CoAmDA. Et nyt dansk talesprogskorpus

Af KAROLINE KÜHL, JAN HEEGÅRD PETERSEN, GERT FOGET HANSEN og FRANS GREGERSEN

Denne artikel beskriver 'Corpus of American Danish' (CoAmDa), et korpus af dansk talesprog fra USA, Canada og Argentina, som er blevet opbygget siden sommeren 2014 inden for projektet 'Danske Stemmer i USA og Argentina' ved Københavns Universitet. Artiklen skal forstås som en redegørelse for viden, der er nødvendig for brug af korpusset. Den omfatter en opgørelse over de forskellige data og dataindsamlinger, der ligger til grund for korpusset, og en beskrivelse af de talertyper, der er karakteristiske for de forskellige datasæt. Vi inkluderer også sociodemografisk baggrundsviden om de danske samfund i Nord- og Sydamerika, som vi skønner nødvendig for en forståelse af den sproglige udvikling. Foruden disse metadata dokumenterer artiklen selve korpussets struktur og etablering med oplysninger om den kvantitative status, en redegørelse for digitaliseringen af ældre data samt en specifikation af udskrivningsstandarder og opmærkning af korpussets materiale.

1 Indledning

I denne artikel beskrives et nyt talesprogskorpus for dansk, nærmere bestemt for de former for dansk, der blev talt af de første danske immigranter i Nordamerika, og det dansk, der stadig tales af efterkommere af danske udvandrere til Argentina.[1] Korpusset, CoAmDa (Corpus of American Danish), er siden 2014 blevet etableret inden for projektet 'Danske Stemmer i USA og Argentina' ved Institut for Nordiske Studier og Sprogvidenskab ved Københavns Universitet.[2]

[1] Vi vil gerne takke en anonym fagfælle samt redaktionen for kritiske og nyttige kommentarer til artiklen samt korpusadministratorerne Iben Nyholm Debess og Nicolai Boston Simonsen for deres store hjælp med opbygningen og brug af CoAmDa.
[2] Forskningsprojektet 'Danske Stemmer i USA og Argentina' (2014–2018) er finansieret af A.P. Møller og Hustru Chaste Mc-Kinney Møllers Fond til almene Formaal, Carlsbergfon-

Vi vil i denne artikel for nemheds skyld kalde de to former for dansk for *amerikadansk* og *argentinadansk*, uden dog derved at tilsigte mere end en geografisk samlende betegnelse. I både det ene og det andet tilfælde dækker betegnelserne over stor intra- og interindividuel sproglig variation. Amerikadansk og argentinadansk omfatter elementer, som repræsenterer ældre stadier af standarddansk og moderne standarddansk, stærkt traditionelt dialektpræget dansk og svagt regiolektpræget dansk. Det mest fremtrædende kendetegn ved både amerika- og argentinadansk er dog, at de er præget af det omgivende og samfundsmæssigt dominerende sprog, dvs. amerikansk engelsk, canadisk engelsk og argentinsk spansk; men også her er der stor inter- og intraindividuel variation.

Formålet med artiklen er en beskrivelse af CoAmDa, som kan bruges som udgangspunkt for videre forskning.[3] Beskrivelsen omfatter en redegørelse for de forskellige data og dataindsamlinger, der ligger til grund for korpusset, samt en beskrivelse af de talertyper, der er karakteristiske for de forskellige datasæt. Vi inkluderer også en del sociodemografisk baggrundsviden, som vi skønner er nødvendig for at kunne vurdere amerika- og argentinadansk sprogligt. Foruden disse metadata dokumenterer artiklen selve korpussets struktur og etablering, med oplysninger om den kvantitative status, en redegørelse for digitaliseringen af ældre data samt en specifikation af format samt udskrivningsstandarder, som er relevant viden for brug af korpus.

Artiklen er bygget op efter følgende struktur: Afsnit 2 giver et kortfattet overblik over CoAmDa i korpussets nuværende form, afsnit 3 og 4 er hhv. en beskrivelse af de nord- og sydamerikanske data og deres oprindelse. I afsnit 5 redegør vi for lydoptagelserne, principperne bag digitaliseringen af ældre data samt principperne for udskrivning af optagelserne og opmærkning af data. Afsnit 6 henviser til nogle perspektiver for brug af CoAmDa.

---

det som *Semper Ardens*-projekt, Institut for Nordiske Studier og Sprogvidenskab og Nordisk Forskningsinstitut. Vi takker for den økonomiske støtte til projektet.

[3] I skrivende stund (juni 2017) er korpusset ikke generelt tilgængelig for den akademiske offentlighed, men vi håber at der i fremtiden vil blive åbnet for at andre kan arbejde uhindret med CoAmDa.

2 CoAmDa i 2017

CoAmDa huses af Sprogforandringscentret ved Københavns Universitet (KUCSS), og det indgår i Sprogforandringscentrets LANCHART-korpus og dets korpusstruktur (jf. Gregersen & Kristiansen 2015). Ligesom LANCHART-korpussets data er optagelser, der indgår i CoAmDa, udskrevet i programmet Transcriber og valideret ved flere processer (jf. afsnit 5).

I skrivende stund indeholder CoAmDa samlet 184 timers optagelser. Tabel 1 viser en opgørelse over fordelingen på de forskellige datasæt:

Tabel 1: Kvantitativ status over CoAmDa (juni 2017)[4]

Delkorpusser	Timer	Talere	Ord
USA	58	194	512.991
Canada	22	45	132.154
Argentina	83	93	891.708
samlet	**218**	**322**	**1.536.853**

Tabellen viser en skæv fordeling mellem antal talere og optagelser. Optagelserne fra Argentina er væsentligt længere end de ældre optagelser fra Nordamerika. Derfor dominerer de nordamerikanske talere kvantitativt, mens de argentinske data udgør størstedelen af data. For en nærmere beskrivelse af dataindsamlingerne og informanter henviser vi til afsnit 3 og 4. Her skal det blot nævnes at alle kendte oplysninger om talerne, så som navn, fødested, fødeår, udvandringsår og bopæl, indgår i korpusset.

Som nævnt stammer CoAmDas materiale fra forskellige dataindsamlinger: Danske Stemmer-projektet har i flere omgange i 2014 og 2015 indsamlet data forskellige steder i Argentina, men de nordamerikanske

[4] Danske Stemmer-projektet råder som alle andre talesprogskorpora over mere materiale fra flere talere, end hvad der er udskrevet. Differencen mellem det transskriberede materiale og det samlede antal timer og informanter hviler på en beslutning om ikke at udskrive alle data på grund af fx manglende metadata om talerne eller for høje omkostninger pga. dårlig lydkvalitet. Artiklens fokus ligger på CoAmDa, og af denne grund redegøres der i teksten ikke for data, der ikke indgår i korpus. Antal ord defineres her som antal udfyldte intervaller, dvs. at fx tøvelyde og tilbagekanaliseringer tæller med.

data er arvegods fra forskellige andre forskere og forskningsprojekter. Det følgende afsnit redegør for de nordamerikanske datas oprindelse.

3 Corpus of North American Danish: Dansk i Nordamerika

Corpus of North American Danish (CoNAmDa), den nordameri-kanske del af CoAmDa, omfatter ca. 96 timers optagelse af 231 talere. Disse data har vi overtaget efter andre forskere og projekter:[5]

1. Lektor Iver Kjær (Københavns Universitet) og professor Mogens Baumann Larsens (Aalborg Universitet) store dataindsamling fra USA og Canada fra 1960'erne og til 1980'erne. Kjær og Baumann Larsen optog 179 dansktalende udvandrere eller disses efterkommere i USA samt 23 i Canada. Heraf indgår 194 personer i korpus.
2. Professor Tore Kristiansens (Københavns Universitet) optagelser af i alt 40 personer i Solvang, Californien, fra starten af 1990'erne. Heraf indgår optagelser med 12 personer i korpusset.
3. Professor emeritus Christopher Hales (University of Alberta, Edmonton) optagelser af 24 dansktalende i New Denmark, New Brunswick, Canada, sidst i 1990'erne, som alle indgår i korpus.
4. Lektor Anne Nesser og ph.d.-studerende Mikala Jørgensen (Aarhus Universitet) optog i årene 1999–2001 knap 100 dansktalende personer i Californien og indsamlede omfattende spørgeskemabesvarelser fra i alt 77 informanter inden for projektet DANA (Danes in North America). Heraf indgår kun optagelser af fem personer (knap fire timer) i korpus.[6] Resten er ikke udskrevet pga. uoverensstemmelse med Danske Stemmer-projektets øvrige talertyper: Projektet DANA har mest fokuseret på nutidens danske *expats* i Californien. Disse talere har større muligheder for vedvarende tilknytning til Danmark også under migrationen end udvandrerne fra omkring det sidste år-

[5] Vi vil i denne forbindelse gerne takke følgende for generøst at stille data til rådighed: Inger Kjær, Tore Kristiansen, Christopher Hale og Mikala Jørgensen. Endvidere vil vi gerne takke Jakob Steensig for samarbejde om amerikanske data. Kjær og Baumann Larsens data blev også bearbejdet af Karoline Kühl i hendes FFK-støttede post doc-projekt ('Dansk uden for Danmark', 2012–2014).

