

”Mit sprog er en hybrid: norske ord og sætninger i en halvdansk indpakning” - sproglig tilpasning til norsk blandt danskere i Norge som sociolingvistisk eksempel på brug af korrespondanceanalyse

Randi Benedikte Brodersen

Forskningsspørgsmål, formål, afgrænsning og korrespondanceanalyse

Hvad sker der med danskeres sprog i mødet med norsk og nordmænd? Hvordan, hvor meget og hvorfor tilpasser danskere sig sprogligt til norsk? Er der forskel på, hvordan danskere tilpasser sig i Bergen, Kristiansand og Oslo? Er der tilpasningsforskelle knyttet til uddannelse, alder, alder ved flytning til Norge, antal år bosat i Norge, civilstatus og køn? Og hvilken rolle spiller national identitet? Disse spørgsmål ønsker jeg at belyse, og det gør jeg i denne artikel ved hjælp af korrespondanceanalyse.

Som det fremgår af titlen og forskningsspørgsmålene ovenfor, er mit ene emne her danskeres sproglige tilpasning til norsk i Norge. Mit andet emne er korrespondanceanalyse, som er navnet på den statistiske metode, jeg bruger i min analyse.

Citatet i titlen er en sproglig selvkaraktistik fra en af mine danske informanter, der som så mange andre danskere i Norge tilpasser sig leksikalsk ved at bruge norske ord og udtalemæssigt ved at ændre sin udtale i retning af norsk, men beholder et dansk grundprog, et dansk substrat, så sprogformen bliver ”en halvdansk indpakning”. Sproglig tilpasning er et af mine emner i en større sociolingvistisk undersøgelse hvor jeg bl.a. har indsamlet data ved at lade en norsk interviewer telefoninterviewe danske informanter fra Bergen, Kristiansand og Oslo og ved at få baggrundsplysninger fra informanterne ved brug af et spørgeskema. Brug af norsk interviewer er en metodologisk pointe og en enkel måde at få adgang til data som egner sig til at belyse danskeres tilpasning til norsk i Norge, fordi det giver en kommunikationssitu-

ation hvor de danske informanter tilpasser sig til norsk, som danskere i Norge også gør ellers i deres daglige kommunikation med nordmænd.



Figur 1. De tre norske informantbyers beliggenhed: Kristiansand ved sydkysten, Bergen ved vestkysten og Oslo i øst, inderst i Oslofjorden¹ (Kartverket 2009)

¹ Luftlinje-afstanden fra Bergen, Kristiansand og Oslo til havnebyen Hirtshals i Nordjylland er henholdsvis ca. 600, 140 km og 300 km.

Den sproglige tilpasning drejer sig om leksikalisk og udtalemæssig tilpasning, det vil sige: Hvor stor en andel norske ord bruger de danske informanter i telefoninterviewet med den norske interviewer og hvor stort er omfanget af udtalemæssig tilpasning i et uddrag af interviewet, eksemplificeret ved ændringen af udtalen fra lyse danske *a*-lyde til mørke norske *a*-lyde (om min begrundelse for valg af netop denne vokal se Brodersen 2010: 219).

Formålet med denne artikel er dels at vise brugen af multivariat korrespondanceanalyse i en sociolingvistisk undersøgelse, dels at udbrede kendskabet til denne variant af korrespondanceanalysen som et stærkt alternativ og et nyttigt supplement til traditionelle statistiske metoder, som fx korrelations- og regressionsanalyse. Især korrelationsanalyse er meget brugt i sociolingvistiske undersøgelser. Formålet er ikke at sammenligne multivariat korrespondanceanalyse, som er en form for geometrisk dataanalyse (GDA), med traditionelle statistiske metoder som korrelationsanalyse eller regressionsanalyse. Min behandling af traditionel statistik er derfor sparsom. Men jeg vil allerede nu nævne at der findes en sammenligning af korrespondanceanalysen og korrelationsanalysen, og den omtaler jeg.

Forhåbentlig kan et konkret eksempel på brug af korrespondanceanalysen og en kortfattet præsentation af denne metode inspirere andre til at bruge den og være til nytte for læsere der ønsker at vide, hvordan korrespondanceanalysen fungerer, og hvordan vi tolker analysens grafiske resultat som vi ser på et såkaldt korrespondanceanalysekort. Samtidig kan min brug af multivariat korrespondanceanalyse i en sociolingvistisk kontekst måske også bringe metoden og dens visuelle helhedsanalyse og andre fordele ud til en bredere kreds af kultur-, samfunds- og sprogforskere i Danmark og Norden. Multivariat korrespondanceanalyse kan give os helhedsfokuserede analyser og en anden forståelse af (sociolingvistiske) data end fx korrelationsanalyse og vise hvordan flere variabler (end to) samvirker. En korrespondanceanalyse af sociolingvistiske data kan måske åbne vores øjne for fordelene ved og behovet for at kombinere og supplere traditionelle bivariate eller multivariate metoder som fx henholdsvis korrelationsanalyse og regressionsanalyse med en multivariat geometrisk metode.

