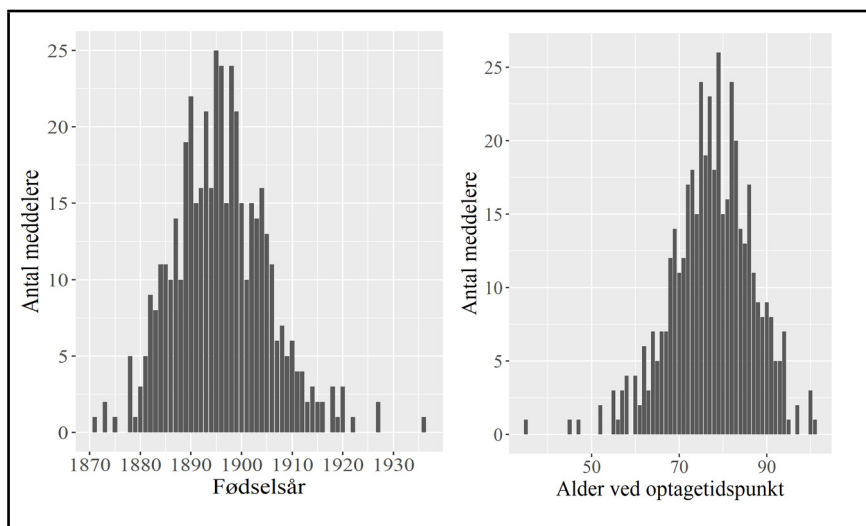
De nødvendige metadata til overskue materialet findes på Peter Skautrup Centret, organiseret i programmet FileMaker Pro. I programmet er der ét opslag for alle optagelser. Til dette opslag knytter der sig udover CD- og båndnummer oplysninger om hvor materialet er optaget, hvem der er meddeler på optagelsen, hvem der har stået for interviewet, noter om indholdet i optagelsen og evt. andre bemærkninger til optagelsen. Lokationen angives både med navnet på det sted (oftest sogn) hvor optagelsen er lavet samt stedets topografiske nummer. De topografiske numre er den måde at angive steder i Danmark, som benyttes af Dansk Folkemindesamling, og en oversigt kan findes i Skjelborg (1967).

2. Optagelsernes indhold og meddelere

Langt de fleste optagelser består af sociolingvistiske interviews med ældre dialekttalende meddelere i deres hjemsgn. Formålet har været at eliciterer ordforråd som kun findes i dialekterne. Interviewene er struktureret som samtaler om emner som fx landbrug, tilberedning af mad og højtider. Der er primært tale om meddelere fra landsogne, og oftest har disse været beskæftiget med landbrug eller fiskeri. Hvis man ser specifikt på materialet indsamlet i 1971-1976, har cirka $\frac{2}{3}$ af optagelserne én mandlig informant, cirka $\frac{1}{4}$ har én kvindelig informant og resten består af gruppesamtaler med mere end én meddeler. Langt de fleste informanter er over 70, og enkelte er over 90 år gamle. De fleste er derfor født omkring århundredeskiftet.

Materialet fra de øvrige perioder er noget mere forskelligartet, da det enten har været indsamlet sporadisk eller med bestemte formål for øje. Blandt det ældre materiale findes en del radiooptagelser af dialekttalende, men også en lang række dialektinterviews foretaget af Poul Andersen, Ella Jensen, Niels Åge Nielsen, K. Ringgaard m.fl. mellem 1940'erne og 1960'erne. Der findes også noget materiale fra 1980'erne og 1990'erne, selvom antallet af optagelser pr år er støt faldende fra slutningen af 1970'erne. En del af det senere materiale adskiller sig fra det ældre ved ikke at være fra landsogne men derimod fra byer som Esbjerg og Aalborg, og talerne er af og til yngre. Der er eksempler på interviews med gymnasieelever og enkelte eksempler på gruppeinterviews med meddelere fra flere generationer. Tilførslen

af nyt materiale stopper stort set i slut-1990'erne med enkelte nye tilførsler i 2002, 2003 og 2006.