[6] Spørgeskemabesvarelserne har vi dog analyseret i Heegård Petersen, Thøgersen & Kühl (2015).

hundredskifte og deres efterkommere. Siden DANA-projektets data udgør en så lille del af det udskrevne korpus, redegøres der i det følgende ikke særskilt for disse data.

CoNAmDas materiale repræsenterer talere født mellem 1870 og 1941, både danskfødte udvandrere og deres amerikansk- eller canadiskfødte efterkommere.

3.1 Sociodemografisk baggrund

Den danske udvandring til USA og Canada er beskrevet grundigt i flere akademiske udgivelser (Hvidt 1971, Grøngaard Jeppesen 2005, 2010; Bredmose Simonsen 1990; Larsen & Bender 1992; for en kortfattet opsummering henviser vi til Kühl 2015c), men også i flere populærhistoriske udgivelser og dokumentarfilm.

En kortfattet version af hovedresultaterne lyder: Til trods for at emigrationen til USA og senere Canada har været betragtelig set med danske øjne, så var den danske gruppe en mindre gruppe set med amerikanske øjne og sammenlignet med de andre nordeuropæiske udvandrergrupper. De danske indvandrere og især deres efterkommere viste, overordnet set, kun en svag tendens til at samles i og om danske bosættelseskolonier, dansk kirke og anden dansk(sproglig) kultur. Der fandtes danske bosættelser, hvor dansk sprog og kultur blev opretholdt, fx Racine County i Wisconsin, New Denmark i Canada, visse kvarterer i Chicago og ikke mindst bosættelserne grundlagt inden for den dansk-amerikanske 'kolonisag'. Sidstnævnte betegner det grundtvigiansk prægede initiativ 'Dansk Folkesamfund', grundlagt i 1883–84, som arbejdede for oprettelsen af danske bosættelser med danske børne- og højskoler, dansk kirke, dansk sprog og dansk levevis. Det lykkedes at oprette en hel del af sådanne bosættelser, fx Tyler (Minnesota, grundlagt 1885), Danevang (Texas, grundlagt 1890), Ringsted (Iowa, grundlagt 1899), Askov (Minnesota, grundlagt 1906) og, som den sidste, Solvang i Californien (grundlagt 1911). For en grundig beskrivelse af 'kolonisagen' og livet i Tyler henviser vi til Bredmose Simonsen (1990).

Men størstedelen af dansk-amerikanerne var socialt og geografisk mobile og tilpassede sig hurtigt i den amerikanske "smeltedigel". Den sproglige konsekvens har været, at dansk (overordnet set) tidligt ophørte med at blive brugt som dagligsprog. Mens der i 1900 kun var 5,3 % af de danskfødte i USA, der i de amerikanske censusbesvarelser angav, at de ikke kunne tale engelsk, var dette tal i 1920 allerede faldet til 2,8 % (Grøngaard Jeppesen 2005: 283; Kühl 2015: 39–41). Baseret på demografiske analyser kan man altså generelt konkludere, at hvis dansk-amerikanere i begyndelsen af det 20. århundrede bibeholdt dansk som dagligsprog i mere end én generation, var det på trods af omstændighederne, ikke på grund af dem. Dette billede understøttes af Kjær & Baumann Larsen (1974: 25), som konkluderer, at de første danske indvandrere i midten af 1970erne ikke længere brugte dansk som et aktivt dagligsprog, men at dansk for dem var blevet til et husket sprog. I dag er der formentlig meget få efterkommere af udvandrerne fra udvandringen før 2. Verdenskrig tilbage, der stadig taler dansk (jf. også Kristiansen, Harwood & Giles 1991).[7] Mens de andre fastlands-skandinaviske sprog i Nordamerika er blevet beskrevet tidligt og løbende, er det påfaldende, så lidt vi ved om, hvordan dansk har udviklet sig som minoritets- og arvesprog i Nordamerika. CoNAmDa giver en solid empirisk baggrund for undersøgelser af dette forskningsspørgsmål.

Den nordamerikanske del af korpus består som nævnt af arvet datamateriale fra tre dataindsamlinger, som vil blive beskrevet nærmere i de følgende afsnit.

3.2 Iver Kjær og Mogens Baumann Larsens data

Størstedelen af optagelserne i CoNAmDa stammer fra de danske forskere Iver Kjær (1938–2002) og Mogens Baumann Larsens (1930–2001) rejser til USA og Canada 1963–1980, som Danske Stemmer-projektet har fået

[7] Udvandringen til USA fortsatte naturligvis efter 2. Verdenskrig, den foregår stadig den dag i dag, og der tales formentlig dansk rundt omkring i USA i dag, dvs. at dansk vedbliver at være et arvesprog i USA. Men hvor, hvor meget, af hvem og i hvilke sammenhænge er os bekendt kun sparsomt belyst (men se Heegård Petersen, Thøgersen & Kühl 2015), ligesom det heller ikke er belyst, i hvilket omfang dansk videregives som et aktivt sprog til 'senudvandrernes' børn.

overdraget fra Iver Kjærs enke, Inger Kjær. Optagelserne omfatter samlet 95 mænd og 108 kvinder født i perioden 1870–1933, og heraf indgår størstedelen (194 personer) i CoAmDa. Ca. ⅔ er immigranter fra Danmark, og langt størstedelen er udvandret i perioden 1887–1929. Af alle amerikansk- eller canadiskfødte informanter i dette datasæt er der kun 13 talere af som har (mindst) en forælder, der også er født i USA eller Canada. Denne talerfordeling repræsenterer dansk-amerikanernes forholdsvis hurtige tab af det danske sprog.

Støttet af Carlsbergfondet var hovedærindet med Kjær og Baumann Larsens rejser at indsamle og bevare mundtlige beretninger om indvandringer fra 1. generationsindvandrerne og at dokumentere eventuelle tilfælde, hvor gamle dialekter havde overlevet grundet isolation og konservatisme (Baumann Larsen 1981: 4). Sigtet lader dog til med tiden at skifte hen imod en interesse for det amerikadanske blandingssprog, bl.a. på grund af at Kjær og Baumann Larsen ikke fandt mange dialektrester. En undtagelse er dog området omkring Lake Norden i South Dakota, hvor thybomål var blevet bevaret (Baumann Larsen 1981: 4–5; Kjær & Baumann Larsen 1981; Heegård Petersen, under udgivelse). Kjær og Baumann Larsens interview varierer i længde, de kan vare alt mellem 7–45 minutter. Mange af dem har en noget formel spørgsmål-svar-struktur, i nogle af dem gives der dog plads til længere narrativer fra informanterne. Gennemgående spørgsmål har drejet sig om erindringer fra den første pionertid (på prærien), den store depression i 1930'erne, religionskonflikten mellem grundtvigianere og Indre Mission i det dansk-amerikanske samfund[8] motivet bag afrejsen og rejsen over Atlanterhavet med skib. De fleste interview er solointerview, i enkelte tilfælde er ægtefæller eller to bekendte interviewet samtidig. Cirka halvdelen af informanterne er optaget én gang, resten er optaget to gange på forskellige rejser, fem er også optaget en tredje gang på en tredje rejse, og tre er optaget fire gange. Til en vis grad har vi altså i denne del af CoNAmDa mulighed for at vise sprogforandring i virkelig tid, men størstedelen af data tegner et synkront billede.

[8] Baumann Larsen havde religion som bifag og var interesseret i forholdet mellem religion og sprog, jf. Baumann Larsen (1973).

Kjær og Baumann Larsen har omhyggeligt dokumenteret metadata (navn, fødeår, -sted, udvandringsår, bopæl mm.) om deres informanter. Disse informationer indgår i korpus.