Artiklen er også et led i min tilegnelse og forståelse af (multivariat) korrespondanceanalyse ud fra min sociolingvistiske baggrund, og denne metode har jeg valgt at bruge som statistisk analysemetode på grund af dens særlige egenskaber og muligheder, heriblandt dens indbyggede helhedsfokus på data, dens relationelle ”tænkemåde” og dens visuelle dataoverblik. I

øvrigt kombinerer jeg korrespondanceanalysen med andre analysemetoder og får bl.a. gennem informantcitater fra telefoninterview og fra skriftlige sprogbiografiske beretninger – hvor informanterne bl.a. fortæller om deres liv som danskere i Norge og deres sproglige tilpasning – dokumentation og belæg, som støtter mine korrespondanceanalytiske resultater.

I dansk sociolingvistik er korrespondanceanalysen ny. I nordisk sociolingvistik er metoden, mig bekendt, ind til videre kun blevet brugt i to større sociolingvistiske undersøgelser, foruden min egen (Knutsen 2007 og Rønne 2005). I sociologien er korrespondanceanalysen velkendt og veletableret (se fx Brodersen 2010: 228), og flere indføringer i korrespondanceanalyse findes på norsk, af nordiske sociologer: Clausen 1989, Hjellbrekke 1999 og Rosenlund 1995. Den nyeste indføring på dansk er Høyen 2010. Af indføringer på engelsk kan nævnes: Clausen 1998, Greenacre 1993 og Le Roux & Rouanet 2010. Gratis software som kan bruges, er R, på siden *Statistics with R*: http://zoonek2.free.fr/UNIX/48_R/all.html. Andre programmuligheder er SPAD, SPSS, SAS og Excel (Høyen 2010: 95) og Analytica Toolbox for MATLAB, som er udviklet af den franske forsker Daniel Apollon, Universitetet i Bergen, og som er brugt i min analyse senere i artiklen.

Artiklens indhold og opbygning er todelt. Første del drejer sig om korrespondanceanalysen og begynder med en kort og kontekstualiserende sammenligning af geometrisk og traditionel statistik og af korrespondanceanalysen og korrelationsanalysen. Derefter følger en karakteristik af korrespondanceanalysen, dens fordele og udfordringer og en vurdering af hvad den giver sociolingvistikken. Anden del drejer sig om min undersøgelse af danskeres tilpasning til norsk og min første korrespondanceanalyse af denne tilpasning. Først præsenterer jeg min undersøgelse og de data der indgår i korrespondanceanalysen. Derefter kommer selve korrespondanceanalysen af 49 danskeres tilpasning, og jeg foretager en visuel inspektion og en tolkning af mine korrespondanceanalysekort. Afslutningsvis vender jeg tilbage til de indledende spørgsmål ovenfor og bringer en samlet opsummering af artiklens to dele.

Geometrisk og traditionel statistik

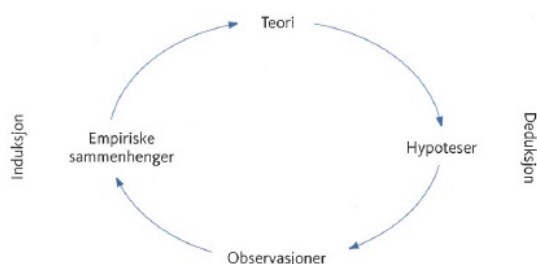
Fordi geometrisk dataanalyse (GDA) stadig er ny og ukendt for mange, vil jeg nu kort nævne centrale ideer og kendetegn der er forbundet med denne type analyse. Jeg nævner også enkelte paralleller og forskelle i forhold til traditionel statistik.

I GDA – som korrespondanceanalysen er et eksempel på – er de tre centrale ideer:

- *Geometrisk modellering.* Vi begynder med en tovejstabel med rækker og kolonner, med informanter og variabler, og tabellens enheder bliver punkter i et geometrisk rum, og vi får to punktskyer, én med informanter og én med variabler, som i figur 4, 5, 6 og 7 nedenfor.
- *Formel tilgang.* Teknikken bygger på matematisk teori i form af lineær algebra.
- *Beskrivelse først!* Filosofien er induktiv, og resultaterne af geometriske metoder er *deskriptiv statistik*, dvs. at resultaterne ikke afhænger af datasættets størrelse (Le Roux & Rouanet 2010: 1–2).