Søjlediagram med meddeleres alder og fødselsår fra bevillingen 1971-1976

3. Udskrift af jyske optagelser

Det ideelle ville selvsagt være at have udskrifter af hele materialet i fonematisk lydskrift. Sådanne udskrifter findes for dele af materialet i Thy, udarbejdet af Torsten Balle, og men også for (dele af) enkelte andre optagelser udarbejdet i forbindelse med udgivelse af dialekttekster. Umiddelbart efter det femårige indsamlingsprojekt i 1970'erne blev der i Aarhus lavet en række udskrifter af jyske optagelser i et format, der var tænkt at kunne bruges i forbindelse med arbejdet med Jysk Ordbog. Sådanne udskrifter findes fra ca. 50 sogne angivet ved punkterne på kort 3.

I disse udskrifter står ordene i deres opslagsform efterfulgt af ordets bøjning. Ord med særlig interesse for ordbogsarbejdet blev i teksten markeret med en stjerne, så man let ville kunne finde dette ord igen. Et eksempel på en sådan transskription ser ud som nedenstående:

de (pron) kalde (v-præt) det (pron-ubest) en (art) *skødsel (s)

Dette eksempel kommer fra en optagelse af Kathrine Wulf (f. 1889) fra Kegnæs Sogn, der fortæller om det redskab man brugte til at sætte



Kort 3: Steder hvorfra der findes transskriptioner ved Jysk Ordbog

brød i ovnen. I almindeligt skriftsprog ville samme linje se ud som følger:

de kaldte det en skødsel

Den valgte transskription med adskilte opslags- og bøjningsformer har den fordel, at den udelukker at homofoner blandes sammen. Fx ville substantivet *dør* transskriberes *dør (s)*, mens verbet *dør* ville transskriberes *dør (v-præs)*. Denne udskriftsmetode er også let at bruge i ordbogsredigering, da den relevante information allerede er knyttet til de enkelte ord. Skriftsprogsudskriften er dog meget lettere at læse. I fonematisk lydskrift kunne samme ytring gengive:

di kålt de en skysøl

Denne udskriftsmetode har den fordel, at den vil indfange en stor del af variationen mellem taleres sprog, men den ulempe, at fonemisk transskription er enormt tidskrævende. Den ordbogsorienterede transskription findes på Peter Skautrup Centret i papirform eller som tekstdokumenter.

Alle udskrifterne er organiserede med linjenumre og hver linje er tilknyttet et triptællernummer. Triptællernummeret refererer til en fysisk placering på det oprindelige bånd, og kunne bruges til at finde en given ytring på båndet. Det svarer cirka til sekunder, men der er umiddelbart ingen præcis korrespondance mellem triptællernummeret og de digitalt tilgængelige lydfiler – jo længere, man kommer hen i en optagelse, jo mere forskel bliver der på triptællernummeret og ytringens placering i lydfilen. Vi kender ikke den præcise grund til dette.

I samarbejde med IT-centret ved Aarhus Universitet blev der i 1970'erne udviklet computerprogrammer, der kunne ordne ordene i forskellige lister. Der blev lavet alfabetiske lister over alle ord (såkaldte

Al-lister) og lister ord de særligt interessante ord, fx **skødsel* (*s*) (såkaldte S-lister). Listerne angiver i hvilke udskrifter de enkelte ord fandtes, på hvilken linje i udskriften og på hvilket triptællernummer i optagelsen. Disse lister blev samlet i bøger, der i dag står på Peter Skautrup Centret.

I forbindelse med vores arbejde med materialet har vi lavet nogle forsøg med at oprette lignende ordlister på baggrund af udskrifterne. Ved hjælp af programmering i Python-miljøet lykkedes det på et par arbejdsdage at lave lignende lister på baggrund af et udsnit af materialet. Hvis udskrift og lyd blev knyttet sammen, ville man relativt hurtigt kunne udarbejde søgbare ordlister med direkte adgang til ordenes placering i lydfileerne.