3.3 Tore Kristiansens data fra Solvang, Californien

Tore Kristiansen (Københavns Universitet) optog i starten af 1990'erne i alt 40 personer med danske rødder i Solvang, Californien. Heraf indgår knap 8 timers optagelser af 16 personer i CoNAmDa. Grunden til den forholdsvis lille del af processeret materiale i korpus er manglen på information om de fleste talere. Vi har umiddelbart kun haft den information om informanterne, der var noteret på båndene eller deres æsker samt information som fremgår af lydoptagelserne selv. Vi har derfor fokuseret på de talere, vi ved mest om. Af de talere, der indgår i korpus, er den ældste født i 1899. Mindst 7 er amerikansk-fødte og flyttet til Solvang fra Midtvesten sent i deres arbejdsliv eller som pensionister. Sigtet med Kristiansens optagelser var en undersøgelse af det danske sprogs 'ethnolinguistic vitality' (jf. Kristiansen, Harwood & Giles 1991), hvilket også omfattede anden evidens end optagelserne, men ikke en decideret lingvistisk undersøgelse af det amerikadansk, der tales. Optagelserne omfatter løst strukturerede interview og gruppesamtaler mellem to eller tre informanter.

3.4 Christopher Hales data fra New Denmark, Canada

Christopher Hale (University of Alberta, Edmonton) optog i 1998 24 dansktalende canadiere i New Denmark, New Brunswick, Canada. Optagelserne er løst strukturerede enkeltinterviews, enkelte gange to informanter sammen. Alle disse optagelser indgår i CoNAmDa. Talerne er født i perioden 1917–1941, og der er 8 kvinder og 16 mænd. Også her kender vi kun til de metadata om talerne, som har været noteret på båndene og deres æsker samt hvad der fremgår af interviewene.

New Denmark, grundlagt i 1872, er en af de ældste danske bosættelser i Nordamerika (Lang 2005), som derudover har været forholdsvis standhaftig mht. bevarelse af dansk. Optagelserne fra New Denmark bidrager således

med nogle dansksprogede talere med mindst én forælder født i Nordamerika), noget som ellers er sjældent i den nordamerikanske migrationssituation.

4 Corpus of South American Danish: Dansk i Argentina

Corpus of South American Danish (CoSAmDa), den argentinske del af CoAmDa, omfatter i skrivende stund (juni 2017) ca. 93 timers optagelse af 138 talere, som er født mellem 1911 og 1971. Heraf udgøres langt størstedelen af kvinder, gennemsnitsalderen er forholdsvis høj (72 år), og kun få er selv udvandret fra Danmark. Vi kender ikke til nogen tidligere optagelser af dansk-argentineres sprog, og som sådan er opbygningen af CoSAmDa et vigtigt første skridt i dokumentation af et døende sprog. Det betyder dog også, at vi kun kan tegne et synkront billede af det dansk-argentinske sprog.

I det følgende vil vi kortfattet beskrive Danske Stemmer-projektets dataindsamling samt give et ligeledes kortfattet overblik over den historiske og aktuelle sociodemografiske baggrund for det danske sprog i Argentina.

4.1 Danske Stemmer-projektets dataindsamling

Optagelserne af dansktalende dansk-argentinere er foretaget på samlet tre rejser i 2014 og 2015 af deltagere af Danske Stemmer-projektet, Jan Heegård Petersen, Karoline Kühl og ph.d.-studerende Anna Sofie Hartling samt konsulent Jon Albris (i skiftende kombinationer). Der blev optaget i Buenos Aires, i Eldorado i det nordlige Argentina og særligt i 'Danskertrekanten' syd for Buenos Aires. I korpusset indgår desuden optagelser foretaget i marts 2015 af specialestuderende Amalie Glargaard, med såkaldte 'frøhistorier' (Berman og Slobin 1994), som var grundlag for en undersøgelse af bevægelsesverber i argentinadansk (Glargaard 2015).

De indledende kontakter til det argentinadanske samfund er foregået gennem Den Danske Kirke i Buenos Aires, Den Danske Ambassade i Buenos Aires og gennem en af projektdeltagernes familierelationer. Kontakten med det danske samfund i Eldorado er kommet i stand gennem Den Dansk-Argentinske Forening i Danmark. Fra disse første kontakter er kredsen af informanter blevet udvidet gennem 'venners venner'-princippet og gennem deltagelse i den dansk-argentinske 'sommerfest' ved La Dulce og i højsko-

letræffet *Cinco Días*. Ved disse lejligheder blev der uddelt en (spansk- og dansksproget) pjece med information om projektet.

Som indsamlingsmetode har vi anvendt en version af det sociolingvistiske interview. Der er fra to til fem deltagere i interviewene, typisk en informant og to interviewere, men også tit en informant og en interviewer. I gruppeinterviewene er der oftest to-tre informanter, men nogle gange også fire, og en til to interviewere. Gruppeinterviewene er kommet i stand ved fx at ægtepar er blevet interviewet sammen, eller ved at en person har inviteret en bekendt eller et familiemedlem med. Andre gange har vi oplevet, at kontaktpersonen har arrangeret sammenkomst med flere deltagere og opdækning. I disse tilfælde har samtalen ofte været meget overlappende og temmelig fri. Interviewerens opgave i sådanne situationer har været at forsøge at fordele taletiden ligeligt og at få informanterne til ikke at tale i munden på hinanden. Dette har ikke altid været en nem opgave.

Vi har haft som overordnet princip at få informanterne til at fortælle om deres liv fra vugge til grav. Det har på flere måder været givende: Emnet er umiddelbart relevant for informanten, princippet giver en vis ensartethed med henblik på teksttype og ikke mindst er det muligt at elicitere nødvendig faktaviden så som fødested og -år, sprogbrug i barndommen, ophold i Danmark og lignende, uden at det munder ud i en ren spørgsmål-svar-struktur. Den kronologiske orden giver desuden intervieweren god mulighed for at orientere sig i interviewet og finde på impulser til at holde samtalen i gang. Det har været tilstræbt at elicitere så meget fri tale fra informanten som muligt og at få vedkommende til at tale om forskellige emner og episoder for på den måde at få belyst, om der var emner som vedkommende sprogligt set, dvs. med henblik på dansk ordforråd og generelt 'flydende sprog', var mere sikker i end andre. Af samme grund har vi i de fleste interview været tilbageholdende med at tilkendegive nogen form for spanskkompetence, netop for at få informanterne til at udnytte deres danske sprogkompetencer til det yderste. Generelt har fortælleglæden været stor, og interviewene er blevet tilsvarende lange. De korteste interview varer omkring 40–45 minutter, de længste er på 90–100 minutter (se afsnit 5 for teknisk specifikation af optagelsesudstyr og -principper).

Alle informanterne blev informeret om optagelsens formål, og alle har været positivt stemt over for projektet og over for det at blive optaget med henblik på at undersøge det danske sprog i Argentina. Alle informanter er blevet lovet anonymitet, og alle har givet mundtlig tilladelse til, at optagelserne kan bruges til forskning.

For hvert interview er der blevet ført logbog med fx interviewerens bedømmelse af forløbet og særlige elementer i informantens/-ernes livsforløb. Desuden blev der udfyldt et spørgeskema med metadata om informanterne, fx informantens bopæl, kontaktinformation, fødested, fødeår, eventuelle ophold i Danmark, uddannelse og beskæftigelse, ægtefælles navn og 'etnicitet' samt forældrenes og bedsteforældrenes fødested, for så vidt dette var kendt. Sidstnævnte tjener til belysning af grad af slægtsmæssig forankring i Argentina (hvis forældre og bedsteforældre er født i Argentina) og eventuel dialekt- eller regiolektpåvirkning (hvis forældre er født i Danmark). Hvad angår forældres eller bedsteforældres fødested i Danmark, er oplysningerne ikke altid lige præcise. Ofte nævner informanterne blot "Sjælland" eller "et sted i Jylland", og nogle gange er det uoplyst. Metaoplysningerne indgår i CoAmDa.

4.2 Sociodemografisk baggrund for CoSAmDa

Den danske udvandringshistorie til Argentina er langt fra så kendt i den danske offentlighed som udvandringen til Nordamerika, og den er sparsomt beskrevet, med den argentinske historiker María Bjerg (bl.a. 1996) som en spansksproget undtagelse (se også Bækhøj 1948). Der findes dog en del erindringer og personportrætter (Johansen 1934; Engwald 1938) og en del nyere antropologiske og historiske specialer (fx Worgall 2011, Agertoft 2005 og Jarnum 2006). Hertil kommer oplysningerne fra vores informanters livshistorier og fortællinger om deres familiers migration, der med et vist personligt præg afspejler etableringen og udviklingen af de danske kolonier i Argentina.