Geometrisk og traditionel statistik repræsenterer forskellige måder at forholde sig til data på, og til videnskabens verden. Geometrisk statistik er deskriptiv, udforskende og åben, og vi har ikke et enten-eller-svar parat på forhånd. Det induktive grundlag kommer også til udtryk grafisk idet alle analysens variabler og informanter er repræsenteret visuelt, og alle mønstre i et materiale kommer frem i lyset. Formålet er (helheds)beskrivelse og udforskning – og desuden grafisk gengivelse af numerisk information for at lette tolkningen af informationen. Traditionel statistik, det vil sige den dominerende angloamerikanske statistik, repræsenterer en hypotetisk-deduktiv tænkemåde, er konfirmerende, lineær og styret af forhåndsantagelser og krav til variabelers fordeling, fx kravet om normalfordeling. Formålet er ofte hypotesetestning.

I videnskabscirklen nedenfor hører geometrisk statistik hjemme i den databaserede induktive afdeling i venstre side, og traditionel statistik i den teoristyrede deduktive del i højre side.



Figur 2. Videnskabscirkel (Halvorsen 2008: 24, efter Ringdal 2001: 68)

Et fællestræk ved de to typer statistik er udgangspunktet i en tabel med rækker og kolonner, en data-matrix. Men der stopper også enhver lighed mellem geometrisk og traditionel statistik. Går vi ind i korrespondanceanalysen, går vi ind i en geometrisk flerdi-

mensional tænkemåde og videre ind i et rum med et ubegrænset antal dimensioner. Vi bevæger os fra en tovejs tabel til et rum med punkter, og tabellens informationer bliver transformeret til punkter i et flerdimensionalt rum.

Resultatet af geometrisk dataanalyse er dels statistiske tabeller med analysens talmateriale, dels et visuelt billede af data, fx et korrespondanceanalysekort, hvor data er repræsenteret som punkter i en punktsky i et optimalt lavdimensionalt rum, eller rettere sagt: to rum med hver sin punktsky, men samlet på ét korrespondanceanalysekort.

Værd at nævne er også to ofte stillede spørgsmål om geometriske metoder og dermed korrespondanceanalysen:

1. Are geometric methods amenable to statistical inference?
 2. Can they be used for explanatory reasons?
- The answer to both questions is "Yes, definitely."
(Le Roux & Rouanet 2010: 12)

I tilknytning til det første spørgsmål er det relevant at fremhæve den gyldne regel: "that inference procedures should always be subordinate to substantive conclusions. Thus the model of data, already enriched by structured data analysis, will be further substantiated by inductive data analysis", i samsvar med denne skematiske fremstilling af analyseprocessen:

Tabel → Punktsky → Strukturer → Inferens
(Le Roux & Rouanet 2010: 12).

Traditionel statistik er måske nok umiddelbart "enklere" at have med at gøre end geometrisk statistik, fordi mange sociolingvister er socialiseret ind i en amerikansk og labovsk statistik-tradition (og bruger samme type statistik som sociolingvistikens grundlægger William Labov) og derfor er fortrolige med statistiske tabeller, signifikansværdier og testresultater. En anden faktor som også spiller en rolle er, at den traditionelle statistik er den eneste mange af os kender og er vant til, hvis vi bruger statistiske metoder. Det er ofte en udfordring at tænke anderledes og gå imod strømmen og traditionen og vælge en anden analysemetode end den gængse, men ved at vælge en geometrisk teknik som fx korrespondanceanalysen bidrager vi til berigelse, fornyelse og videreudvikling af en disciplins metodegrundlag samtidig med at vi stimulerer andre til nytænkning.

Korrespondanceanalysen og korrelationsanalysen

Korrespondanceanalysen er som nævnt endnu ret ny og ret ukendt i sociolingvistikken, og i andre sprogvidenskabelige discipliner, og jeg bruger derfor i dette afsnit mere plads på korrespondanceanalysen end på den i sociolingvistikken veletablerede korrelationsanalyse,² og siden den norsk-svenske sociolog Lennart Rosenlund har sammenlignet korrespondanceanalysen og korrelationsanalysen, citerer jeg de centrale punkter i Rosenlunds sammenligning.