Det er videre lykkedes med nogen succes at automatisere oversættelsen af nogle transskriptioner til et mere læseligt skriftsprog-dansk, selvom dette dog vil kræve noget manuelt forarbejde. Automatisering af fonemisk lydskrift vurderes ikke umiddelbart at være muligt.

4. Udskrift og anvendelse af ømålsoptagelser

En række dialektoptagelser er blevet udskrevet på Københavns Universitet i et andet format (Pedersen 1973, 1983). De københavnske udskrifter har fundet noget bredere anvendelse end de jyske. Udskrifterne findes i vid udstrækning online i søgbart format som del af korpusset CorDiale. Gudiksen & Hovmark (2008) beskriver hvordan optagelserne understøtter arbejdet med redigeringen af Ømålsordbogen især når det gælder funktionsord og højfrekvente ord, der ikke varierer særligt meget fra rigsmålets. Disse ord er sjældent særligt frekvente i seddelsamlingerne, der ellers ligger til grund for ordbogsarbejdet. Som eksempel bruges blandt andet artiklen om ordet 'ja', der i Ømålsordbogen (bind 9) fylder 16 spalter med en mængde underinddelinger efter funktion. I Jysk Ordbog findes der to ret korte artikler om 'ja' som hhv. interjektion og substantiv. Hovmark (2004) beskriver videre hvordan båndmaterialet er en kilde til citater i ordbogen. Ifølge Karen Margrethe Pedersen (personlig kommunikation) findes der cirka 170 tekster fra 150 sogne, $\frac{2}{3}$ af dem fra Øerne og resten fra Jylland.

5. Anvendelse af de jyske optagelser

Det var oprindeligt ideen, at de jyske optagelser skulle fungere som supplement til det skriftlige materiale i udarbejdelsen af Jysk Ordbog. Andersen (1981) nævner eksempler på hvordan de kan bruges til at understøtte ordbogsarbejdet, men de er indtil for nylig kun i meget ringe grad blevet anvendt i videnskabeligt arbejde.

Optagelserne er blevet anvendt i en række universitetsopgaver og specialer. Kopier af disse opgaver findes dog ikke systematisk indsamlet. Optagelserne er desuden blevet brugt i formidlende arbejde, bl.a. på de interaktive dialektkort på hjemmesiden *Dialekt.dk*.²

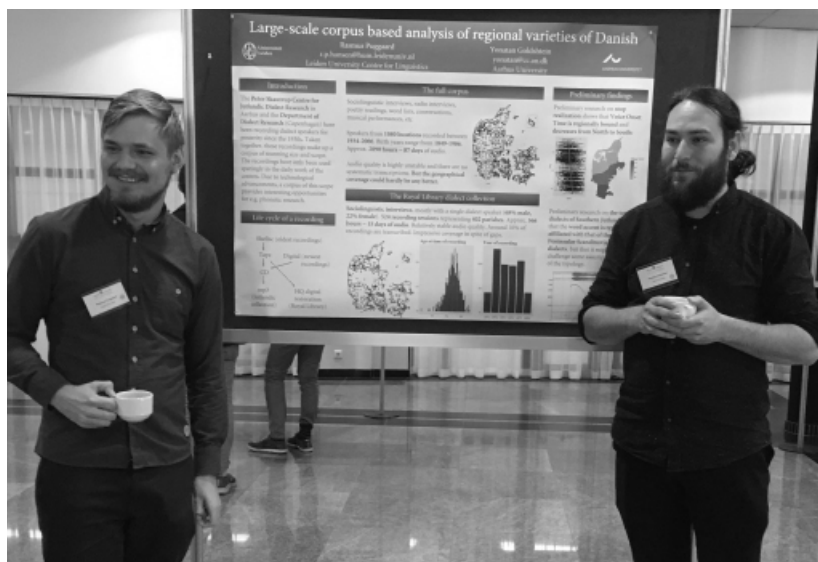
Siden 2018 har denne artikels to forfattere været i gang med hver deres PhD-projekt, hvor lydoptagelserne anvendes til akustisk-fonetiske undersøgelser. Rasmus Puggaard arbejder bl.a. med konsonanter og stavelsesstrukturer i jysk ved Leiden Universitet og har udgivet og præsenteret forskellig forskning om variation i jyske klusiler (Puggaard 2018a, 2018b, 2019). Yonatan Goldshtein skriver om intonation, accentuation og tonale forhold i jysk ved afdelingen for Nordisk sprog og litteratur på Aarhus Universitet. Han har indtil videre primært beskæftiget sig med ordaccenter på østsønderjysk (Goldshtein 2019). Et af vores fælles mål er desuden generelt at udbrede viden om og inspirere til brug af det imponerende korpus af lydoptagelser, og indholdet i denne artikel lægger sig tæt op af en conferencepræsentation med netop dette formål (Puggaard & Goldshtein 2018).