Den danske udvandring til Argentina kan opdeles i to bølger. Den første bølge var beskeden i antal og varede fra ca. 1860–1880. Bølgen omfattede hovedsagelig folk fra Lolland og andre Sydhavsøer, der, ansporet af pione-

ren Juan (Hans) Fugl, bosatte sig på pampaen omkring byen Tandil, ca. 350 km sydvest for hovedstaden Buenos Aires (Bjerg 1996: 147; Hansen 2016: 6–7). Den anden udvandringsbølge fandt sted i perioden fra ca. midt 1880 til 1920, med de fleste immigranter registreret i perioden mellem 1910 og frem til første verdenskrig (Bjerg 1996: 134). Denne bølge var mere omfangsrig end den første og omfattede udvandrere fra hele Danmark, men især fra Midt- og Nordjylland og fra København. Disse udvandrere fandt området omkring Tandil opdyrket og søgte derpå længere sydpå, til landområderne omkring havnebyen Necochea, Tres Arroyos og Coronel Dorrego. Her dannede de en særdeles stærk og levedygtig dansk koloni, og området mellem byerne Necochea, Tres Arroyos og Tandil skulle blive kendt som Danskertrekanten (Bjerg 1996: 134–147; Hansen 2016: 6–7). Andre områder, hvor de danske udvandrere og deres efterkommere bosatte sig forholdsvis koncentreret, var hovedstaden Buenos Aires, byen Eldorado i den nordlige provins Misiones og byen San Carlos de Bariloche i Andesbjergene i det vestlige Argentina.

Antallet af dansk-argentinere, for hvem dansk som arvesprog stadig i dag er et aktivt sprog, er ukendt. Selvom det danske sprog stadig står relativt stærkt i de dansk-argentinske samfund, er det mest personer i den ældre generation, der stadig bruger dansk i hverdagen, og der er ikke nogen, der ikke også bruger spansk. På feltarbejdet i 2014 og 2015 fik vi det indtryk, at vi langtfra udtømte mulighederne for at finde flere end de 126 talere, som vi nåede at optage, men også at talergruppen af naturlige årsager vil svinde ind hurtigt i de næste år. For så vidt er dansk i Argentina et døende sprog. Den yngre generation (mellem 40 og 60 år) er tit opvokset med dansk som familiesprog, men for dem er spansk hurtigt blevet det dominerende sprog. Det danske, de behersker i dag, er typisk et inaktivt rudimentært husket sprog, eller et dansk, som de har lært sig (eller reaktiveret) gennem arbejde, studier og højskoleophold i Danmark.

CoSAmDa repræsenterer fortrinsvis talere af den førstnævnte type, dvs. ældre personer (60+), som er tosprogede med dansk og spansk, som har lært dansk i deres barndom, og som taler nok dansk til at kunne gennemføre en lang samtale med dansktalende interviewere. En overordnet hypotese, som

vi har fulgt i dataindsamlingen, er, at hvis der skulle have udviklet sig en særlig argentinadansk varietet af dansk, må det være i denne gruppe talere, som har brugt og stadig bruger dansk indbyrdes. Det er derfor disse talere, der har været i fokus for forskningsprojektet.[9]

I dagens Argentina bruges dansk som arvesprog kun i begrænset omfang: Det danske sprog spiller i forskellig form og forskellige sammenhænge en daglig rolle for mange af vores informanter, fx til dels som familiesprog (sammen med spansk), til sociale sammenkomster (sammen med spansk), i de danske ugeblade, der læses af danskargentinerne, i forbindelse med danske traditioner omkring fx jul, i de efterhånden sjældne danske gudstjenester og, undertiden, som hemmeligt sprog blandt andre argentinere. Men spansk er for de allerfleste blevet det dominerende hverdagssprog. Ved sociale arrangementer og i foreningsregi bruger man hovedsagelig spansk, idet der her ofte også deltager gæster og familiemedlemmer, som ikke taler dansk. Inviterede danske forelæsere, fx i tilknytning til kirken i Buenos Aires, taler naturligvis på dansk, men der er ofte tolkning til spansk. De allerfleste informanter taler et flydende, undertiden måske lidt tøvende dansk med forskellige former for og grader af afvigelser fra danmarksdansk. Sammen med vores iagttagelse af og informanternes egne beretninger om sprogbrugen i det dansk-argentinske samfund, viser dette, at de interviewede dansk-argentinere godt kan tale dansk, men de gør det ikke så meget længere.

Årsagerne til dette sprogskift skal formentlig findes i 1970'erne og 1980'erne, hvor det interne sammenhold i den dansk-argentinske gruppe og opbakningen omkring dyrkelsen af de danske værdier svækkedes, og hvor fx de dansk-argentinske skoler og højskoler blev nødt til at lukke. Mange af vores informanter beretter, at den familieinterne kommunikation skiftede til spansk med deres børns start i en almindelig argentinsk (dvs. spanskproget) skole. Fordi børnene heller ikke indgik i andre danskprogede sammenhæn-

[9] En anden type talere, der kan siges at repræsentere en anden type argentinadansk, er de dansk-argentinere, der er flyttet (tilbage) til Danmark, og som nu bor fast her i landet. Mange af dem er organiseret i Dansk-Argentinsk Forening, som er base for træf og højtidsfejring i Danmark. Disse taleres danske kunne tænkes at ligge ret nær standarddansk, men med reminiscenser af argentinadansk eller spanskpåvirkning. Dette er ikke undersøgt.

ge, blev den intergenerationelle sproglige socialisering således afbrudt. Også ægteskaber med argentinere af ikkeeuropæisk afstamning, som tidligere havde været ildeset i Danskertrekanten, blev mere acceptable i denne periode, og dette førte til en yderligere svækkelse af dansk som familiesprog. Endelig er der ikke foregået nogen bemærkelsesværdig 'fornyelse' af sprogsamfundet ved nyindvandring fra Danmark. Og sidst, men ikke mindst, har Danmark ikke understøttet de dansk-argentinske samfund finansielt og kulturelt (fx med sprogskoler), som andre nationer har gjort det med deres minoriteter i Argentina. Det er faktorer som disse, der er årsag til at det danske sprog i Argentina formentlig forsvinder inden for få år.

4.3 Danskertrekanten

Det er en del af den dansk-argentinske fortælling om Danskertrekanten, som den kommer til udtryk i de nævnte portræt- og erindringsbøger samt i Danske Stemmer-projektets optagelser, at danskerne i dette område har været driftige og udgjort et stærkt netværk, fx oprettede dansk-argentinerne tidligt et gensidigt forsikringsselskab, *Associación Mutual Dan*, til forsikring mod fx skader på afgrøder ved haglstorme. Nytilkomne danskere blev tilbudt arbejde på etablerede landbrug med henblik på at kunne tjene penge til selv at investere i et jordstykke, evt. med lån fra andre danske. Udvandrere, der ikke slog sig ned som jordbrugere, etablerede sig inden for mejeribranchen eller som håndværkere, og den dag i dag ses der i Danskertrekanten, og især omkring Tres Arroyos, mange danske firmanavne, og der er mange efterkommere af danskere blandt de store jord- og virksomhedsejere i området. Danskertrekanten har dog ikke været et slaraffenland. Som det ofte fremhæves både i litteraturen og i interviewene, har der ligget stor flid og et hårdt arbejdsliv bag den økonomiske succes. Mange har også været ofre for de vekslende økonomiske forhold i Argentina og har måttet afstå jorden eller virksomheden for at kunne ernære sig på anden vis. Dog er der i det overordnede billede ingen tvivl om, at Danskertrekanten har udgjort, og stadig i dag er, et stærkt center for dansk-argentinsk liv, sprog og kultur, ikke mindst på grund af udpræget endogami. Det danskes levedygtighed i Danskertre-

kanten er beskrevet af flere antropologer, både med et nutidigt og et historisk blik (fx Agertoft 2005, Jarnum 2006 og Erdem 2007).

Et samlingspunkt for danskerne på pampaen har været den protestantiske tro, men gudstjenesterne i de fire kirkemenigheder Tres Arroyos, Tandil, Necochea og Buenos Aires foregår i dag mest på spansk, enten af hensyn til ikkedansktalende medlemmer af trossamfundene eller fordi præsterne selv ikke taler dansk. Dog står den store protestantiske (danske) kirkegård i Tres Arroyos med dens dansk udseende kapel og de dansk-argentinske gravsten som et talende levn fra det dansk-argentinske samfund.

Det danske foreningsliv har været et andet stærkt samlingspunkt for dansk-argentinerne i Danskertrekanten. Det har omfattet bl.a. håndbold, roning, gymnastik, folkedans og kortspil. Lidt nord for Necochea, i Lumb (i øvrigt det eneste danske bynavn i området), findes samlingsstedet *Dannevirke*, hvor der stadig holdes sommerfester og andre træf.