Korrespondanceanalysen kan bruges på alle typer data og variabler, der er homogene og kategoriske og forudsætter ingen normalfordeling af data. Korrespondanceanalysen behandler alle typer data på samme måde, dvs. den skelner ikke mellem informanter og variabler og ikke mellem afhængige og uafhængige variabler, og alle variabler, både sproglige og sociale variabler, bliver behandlet som egenskaber:

Den store og viktige forskjellen mellom denne form for dataanalyse [dvs. korrespondanceanalyse, RBB] og de teknikker vi [dvs. sociologer, RBB] vanligvis bruker (korrelasjonsanalyse, faktoranalyser, regresjonsanalyse og beslektede teknikker) er at korrespondanceanalysen i utgangspunktet behandler hver variabels ulike verdier som selvstendige statistiske kategorier. I den franske terminologi kalles disse for modaliteter. Analysen avdekker de statistiske relasjonene mellom disse modalitetene (alle variabelers alle verdier) (Rosenlund 1995: 65).

Sammenligner vi de to metoder, udmærker korrespondanceanalysen sig ved at være fleksibel og ved at kunne afdække mønstre i et materiale og relationer og sammenhænge mellem mange variabler samtidig, mens korrelationsanalysen tester sammenhængen mellem to eller evt. tre variabler ad gangen. Der findes andre traditionelle statistiske metoder til at undersøge sammenhænge mellem flere variabler samtidig, fx regressionsanalyse. Et problem og en ulempe ved begge disse metoder er imidlertid at de er lineære. Derfor konkluderer Lennart Rosenlund i sin sammen-

ligning af korrespondanceanalyse og de traditionelle metoder korrelationsanalyse og regressionsanalyse

at korrespondanceanalysen er overlegen korrelationsanalysen når det gjelder å beskrive strukturer i et datamateriale og å frilegge samspillseffekter mellom analysevariablene. Utfra den kunnskap korrespondanceanalysen gir, kan vi også slutte oss til at en videre analyse av disse variabler, uten denne kunnskap, basert på lineære metoder – for eksempel en tradisjonell regresjonsanalyse – ville kunne føre galt av sted og gi feilaktige resultater (Rosenlund 1995: 72).

Et sidste vigtigt punkt som fortjener at blive nævnt, når vi sammenligner korrespondanceanalysen og korrelationsanalysen, er den arbejdsmæssige besparelse som korrespondanceanalysen giver i forhold til korrelationsanalysen: Hvis vi skulle udføre korrespondanceanalysens

kortleggingsarbeid på tradisjonelt vis, ville arbeidet bestå i å vurdere analyser av 28 krysstabeller med tilsammen 1347 enkelte frekvensangivelser [i Rosenlunds eksempel, RBB]. Videre måtte sammenhengene i disse tabeller sjekkes for eventuelle samspillseffekter, ved å kontrollere for innvirkning av tredje variabel. Kartet [dvs. korrespondanceanalysekortet, RBB] kan [...] på en enkel måte avsløre hvor i materialet vi vil finne de viktige interaksjonseffektene mellom variablene (Rosenlund 1995: 70).

Korrespondanceanalysen afdækker samspil og relationer mellem alle variabler og får et mangefold af relationer og nuancer frem og giver os et sammenfattende helhedsbillede af data. Det er specielt værdifuldt, når vi har at gøre med et forholdsvis kompleks datamateriale som i mit analyseeksempel nedenfor med 49 informanter og 51 variabelkategorier. Her kommer traditionelle metoder til kort.

En kort karakteristik af korrespondanceanalysen

Modellen må følge data, og ikke omvendt. Dette er [...] eit leitt³ trekk ved tilpassinga av matematikken til samfunnsvitskapane; ein overflod av a priori spesifiserte modellar, som deretter vert

² Korrelationsanalyse er en form for bivariat analyse som har til formål at finde ud af om der er en lineær sammenhæng mellem to variabler eller ej. Sammenhængen kan være positiv eller negativ og måles ved hjælp af den lineære korrelationskoefficient som ligger mellem 1 og -1, og hvor 1 er perfekt positiv sammenhæng, og -1 perfekt negativ sammenhæng, og 0 betyder ingen sammenhæng. Korrelationsanalysen forudsætter en lineær sammenhæng mellem variabler. Emnet korrelation er behandlet i alle indføringer i traditionel statistik.

³ Ordet *leitt* (tt), fra det norrøne *leið*, kan oversættes med "kedelig", jf. *Bokmålsordboka* og *Nynordordboka* 'ubehageli(e)g', 'kjedelig', 'pinli(e)g': www.nob-ordbok.uio.no.

konfronterte med data gjennom det ein kallar testar. Snart tener testen til å rettferdiggjere ein modell på eit felt der det finst fleire parametarar å tilpasse enn dei ein sjølv har fastsett ut frå data. Han tener til å strengt forkaste dei klokaste innvendingane ein forskar kan ha. Men det vi har bruk for, er ein rigorøs metode som er i stand til å trekke ut strukturer med utgangspunkt i data (Benzécri 1973II: 6, citeret efter og oversat af Hjellbrekke 1999: 8).