Litteratur

- Andersen, Torben Arboe (1981): Dialektbånd og databehandling. I: *Ord & Sag* 1:11-18.
- Goldshtein, Yonatan (2019): Tonal accent in Felsted, South Jutland. Præsenteret ved *10th International Conference on Language Variation in Europe*, Fryske Akademy, Leeuwarden.
- Gudiksen, Asgerd & Henrik Hovmark 2008. Båndoptagelser som kilde til Ømålsordbogen. *Nordiske Studier i Leksikografi* 9:173-182.

² <https://dialekt.ku.dk/dialektkort/>

- Hovmark, Henrik 2004. Jagten på det autentiske citat. *Danske Talesprog* 5: 3-21.
- Jørgensen, Jens Normann (1998): Introduktion til Køge-projektet. I: Janus Spindler Møller (red.): *Tosproget udvikling. Studier i Køge-projektet*, 5-19. Aarhus: Danmarks Lærerhøjskole.
- MacWhinney, Brian & Johannes Wagner (2010): Transcribing, searching and data sharing. The CLAN software and the TalkBank data repository. I: *Gesprächsforschung* 11:154-173.
- Maegaard, Marie, Malene Monka, Kristine Køhler Mortensen & Andreas Candefors Stæhr (red.) (2019): *Standardization as sociolinguistic change. A transversal study of three traditional dialect areas*. London & New York: Routledge.
- Pedersen, Karen Margrethe. 1973. Dialekttekster i rigsmålsnotation med becifring: et EDB-projekt. I: *Danske Folkemål* 20:29-46.
- Pedersen, Karen Margrethe 1983. Om transskription af båndoptagelser og anvendelse af teksterne til leksikografiske og syntaktiske undersøgelser. Et EDB-projekt. I: *Folkmålsstudier. Meddelanden från Föreningen för nordisk filologi* 28:131-139.
- Puggaard, Rasmus (2018a): Realizations of /t/ in Jutlandic dialects of Danish. I: *Linguistica Lettica* 26:368-393.
- Puggaard, Rasmus (2018b): Telefoniske båndpasfiltres indflydelse på talergenkendelse. Høje frekvenser og dansk /t/. I: *Nydanske Sprogstudier* 54:129-155.
- Puggaard, Rasmus (2019): Microvariation in stop realization as a consistent regional feature of Danish. Poster præsenteret ved *10th International Conference on Language Variation in Europe*, Fryske Akademy, Leeuwarden.
- Puggaard, Rasmus & Yonatan Goldshtein (2019): Large-scale corpus-based analysis of regional varieties of Danish. Poster præsenteret ved *10th International Conference on Language Variation in Europe*, Fryske Akademy, Leeuwarden. Link: <http://kortlink.dk/yp35>.
- Skjelborg, Åge (1967): *Topografisk ordningen af Danmarks egne med alfabetisk nøgle*. København: Statens Trykningskontor.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer. 1-. København: Institut for Dansk Dialektforskning og C.A. Reitzels Forlag 1992-.



Rasmus Puggaard og Yonatan Goldshtein præsenterer båndsamlingen på ICLaVE 10 i Leeuwarden