Undervisning i og på dansk har tidligere været et vigtigt element i både børneopdragelsen og i den videre dannelse i ungdomslivet. Lidt uden for Cascallares findes den nu lukkede danske kostskole, hvor mange dansk-argentinske børn har tilbragt deres grundskoleår. Mange af de interviews, der indgår i CoSAmDa, indeholder erindringer om denne skolegang. Den danske skole i Cascallares fungerede i mange år også som base for *Los Ocho Días*, 'De otte dage', som er det årlige dansk-argentinske højskoleophold med foredrag (på dansk og spansk), kulturelle aktiviteter og socialt samvær. Næsten alle interviewede personer fra Danskertrekanten har erindringer fra *Ocho Días*-tiden, som er videreført som *Los Cinco Días* 'De fem dage' og i dag afholdes nær Tandil. I Gowland, vest for Buenos Aires, fandtes den danske højskole *Alta Mira*, der i en lang periode fungerede som base for sommer- og vinterskoler.

Af de talere, som Danske Stemmer-projektet har optaget i Danskertrekanten, indgår (i skrivende stund) 62 i CoSAmDa.

4.4 Eldorado

Eldorado i den nordlige argentinske provins Misiones ligger mellem Paraguay mod vest og Brasilien mod nord og øst. Der har boet danskere her si-

den den industrielle kolonisering af dette klimamæssigt barske argentinske jungleområde tog fat i 1920'erne. To danskere, Carl Engwald og Oluf Johansen, var med deres familier blandt de første pionerer (Hansen 2016: 13–15),[10] og det danske er stadig tydeligt i byen den dag i dag i form af mindesten, et bygningskompleks kaldet *casas de Johansen*, et hus ved navn *casa Danés*, danske familiers repræsentation i bymuseet, forsamlingshuset *Haabet*, vejnavne som *Christiansen* og *Dinamarca* og en dansk kirkegård.

De danskere, der slog sig ned i Eldorado, kom til dels fra Danmark, men for en stor dels vedkommende også fra Buenos Aires og Danskertrekanten, hvor den gode jord var blevet opkøbt i 1920'erne, og mulighederne for at etablere sig og skabe sig et økonomisk uafhængigt liv derfor var mindre. Ifølge beretningerne lod de sig lokke af udsigten til 'det grønne guld', dvs. dyrkningen af *yerba*, der bruges til den argentinske nationaldrik *mate*. Omlægningen af jungle til landbrugsland eller tømmerindustri står centralt i Eldorados og Eldoradodanskernes historie.

Vi kender ikke præcise tal, men det er tydeligt, at den dansk-argentinske gruppe ikke har været dominerende i Misiones-provinsen: Nogle af informanterne taler om 30 danske familier i 1930–40'erne. Der har aldrig været hverken dansk skole (dog privat danskundervisning), dansk menighed og kirke eller særlige danske foreninger. Derimod har især tysksprogede indvandrere præget området, der også har tiltrukket mange andre (bindestregs-) argentinere, brasilianere og paraguayanere og desuden huset en oprindelig befolkning, guaraní. Området har altså været flerkulturelt og flersproget (og er det stadig), og danskerne har tidligt integreret sig i det omgivende samfund: Nogle af vores informanter nævner, at alle 'nationaliteter' blev inviteret med til de store fester. Dog har en lille kerne holdt dansk kultur (fx julefester i *Haabet*) og det danske sprog i live. Bemærkelsesværdige er en lille gruppe kvinder i Eldorado, som mødes en gang om måneden til en danskaften, hvor de øver sig i at tale dansk.

[10] Se Engwald (1938) og Johansen (1934) for indlevende og nøgterne beskrivelser af de første hårde pionerår omkring Eldorado i Misiones-provinsen, fx Johansen (1934: 40): "dette var rigtignok ikke den frie, eventyrlige Tilværelse, jeg havde drømt om".

Vi anslår på basis af informationerne fra vores informanter, at der i skrivende stund er ca. 15 personer tilbage i Eldorado, der kan tale dansk. Heraf optræder størstedelen på Danske Stemmer-projektets optagelser og indgår i CoSAmDa.

4.5 Buenos Aires

Siden den første udvandringstid har danskere også slået sig ned i og omkring den argentinske hovedstad Buenos Aires, enten for at etablere sig i selve storbyen med dens udbud af fx industrielt arbejde, kontorarbejde og håndværksindustri, eller i Paraná-flodens delta nord for Buenos Aires, hvor mange har arbejdet med tømmerdrift. I en storby som Buenos Aires har danskerne naturligt integreret sig i det omgivende samfund, og på den måde ligner den danske gruppe her kolonien i Eldorado. Men forskelligt fra forholdene i Eldorado og mere lig dem i Danskertrekanten blev der i Buenos Aires tidligt grundlagt foreninger, fx *Den Danske Ungdomsforening* og *Den Skandinaviske Roklub* i Tigre, nord for Buenos Aires. Foreningerne spiller en rolle for kulturelt samvær om fx sport (især håndbold og roning), folkedans og amatørteater (den dag i dag er der en aktiv og turnerende dansk folkedanserforening i Buenos Aires). Centralt for mange dansk-argentinere i Buenos Aires står også *Den Danske Kirke, Iglesa Danesa*, i bydelen San Telmo (om de danske kirker i udlandet, herunder kirken i Buenos Aires, se Warburg 2015). Kirken huser ud over kirkerum også et dansk bibliotek, bl.a. med en samling af lokalhistoriske skrifter samt en fællessal, hvor der tidligere har været afholdt amatørteater og fester, og hvor der stadig afholdes sociale sammenkomster og folkedans.

Vi ved ikke, hvor mange dansk-argentinere, der i dag bor i og omkring Buenos Aires, og det er ligeledes uklart, hvor stor gruppen talmæssigt har været gennem tiderne. Det står dog fast fra fortællingerne, at det danske samfund i storbyen har haft en stadig tilførsel af dansk-argentinere fra andre steder i Argentina, som i perioder er flyttet hertil. I alt indgår 14 talere fra Buenos Aires i CoSAmDa.

5 Etablering af CoAmDa

5.1 Optageudstyr og principperne bag optagelser og digitalisering

Lydoptagelsers tekniske kvalitet har altafgørende betydning for mulighed, kvalitet, omfang og art af lingvistiske undersøgelser af talt sprog. Derfor har det ligget os meget på sinde at sikre, at de optagelser, som Danske Stemmer selv har foretaget i Argentina (CoSAmDa), er af så høj teknisk kvalitet, som det er praktisk muligt, når man laver optagelser i felten. Stort set alle optagelser er foretaget med professionelt udstyr, og vi har tilstræbt at anvende hovedbårne miniaturemikrofoner (headsets). I enkelte tilfælde har vi været nødt til at placere optagemikrofonerne med nogen afstand; typisk på et bord foran informanterne. Det gælder tilfælde, hvor der har været vanskeligheder med at tilpasse headsettet til en informant eller i tilfælde, hvor der har været for mange deltagere til, at der kunne anvendes headset på alle. Endelig er enkelte optagelser foretaget med sekundært udstyr (mobiltelefon, videokamera eller diktafon), enten pga. pludseligt opståede problemer med det primære optageudstyr eller fordi situationen påbød brug af enklere udstyr. Generelt er optagelserne i CoSAmDa af så høj teknisk kvalitet, at det ikke har givet væsentlige begrænsninger i forhold til udskrivningen eller de auditive og akustiske analyser.

Det nordamerikanske materiale, som udgør CoNAmDa, er teknisk set langt mere heterogent, fordi det er indsamlet over en lang periode og af flere uafhængige personer. Baumann Larsen og Kjærs materiale er optaget på spolebånd i tre forskellige formater, og enkelte optagelser er på kassettebånd. Tore Kristiansens og Christopher Hales data er optaget på kassettebånd, mens DANA-materialet er optaget på mini-DV digitale videobånd. De fleste af Baumann Larsen og Kjærs optagelser er foretaget på relativ kort afstand med en håndholdt mikrofon. Det giver indimellem noget forstyrrende støj fra håndteringen. Nogle optagelser kan være lavet med mikrofonen placeret i et stativ, muligvis på bordet. Både Kristiansens og Hales optagelser formodes at være lavet med en mikrofon indbygget i de båndoptagere, som er benyttet. Båndoptageren lader til at have været placeret på et bord mellem interviewere og informanter. Lyden på DANA-projektets optagelser er lavet med videokameraets indbyggede mikrofon. Fordi kameraet har væ-

ret placeret på god afstand af informanterne, lider lydkvaliteten under et stort bidrag af rumklang og anden uvedkommende lyd samt mekanisk støj fra båndtransporten i videokameraet. Sammenlignet med CoSAmDa giver den tekniske kvalitet af optagelserne i CoNAmDa på forskelligartet vis begrænsninger i forhold til både udskrivningen og ikke mindstefterfølgende auditive og akustiske analyser.