Det skarpe citat giver en karakteristik af den metode som samfundsvidenskaberne har brug for – og også sociolingvistikken. Metoden er induktiv, dvs. databaseret: "Modellen må følge data, og ikke omvendt" og være "i stand til å trekkje ut strukturer med utgangspunkt i data" og være fri for testtænkning. Citatets ophavsmand er den franske matematiker og statistiker Jean Paul Benzécri som i 1960'erne ønskede at lave "ein induktiv og eksplorerande teknikk for analyse av lingvistiske data, i protest mot datidas deduktivistiske lingvistik", som især Noam Chomsky repræsenterede (Hovden 1997: 1). Den omtalte teknik er korrespondanceanalysen. Navnet er en paraply for en hel familie af geometriske teknikker, og der findes både bivariate og multivariate varianter. Den første type kan, som navnet antyder, kun udforske sammenhænge og relationer mellem to variabler ad gangen, mens den anden håndterer flere variabler ad gangen.

Korrespondanceanalysen er især blevet kendt og anerkendt internationalt i 1980'erne da den franske sociolog og antropolog Pierre Bourdieus vigtige værker *La distinction* (1979) og *Homo academicus* (1984) udkom på engelsk, i henholdsvis 1984 og 1988. I øvrigt er det førstnævnte værk stadig det mest kendte eksempel på brug af korrespondanceanalysen. De engelske oversættelser af Bourdieus værker fortjener at blive nævnt eftersom de uden tvivl har haft stor betydning for korrespondanceanalysens internationale gennembrud, og det samme har udgivelsen af engelsksproglige indføringer og integreringen af teknikken i store statistikpakker (Greenacre & Blazius 1998: viii). Den første engelsksproglige indføring i korrespondanceanalyse var den sydafrikanske statistiker Michael Greenacres *Theory and Applications of Correspondence Analysis* (Greenacre 1984), og den seneste er de to franske matematikere Brigitte Le Roux og Henry Rouanets *Multiple Correspondence Analysis* (Le Roux og Henry Rouanet 2010). Andre engelsksproglige indføringer er nævnt i Hjellbrekke 1999 og Høyen 2010.

Vi har i samfundsvidenskaben og i sociolingvistikken og andre menneskevidenskaber brug for en

statistisk metode som "følger data" og "trækker strukturer ud med udgangspunkt i data", for at gentage og fremhæve Benzécri's ord i citatet ovenfor, og som er i stand til at afbilde sammenhænge mellem mange variabler samtidig, da mange af os ofte arbejder med komplekse og kvantitative problemstillinger og mange variabler, der samvirker.

Komplekse problemstillinger nyder godt af kombinerede tilgange. Vi har derfor brug for både fransk geometrisk statistik – som er en hel familie af metoder, og heriblandt forskellige varianter af korrespondanceanalyse – og traditionel angloamerikansk statistik med metoder som bl.a. korrelations- og regressionsanalyse. Vi vil ofte have udbytte af at kombinere forskellige typer metoder da de jo har forskellige muligheder, styrker og svagheder. Et godt eksempel er Brinkkjær 2007 og 2010 som viser hvordan geometrisk dataanalyse og regressionsanalyse kan integreres. Et andet godt eksempel er nævnt i Rosenlund 1995 som viser hvordan korrespondanceanalysen kan bruges som redskab i en "'visuell' regresjonsanalyse" (s. 72-74).

Som nævnt indledningsvis er korrespondanceanalysen velkendt i sociologien, men i indføringer i fagområdet statistik er geometrisk statistik og dermed korrespondanceanalysen ukendt. Statistik er traditionel angloamerikansk statistik. At traditionel angloamerikansk statistik har så dominerende en position, er nok en del af forklaringen på at korrespondanceanalyse og andre geometriske teknikker har (eller har haft) svært ved at vinde terræn. Hvordan situationen er for geometrisk statistik i 2010 (da bogen *Multiple Correspondence Analysis* udkom, i øvrigt efter en lang og besværlig tilblivelse), viser den opsummering som Brigitte Le Roux og Henry Rouanet giver i deres indledning:

[I]n the field of multivariate statistics, geometric methods remain rather isolated. For MCA [= multiple correspondence analysis, RBB], the situation is really poor. This method which is so powerful for analyzing full-scale research studies is still rarely discussed and therefore underused (Le Roux & Rouanet 2010: 3).