Det er værd at bemærke, at der ikke er en jævn sammenhæng mellem optagelsernes alder og deres tekniske kvalitet: Selvom de ældste optagelser fra Baumann Larsen og Kjær er over 50 år gamle, er vi ikke stødt på væsentlige problemer i forbindelse med digitaliseringen, der skyldes ældning af materialet. Vi har derimod haft visse udfordringer med at skaffe adgang til velfungerende udstyr til brug for digitaliseringen: Udstyr forældes og forfalder hurtigere, end optagelserne gør, og da der ikke længere produceres nyt udstyr til afspilning af disse historiske formater, er man afhængig af adgang til kvalificerede teknikere og reservedele. Vi har lagt vægt på at anvende udstyr til digitaliseringsprocessen, som er af høj og professionel kvalitet, og at få udstyret serviceret, så vi kan sikre os, at det lever op til de forventede standarder og ikke skader eller belaster originalmaterialet unødigt i forbindelse med afspilningen. I digitaliseringsprocessen har vi tilstræbt at afspille båndene i overensstemmelse med de standarder, de er optaget efter (fx. båndhastighed, sporbredde, equalisering, etc.), men på grund af manglende dokumentation er det i nogen grad baseret på kvalificerede skøn.[11]

Der er blevet ført en logbog, som beskriver digitaliseringen af de enkelte bånd i detaljer, og vi har løbende dokumenteret båndmaskinernes tilstand ved at overføre certificerede testbånd. For de ældre optagelser gælder det, at det pågældende medie (spole- eller kassettebånd) skaber tidsbegrænsninger. Derfor vil der sjældent være en én til én-sammenhæng mellem bånd og interview. Navnlig for de optagelser, som ligger på spolebånd, betyder det, at der typisk er flere interview på ét bånd, men også at ét interview kan være fordelt over flere bånd. En tidskrævende udfordring i etableringen af CoN-

[11] For alt materialet i CoNAmDa gælder det, at der ikke er registreret tekniske informationer om optagelserne. Vi kender derfor kun de tekniske forhold, man kan aflede ud fra optagelserne samt ud fra enkelte fotografier, som er bevaret fra indsamlingsrejserne.

AmDa har altså været at identificere samhørigheden mellem optagelse og talere.

Ud over at stræbe efter at få mest mulig af den originale information overført fra båndene, så har vi også forsøgt at undgå, at man i fremtiden skulle kunne stå med spørgsmål til optagelserne, som kun ville kunne besvares ved at afspille originalmaterialet. Derfor har et led i digitaliseringsprocessen været at fremstille en digital faksimile af hver båndside som helhed og efterfølgende identificere grænserne mellem de enkelte interview og derpå gemme hvert interview som en separat lydfil, der er identisk med den originale.

Afspilningen og digitaliseringen har så vidt muligt fulgt objektive kriterier. I de tilfælde, hvor det har været nødvendigt at 'forbedre' en optagelse (fx for at lette arbejdet med udskrivningen), ligger den bearbejdede version i korpus som en alternativ version ved siden af den ubehandlede version. Dermed er der til enhver tid mulighed for at forsøge sig med nye 'forbedringer' ud fra det originale materiale, afhængigt af de behov og de muligheder, der måtte opstå fremover. Denne distinktion mellem en primær, objektiv, version og sekundære formålsspecifikke versioner lægger sig op ad de principper, som udfoldes i Copeland (2008).

5.2 Relevante principper for opbygning af korpus

I etableringen af CoAmDa er vi blevet nødt til at tage stilling til ikke kun talesprogets idiosynkrasier, men derudover også til talernes tosprogede sprogbrug. Med henblik på at sikre muligheden for søgning på ord og ordfølger har udskrivningen fulgt standardortografien for det pågældende sprog, uanset udtale. Princippet om, at udskriften følger standardortografien, gælder også i ordinterne sprogskift som fx *cattlebåd* eller *campen* (spansk *el campo* 'landet' sammensat med dansk enklitisk bestemthedsendelse). Særdialektale ord følger dialektordbøgerne (jysk *glant*, jf. Jysk Ordbog). Dog er jysk foranstillet bestemt artikel ikke normaliseret til enklitisk suffiks. Den er uanset udtale udskrevet som *æ*.

Den anden undtagelse for dialektalt præget dansk i CoAmDa er varianter af negationen *inte* (fx *int*, *ent*, *inte*, *ente, 'nt*) som, igen uanset udtale, er ud-

skrevet som *inte*. Det har vi valgt fordi de danske dialektale negationsformer med *n* i lyd, form og funktion korresponderer med engelsk negation på verber. Således forekommer der i CoNAmDa sammentrækninger af verber og negation, hvor det er tvivlsomt, om *'nt* eller *'t* er en sammentrukket engelsk negation eller en form af *inte*. Dette er fx tilfældet med [wɑ˗(ˀ)nt], hvor det kan være svært at afgøre, om det er jysk *var inte* udtalt med [w] og og såkaldt rhoticeret /r/, engelsk *weren't* eller jysk *var* med engelsk *not* (*varn't*). Muligvis understøtter netop en sådan lighed mellem sprogene brugen af *inte* (i stedet for *ikke*), og det er derfor en bevidst beslutning pga. søgbarheden at udskrive *inte* som sådan og ikke at normalisere den til *ikke*.[12] Vi har dog efter endt udskrivning kunnet konstatere at *inte* kun optræder med 11 forekomster (produceret af fem talere), og at denne undtagelse af de generelle udskrivningsregler altså ikke har opdækket et generelt mønster af sproglig jævnførelse mellem amerikansk engelsk og dansk.

Især i sprogkombinationen engelsk-dansk har det til tider været umuligt at skelne mellem sprogene uden at ty til resursekrævende fonetiske analyser: Navnlig kognater kan i udtalen ligge så tæt op ad hinanden i de to sprog, at det er uhensigtsmæssigt at skelne – uhensigtsmæssigt, fordi en tvang til at tilskrive ordet enten det ene eller det andet standardsprog ville mørklægge den potentielle talerstrategi med at jævnføre sprogene så vidt som muligt (jf. eksemplet *w/varn't* ovenfor, andre eksempler er kognaterne *formand*/*foreman*, *fader*/*father*). Forekomster som disse er i udskrivningen blevet markeret som 'ambig', en forkortelse for engelsk 'ambiguous'. En sådan markering af et ord som 'ambig' er (i hvert fald i skrivende stund) en vurdering foretaget af udskriveren (jf. nedenfor); der ligger ikke som sådan nogle specifikke lingvistisk-fonetiske kriterier bag.

Et andet resultat af talernes tosprogethed er de hyppigt forekommende ordinterne sprogskift. Disse er markeret som 'hybrid'; en markering som på nuværende tidspunkt er mere deskriptiv end analytisk, idet den ikke tager højde for forskellene fx mellem komposita som skift mellem ordstammer (fx *attenhundred og ninety-five*, *diningrummet*, *campleje* fra spansk *el cam-*

[12] Negationen *itte* er derimod blevet udskrevet som *ikke*, siden vi har skønnet at den ikke har de samme muligheder for sproglige jævnførelse som *inte*.

po 'landet' + dansk *leje*) og skift mellem stamme og fleksionsmorfologi (fx *towelsne, ungjort, peonerne* fra spansk *peon* 'landarbejder'). Markeringen muliggør dog en forholdsvis nem søgning på ordinterne kodeskift til den videre analyse.[13]

Alle ord i CoAmDa er blevet tildelt en kategori, til dels manuelt (i udskrivningsprocessen) og til dels semiautomatisk (efter udskrivningen). Kategoriseringen indebærer, at alle ord er markeret for en af de følgende kategorier: sprog (et standardsprog), 'dialekt' (generisk markering for et dansk dialektord), 'hybrid' (ord sammensat af flere sprog), 'ambig' (sprogligt uentydigt ord eller ytring), 'ambig-hybrid' (et ord sammensat af flere sprog, hvor et element er sprogligt uentydigt) eller som en af de sprogneutrale kategorier 'egennavn',[14] 'latter', 'minimalrespons', 'udbrud', 'selvafbrydelse' eller tøven.

Den semiautomatiske kategorisering har indebåret en korrekturgennemgang af alle ord med henblik på deres kategorisering. Dette implicerer en høj grad af korrekthed i CoAmDas opmærkning for sprog, sprogblandingskategorier og andre kategorier. Ved denne opmærkning har vi således opnået et fuldstændigt og korrekturtjekket inventar af CoAmDas sproglige materiale, som i princippet udgør en ordbog for amerikadansk og argentinadansk,

[13] Der er én gennemgående undtagelse, verbet og substantivet 'farm', også i alle mulige sammensætninger og afledninger (fx *en farmer, vi farmede, farmarbejde, hestefarm*). Ordene *farm, farmer* og *at farme* er kendt i Danmark siden sidst i det 19. århundrede (kilde: ODS; bindet trykt i 1922) og som sådan kan ordformerne gælde som danske. Men ODS kommenterer "næsten kun m.h.t. forhold i engelsktalende lande", og en af kilderne til eksempler er da også Henrik Cavlings rejseberetning fra hans rejse til USA i 1890'erne (jf. *farme 1,* ODS). Konceptet er altså tydeligvis importeret fra USA. Det er derudover ikke til at vide, om CoAmDas emigranttalere har kendt ordene fra Danmark (i så fald skal ordformerne regnes som danske) eller om de har erhvervet ordet i alle dets former i USA og Canada. Denne usikkerhed har en vis betydning, idet *farm* i alle ordets varianter forekommer forholdsvis højfrekvent i CoNAmDas sproglige materiale. At kategorisere det som 'engelsk' (og tilsvarende 'hybrid', hvis det forekommer som sammensætning med et dansk element) eller 'dansk' vil derfor give sig tydeligt og muligvis misvisende kvantitativt udslag i CoNAmDas sproglige profil.

[14] Vi har kunnet konstatere, at proprier, især stednavne, er særlig udsatte for sproglig ambiguitet. Fordi proprier optræder hyppigt i CoAmDa og opmærkningen af +/-'ambig' er tids- og resursekrævende, er 'egennavne' blevet oprettet som særskilt og indtil videre sprogneutral kategori. Deres faktiske lydlige realisation i amerikadansk og argentinadansk afventer kompetente ører.

og som vil kunne danne grundlag for en sammenligning af talernes ordforråd, sprogvalg og brug af fænomener som tøven, selvafbrydelser, latter m.m.

I etableringen af CoAmDa har vi fulgt flere valideringsprocesser: For det første er der blevet foretaget en almen korrekturlytning og -læsning ved en anden end udskriveren selv. Denne korrekturlæsning er møntet på generel overholdelse af udskriftskonventionerne, men også på rettelse af stavefejl. Sidstnævnte er af stor vigtighed i forhold til korpussets søgbarhed, og vi har derfor bestræbt os på at lade studerende med gode engelsk- og spanskkompetencer lytte og læse korrektur på filer, som er udskrevet af andre med knap så høje fremmedsprogskompetencer. I en række tilfælde har korrekturlytningen/-læsningen også ledt til udskrift af passager, der har været uforståelige for udskriveren i første omgang. Både udskriverne og korrekturlæserne har gennemgået en ensartet oplæring i projektets detaljerede udskriftskonventioner med en afsluttende test af deres kompetencer. Udskriftskonventionerne er ledsaget af mange eksempler, og vi har fulgt op på eventuelle spørgsmål ved månedlige møder med hele gruppen af udskrivere og korrekturlæsere.

Desuden har vi på en wiki samlet information, som på forskellig vis var relevant for udskrivningen. Informationerne på wikien omfatter bl.a. relevante danske, argentinske, amerikanske og canadiske stednavne for at sikre konsistens i stavning og minimere tid brugt til søgning på nettet, men også afklaringer om (stavning af) dialektale ord (fx ”*førstningen* er et dansk ord, dog gammelt og dialektpræget). Det er altså ikke nødvendigvis hybriden *first-ning*, men det kan det være, ligesom det også er en oplagt kandidat til ambig”).

Derudover indeholder wikien lister over hyppigt forekommende spanske ord samt informationer om relevante institutioner (fx ”*Hogar El Atardecer*, det dansk-hollandsk plejehjem i Tres Arroyos”) og relevante lokale begreber (fx ”*estancia*, en stor gård eller gods på 10.000–15.000 hektarer; forekommer også i bestemt form: *estanciaen* eller *estancien*”), til dels også med en forklaring på sprogbrug (fx ”*La Pampa* er navnet på den provins, der ligger vest for provinsen Buenos Aires. Når dansk-argentinerne taler dansk, bruger de altid det spanske navn, når de refererer til provinsen, mens de ofte siger

pampaen (på 'dansk'), når de refererer til det som det store sletteområde, der krydser provinsgrænser"). Disse oplysninger har lettet udskrivningsarbejdet og bidraget væsentligt til ensartethed og konsistens i udskrifterne.

Sidst, men ikke mindst, har udskriverne og korrekturlæserne haft mulighed for at sætte kommentaren 'discuss' i udskriften som en markering af usikkerhed mht. hvad der skal stå i udskriften, hvilket så afgøres af forskerne.

6 Perspektiver i brug af CoAmDa

6.1 Kulturhistorisk og sociologisk forskning

Informanterne i CoAmDa taler ikke bare, de taler *om* noget, noget vigtigt: De biografiske narrativer i CoAmDa, som ofte også omfatter forældrenes eller bedsteforældrenes migrationshistorie, kan udgøre et udbytterigt datamateriale til narrativforskningen, særlig narrativer om livshistorie. De er derudover interessante ud fra et personalhistorisk perspektiv, idet de indeholder (overleverede) beretninger fra indvandrernes første tid i USA og Argentina, fortællinger om de danske netværk og traditioner i kolonierne og især om dansk-argentinernes blik på etniske og nationale tilhørsforhold. CoAmDa kan, ud fra dette synspunkt, også tjene som kilde til historieskrivning og sociologiske undersøgelser. En oplagt mulighed er at gennemføre den omvendte øvelse: At indsamle livshistorieinterview med indvandrere til Danmark for derefter at sammenligne: Hvor er der fællestræk mht. tilhørsforhold eller mangel på samme, udlængsel og hjemve, og hvor adskiller de forskellige fortællinger sig fra hinanden?

6.2 Sprogforskning

Udover de kulturhistoriske og sociologiske perspektiver giver opbygningen af CoAmDa selvfølgelig mulighed for et bredt spektrum af sproglige undersøgelser: Vi kan sammenligne, hvad indlejringen i et engelsktalende miljø betyder for det dansk, der tales, og kontrastere det med, hvordan dansk tales i spansktalende omgivelser i Argentina. Derudover er det oplagt at sammenligne de forskelligartede sociolingvistiske udviklinger i Argentina og Nordamerika, som har resulteret i tidligt sprogskift i Nordamerika over for en

vedvarende stabilitet og forholdsvist abrupt sprogskifte i Argentina. Samtidig er der i begge datasæt, men især i CoNAmDA, en vis spredning i alder og udvandringsgeneration, som gør det muligt at efterspore, hvordan migrantgeneration påvirker sproget.

Baseret på CoSAmDas interne opbygning kan man i et vist omfang undersøge de sproglige konsekvenser af at være en sproglig minoritet blandt mange nationaliteter (argentinadanskerne i Misiones) og så den form af argentinadansk, der udkrystalliserede sig i Danskertrekanten, hvor dansk nærmest var majoritetssprog blandt minoriteterne. Hvordan de danske i Buenos Aires skal passes ind i denne ligning kan man diskutere, og det er ærgerligt, at der til sammenligning findes så mangelfuld en dokumentation af danskerne i f.eks. New York og Chicago.

Desværre råder vi ikke over et helt parallelt korpus for dansk i Danmark, som kunne tjene som sammenligningsgrundlag for CoAmDas materiale (kaldt 'baseline' inden for arvesprogsforskningen), og det er en kilde til jævnlig diskussion blandt projektets medlemmer, hvordan vi kan sammenligne talere i USA og Argentina med talere i Danmark mht. dialekttræk eller mangel på samme. Her deler vi skæbne med alle andre forskere, der undersøger en sprogform hos udvandrede repræsentanter for sprogsamfundet, og hvad angår sammenligninger med (kunstige) baselinevarianter af dansk, må vi derfor inddrage dialektoptegnelser, fx i form af Korpus Cordiale, http://cordiale. dyndns.org/, og det noget yngre LANCHART-korpus (Gregersen & Kristiansen 2015). Man skal dog her holde sig for øje, at der er tale om helt forskellige betingelser for talerne: Danskere i Danmark møder ikke op til et interview med forestillinger om, at deres dansk måske ikke er tilstrækkeligt, mens det nærmest vil være en forudsætning for mange af de dansktalende udvandrere og deres efterkommere i Nordamerika og Argentina. Sidstnævnte taler jo ikke først og fremmest dansk til daglig. De må derfor i et vist omfang anstrenge sig på en anden måde end danskere, og der er ingen tvivl om, at det for nogle af dem er en sproglig anstrengelse at deltage i et interview. Netop det har givet anledning til en vis teoriudvikling i projektet og et større arbejde med empirisk baserede talerprofiler, som i sig selv

kan repræsentere en ny måde at kvantificere træk på, som kan give os indblik i hastigt eller langsomt døende sprogformer.