Brigitte Le Roux og Henry Rouanet peger på, at geometriske metoder, og især multivariat korrespondanceanalyse, fortsat er temmelig isolerede og dermed forsømte, oversete og underbrugte. Der er dog i Sverige, Norge og Danmark stærke miljøer, især sociologiske, hvor korrespondanceanalysen bruges, trives og formidles, og selvom vi der bruger korrespondance-

analyse, ikke er en talrig skare, så er det mit indtryk at interessen er stigende.

Korrespondanceanalysen er, som jeg har nævnt ovenfor og som vi kan læse i citatet nedenfor, en grafisk teknik som oversætter numerisk information til et grafisk billede, med tanke på at lette tolkningen:

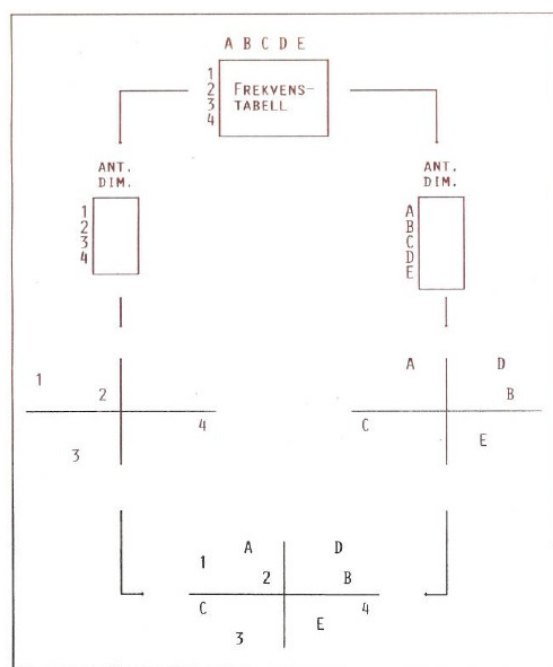
The primary goal of correspondence analysis is to transform a table of numerical information into a graphical display, facilitating the interpretation of this information. This goal is shared by such familiar graphical techniques as histograms, box-plots, star diagrams and various types of scattergrams. The aim of all these methods is to communicate numerical information by expressing it in a different form. These techniques are all explanatory in the sense that they describe [...] the data (Greenacre 1994: 8).

Resultatet af korrespondanceanalyse-kørslen er deskriptiv statistik, i teknisk forstand på den måde at analysens resultat ikke afhænger af datasættets størrelse, og det betyder at hvis vi fx doubler antallet af informanter, forbliver punkterne i punktskyerne uændrede (Le Roux & Rouanet 2010: 2).

Analysen giver os som nævnt et grafisk helhedsbillede af data i form af det karakteristiske korrespondanceanalysekort hvor data er afbildet som punkter, i en såkaldt punktsky, som vi ser nedenfor i figur 4, 5, 6 og 7. Begrebet *punktsky* kræver en kommentar. Det henviser til resultatet af den analyseproces som transformerer den numeriske information – som består af tal i rækker og kolonner (med informanter og variabler) – til vektorer i et multidimensionalt rum, eller rettere sagt, to rum, ét med informantpunkter og én med variabelpunkter. *Vektor* er et matematisk begreb som betyder en størrelse med en vægt eller masse og en retning. Og disse vektorer bliver til en sky af punkter i rummet på kortet, og hvert punkt har en bestemt afstand og retning i forhold til *origo*, som er rummets midtpunkt. Målet på punktskyens spredning og tyngde kaldes *inerti*. Inertien består af bidrag fra alle punkter, og hver dimension/akse i det resulterende kort dækker eller dekomponerer en andel af inertien. Første dimension mest, anden dimension næstmest osv.

De to ovennævnte multidimensionale rum har lige så mange dimensioner som der er kolonner og rækker i den tabel der ligger til grund for analysen. Vi kan bare ikke forestille os et multidimensionalt rum, fordi vores menneskeøjne kun kan se et rum med tre dimensioner. Men matematikken opererer ubesværet med et ubegrænset antal dimensioner. For at komme fra to multidimensionale rum til to lavdimensionale

rum foretager analysen en reduktion i dimensionalitet ved at reducere de to multidimensionale rum til to lavdimensionale rum. Det sker ved at analysen beholder den vigtigste information og fjerner mindre vigtig information med det mål for øje at give den bedste beskrivelse af alle punkter ved så få dimensioner eller akser som muligt. De to rum er så ens (Clausen 1989: 26) at de kan føjes sammen til ét fælles rum, med både informanter og variabler. De korresponderer. Deraf navnet korrespondanceanalyse. Analyseprocessen er vist her:



Figur 3. Illustration af analyseprocessen (Clausen 1989: 14)

Som vi ser i figuren ovenfor, projicerer analysen de to rum ind på ét og samme kort og viser informanter og variabler "oven i" hinanden. Netop derfor kan vi opdage, udforske og vise mønstre, modsætninger, relationer og sammenhænge som traditionelle statistiske metoder ikke giver mulighed for, i en så tydelig og helhedspræget visuel form. Det er metoden der i kraft af sin teknik afdækker mønstre, relationer og modsætninger som resultat af analysen, men det er naturligvis forskeren der læser og beskriver kortet og tolker analysen og de afdækkede mønstres og relationers betydning og som i en tidligere fase i processen har valgt og kodet og omkodet variabelkategorier mm. Forskerens valg, kodning, omkodning, data- og metodeindsigt og arbejde med at raffinere analysen er naturligvis afgørende for analysens kvalitet.

I geometrien skelner man mellem et rum, et plan og en linje. Et *rum* har tre dimensioner (i et koordinatsystem med en x-, en y- og en z-akse), et *plan* har to dimensioner, og en *linje* har én dimension. I korrespondanceanalytisk praksis taler vi dog også om et rum, når vi kun har to dimensioner, knyttet til hver sin akse på kortet. Og det er et bredere og større rum. Det er også et socialt rum. Går vi fx til Bourdieus korrespondanceanalysekort i *Distinktionen*, finder vi et socialt hierarki med akademikere i toppen og arbejdere i bunden og en tilsvarende fordeling af kapital (Bourdieu 1995: 35).

Den grafiske repræsentation af data – som hviler på et geometrisk fundament – er det vigtigste og visuelle særtræk ved korrespondanceanalysen og et af de træk som statistikerne Michael Greenacre nævner i en 4-punkts karakteristik af korrespondanceanalysen. De øvrige kendetegn er: 2) den induktive filosofi som er knyttet til et ønske om at beskrive data uden forhåndsantagelser og uden de krav til variabler som traditionel statistik stiller, at lade data komme til orde, for at kunne finde de mønstre der ligger i data (som allerede nævnt i Benzécri-citatet ovenfor), og derfor bliver korrespondanceanalysen karakteriseret som en deskriptiv og eksplorativ metode, 3) den algebraiske form, som var meget utilgængelig for den angloamerikanske tradition da korrespondanceanalysen blev udviklet og præsenteret, og 4) korrespondanceanalysens unikhed og dens teoretiske slægtskab med mere kendte teknikker som bl.a. principal komponentanalyse og regressionsanalyse (Greenacre 1984: 10–11, efter Rosenlund 2009: 61–62). En passant er det efter min mening værd at fremhæve denne unikhed: At der ikke findes andre teknikker med de samme egenskaber og muligheder med hensyn til visuel styrke, indbygget helhedsfokus, dynamik, fleksibilitet og stabilitet. Alle data, alle informanter og alle variabler er afbildet på ét kort. Det er muligt at fjerne eller tilføje en eller flere variabler i analysen og således udforske variablers betydning en efter en. Analyser med både få og mange variabler og informanter er stabile og at fjerne én eller flere variabler eller informanter påvirker ikke analysens resultater. Brug af såkaldt illustrative variabler, også kaldet supplerende eller supplementære variabler, dvs. inaktive variabler som er med på korrespondanceanalysekortet, men som ikke tæller i analysen, giver mulighed for at udforske særlig interessante variablers betydning i analysen. Et femte kendetegn, som sociologen Lennart Rosenlund tilføjer, er at korrespondanceanalysens "anvendelighed er nærmest ubegrenset" (Rosenlund 1995: 57). Et sjette og sidste kendetegn som jeg vil tilføje, er at korrespondanceanalysen også giver mu-

lighed for hypotesetestning (signifikanstestning) (Greenacre 1984, efter Clausen 1989: 9 og 31, og Le Roux & Rouanet 2010: 12).

I korrespondanceanalysen forholder vi os anderledes til data, analyse og fortolkning end hvad vi er vant til i sociolingvistikken, dels alene på grund af den grafiske fremstilling af analysens resultater, dels fordi analysen i første omgang giver os en yderst komprimeret og detaljeret visuel helhedsfremstilling af data med alle mønstre og relationer og i anden omgang, i en mere raffineret og forbedret analyse, de vigtigste mønstre og relationer på ét korrespondanceanalysekort.