Udover en sammenligning af CoAmDas materiale med nutidigt og historisk talt dansk i Danmark, kan det også sammenlignes med andre foreliggende data om dansk talt andre steder end i Danmark, nemlig på Færøerne og i Sydslesvig (Kühl 2014, 2015a, 2015b). En sådan større sammenlignende forskning om dansk i dets forskellige kontekster vil kunne vise, hvilken betydning sprogsociale forhold har for sprogforandring, og hvilke sproglige træk der ændres i hvilken type af sprogkontaktsituationer – mere specifikt, hvilke sprogtræk der overtages fra det omgivende sprog, hvilke varianter der ændrer fordeling under indflydelse af det omgivende sprog, og hvordan disse processer er relateret til genetisk og typologisk slægtskab mellem sprogene (dansk, engelsk, tysk og færøsk som genetisk og typologisk nært beslægtede germanske sprog vs. kontakt mellem dansk og det romanske sprog spansk). Således vil man kunne indkredse, hvad der udgør danmarksdansk talesprog, men også vise dansk talesprog i mange af dets diatopiske varianter uden for Danmark. Hermed nærmer vi os et svar på spørgsmålet om, hvad der definerer dansk, dvs. hvad der altid må være til stede, for at vi overhovedet kan tale om en dansk varietet.

Litteratur

Agertoft, Per 2005: *Danskheden på pampaen 1875–1950. En undersøgelse af aktører, strukturer og diskurser i udviklingen af en dansk-argentinsk i provinsen Buenos Aires*. Speciale. Aarhus Universitet, Aarhus.

Baumann Larsen, Mogens 1973: The Function of Dialects in the Religious Life of the Bornholm Inhabitants, *Internationales Jahrbuch für Religionssoziologie/International Yearbook for the Sociology of Religion* Vol. VIII: II Religion und Sprache/Religion and Language. 236–243.

Baumann Larsen, Mogens 1981: Dansk i USA – et forskningsprojekt. Ålborg Universitetscenter, Institut for sprog, kommunikation og kulturhistorie. Utrykt rapport. Ålborg.

Berman, Ruth A. & Dan I. Slobin (eds.) 1994: *Relating Events in Narrative*. Opr. Lawrence Erlbaum; fotografisk genoptryk Psychology Press 2012.

Bjerg, María M. 1996: *Entre Sofie y Tovelille. Una historia de los immigrantes daneses en la Argentina (1848–1930).* Buenos Aires: Editorial Biblos.

Bredmose Simonsen, Henrik 1990: *Kampen om danskheden. Tro og nationalitet i de danske kirkesamfund i Amerika*. Århus: Aarhus Universitetsforlag.

Bækhøj, Lars 1948: *Danske i Argentina*. Buenos Aires: Kriegers Forlag.

Copeland, Peter 2008: *Manual of Analogue Sound Restoration Techniques*. London: The British Library.

Engwald, Nina Raben 1938: *Eldorado*. København: Hasselbach.

Erdem, Adil 2007: *Indvandrernes paradis: Danske rødder i Argentina*. Odense: Geografforlaget.

Glargaard, Amalie 2015: *Et frøhop ind i det dansk-argentinske arvesprog. En empirisk undersøgelse af dansk-argentinske frøhistorier*. Speciale. Institut for Nordiske Studier og Sprogvidenskab, Københavns Universitet.

Gregersen, Frans & Tore Kristiansen 2015: Indledning. Sprogforandring i virkelig tid. Frans Gregersen & Tore Kristiansen (red.) *Hvad ved vi nu – om danske talesprog?* København: Sprogforandringscentret, Københavns Universitet. 9–45.

Grøngaard Jeppesen, Torben 2005: *Danske i USA 1850–2000. En demografisk, social og kulturgeografisk undersøgelse af de danske immigranter og deres efterkommere*. Odense: Odense Bys Museer.

Grøngaard Jeppesen, Torben 2010: *Skandinaviske efterkommere i USA. Etniske grupper eller kerneamerikanere?* Odense: Odense Bys Museer.

Hansen, Nadia 2016: En snert af dansk mellem urskov og pampa. Det danske i to danskerkolonier i Argentina. Rapport. Redigeret af Jan Heegård Petersen og Karoline Kühl.
Tilgængelig på http://danskestemmer.ku.dk/resultater/publikationer/

Heegård Petersen, Jan under udgivelse: Strengthening of dialecticity: Phonological conservatism and variation in a former Danish colony in South Dakota. *Journal of Germanic Linguistics*.

Heegård Petersen, Jan, Jacob Thøgersen & Karoline Kühl 2015: Holdninger til dansk blandt nyindvandrede danskere i Californien: En afrapportering af en holdningsundersøgelse. København.
Tilgængelig på http://danskestemmer.ku.dk/resultater/publikationer/.

Hvidt, Kristian 1971: *Flugten til Amerika eller drivkræfter i Masse-udvandringen fra Danmark 1868–1914*. Aarhus: Universitetsforlaget i Aarhus.

Jarnum, Katja 2006: *Mellem Danmark og Argentina*. Speciale. Institut for Antropologi, Københavns Universitet.

Johansen, Oluf 1934: *Nybygger*. København: Hasselbach.

Kjær, Iver & Mogens Baumann Larsen 1974: 'De messy ting'. Om kodeskift i dansk-amerikansk. Poul Andersen, Christian Lisse, Karl Martin Nielsen, Hanne Ruus, John Kousgård Sørensen (eds.) *Festskrift til Kristian Hald. På halvfjersårsdagen 9.9.1974*. København: Akademisk Forlag. 421–430.

Kjær, Iver; Baumann Larsen, Mogens 1981: Dansk i Dakota. *Årsskrift 1981*. Carlsbergfondet. København. 25–31.

Kristiansen, Tore, Jake Harwood & Howard Giles 1991: Ethno-linguistic vitality in 'the Danish capital of America'. *Journal of Multilingual and Multicultural Development* 2(6). 421–448.

Kühl, Karoline 2014: Danske aflæggere i Nordeuropa: Færødansk, sydslesvigdansk og bokmål. *Danske Talesprog* 14. 29–54.

Kühl, Karoline 2015a: South Schleswig Danish. Caught between privileges and disregard. Rudolf Muhr, Dawn Marley, Heinz L. Kretzenbacher & Anu Bissoonauth (eds.) *Pluricentric Languages. New perspectives in theory and description*. Wien: Peter Lang, 227–240.

Kühl, Karoline 2015b: Faroe Danish. An unknown variety. Eivind Torgersen, Stian Hårstad, Brit Mæhlum & Unn Røyneland (eds.) *Language variation – European perspectives V. Selected papers from the Seventh International Conference on Language Variation in Europe (ICLaVE 7), Trondheim, June 2013* [Studies in language variation 17]. Amsterdam/Philadelphia: Benjamins. 157–168.

Kühl, Karoline 2015c: ’Det er easy at tale engelsk også’. Amerika-dansk i 1960’erne og 1970’erne. *NyS* 47. 39–64.

Lang, Erik John Nielsen 2005: *New Denmark, New Brunswick: New approaches in the study of Danish migration to Canada 1872–1901*. MA thesis. Carleton University, Ottawa. Department of History.

Larsen, Birgit Flemming & Henning Bender (eds.) 1992: *Danish emigration to the U.S.A.* Aalborg: Danes Worldwide Archives [Udvandrerhistoriske skrifter, 4].

Ordbog over det danske sprog (ODS) 1918–1956. [Historisk ordbog 1700–1950]. København: Det Danske Sprog- og Litteraturselskab.

Warburg, Margit 2015: De danske kirker i udlandet, en religionsdemografisk model for migrantmenigheder. *Religionsvidenskabeligt Tidsskrift* 62. 45–60.

Worgall, Lena 2011: *Danske udvandrerkvinders liv i Argentina i perioden 1890–1935. En analyse med fokus på baggrunden for og optakten til udvandringen, daglige, familiære, kulturelle, emotionelle og sociale aspekter, indbefattet udfordringer og forhold kvinderne har mødt*. Speciale. Institut for Historie, Kultur og Samfundsbeskrivelse, Syddansk Universitet.

Karoline Kühl
Lektor, ph.d.
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet
karoline.kuehl@hum.ku.dk

Jan Heegård Petersen
Lektor, ph.d.
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet
janhp@hum.ku.dk

Gert Foget Hansen
postdoc, ph.d.
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet
gertfh@hum.ku.dk

Frans Gregersen,
professor, dr.phil.
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet
fg@hum.ku.dk