Fordele og udfordringer ved korrespondanceanalysen

Korrespondanceanalysen har som alle andre metoder fordele og ulemper. Blandt fordelene er korrespondanceanalysens sammenfattende og grafiske helhedsfokus, dens dynamik og fleksibilitet og det faktum at den genererer nye problemstillinger, hypoteser og teorier. Disse og andre fordele kan også opfattes som kendetegn ved metoden, og jeg vender tilbage til dem i næste afsnit om hvad korrespondanceanalysen giver sociolingvistikken.

Nogle af korrespondanceanalysens udfordringer hænger nok sammen med uvanthed og manglende erfaring i forhold til at omgås geometrisk statistik og tænke i geometriske baner. Men en speciel udfordring er detaljerigdommen, en anden faren for fejltolkning. Først detaljerne. Så faren.

Netop detaljerigdommen, informationstætheden, kompleksiteten var det første, der slog mig – og fascinerede mig – da jeg første gang så Bourdieus korrespondanceanalysekort i *Distinktionen*. Kortet virkede overvældende og uoverskueligt ved første øjekast på grund af de mange detaljer og oplysninger – og fascinerende. Men kort bliver lettere at læse, når vi tegner forbindelseslinjer mellem indholdsmæssigt forbundne punkter, dvs. mellem variablers kategorier, eller når vi markerer mønstre grafisk, fx med kulørte manglekanter som i kortene i figur 5, 6 og 7 nedenfor, og også når vi i arbejdet med at forfine analysen fx lægger variabler ind gradvis, fjerner variabler, slår variabelkategorier sammen eller inddrager illustrative variabler (se fx Høyen 2010: 52), for det vil ofte være nødvendigt at foretage flere analyser for at nå frem til et tilfredsstillende resultat, og da kan resultatet også være et mindre detaljeret kort.

Faren for fejltolkning er især til stede ved intuitiv læsning af korrespondanceanalysekortet. For selvom et kort til en vis grad er intuitivt tolkbart, kan punkter se ud til at betyde mere eller mindre end de i virke-

ligheden gør. Punkter som ligger nær hinanden i en lavdimensional fremstilling, kan ligge langt fra hinanden i en fremstilling med højere dimensionalitet (Clausen 1989: 29). Derfor kan det være nødvendigt at køre flere analyser for at se hvor punkter ligger i fremstillinger med forskellig dimensionalitet. Punkter i en fremstilling med 1. og 2. dimension vil fx ofte have en anden placering end i en fremstilling med 1. og 3. dimension. For at kunne foretage en præcis vurdering og tolkning af alle punkternes placering og bidrag har vi derfor som nævnt brug for både punkternes bidragsværdier og repræsentationskvalitet i de tilhørende resultattabeller, og vi har også brug for en række grundlæggende tolkningsprincipper som jeg vender tilbage til nedenfor. Andre udfordringer er knyttet til a) hvordan vi tolker karakteristiske mønstre i punktskyerne, b) brugen af regler og metoder i forbindelse med tolkningsarbejdet (jf. fx Hjellbrekke 1999 og Høyen 2010) og c) problemet med såkaldte udeliggere, punkter der ligger i kortets yderområder og som kommer til at dominere for meget i analyseresultatet. Her kan løsningen være at behandle den slags punkter som supplerende, inaktive punkter (Clausen 1989: 29). Problemet med udeliggere findes i min analyse nedenfor, hvor punkterne Voksne børn i Danmark og Anden civilstand er udeliggere, og de kan med fordel i en følgende forbedret analyse gøres inaktive, så de ikke bidrager til akserne, eller evt. helt udelades.

Korrespondanceanalysens fordele og muligheder vejer dog efter min opfattelse til fulde op i forhold de nævnte udfordringer, selvom fordelene nok er mere åbenlyse for os der kender (lidt til) metoden. For andre som endnu ikke har stiftet bekendtskab med denne eller andre geometriske metoder, kan det være nyttigt præcist at vide hvad korrespondanceanalysen kan som metode og giver i en analyse. Det forsøger jeg at give et sociolingvistisk bud på i næste afsnit.

Hvad giver korrespondanceanalysen sociolingvistikken?

Forhåbentlig viser mit eksempel på en korrespondanceanalyse af danskeres tilpasning til norsk nedenfor hvad denne type analyse giver i en sociolingvistisk undersøgelse, og i andre undersøgelser med mange variabler, men en punktvis opsummering med hensyn til korrespondanceanalysens anvendelighed kan give et indtryk af metodens sociolingvistiske værdi. Korrespondanceanalysen:

- håndterer mange variabler og informanter samtidig i én og samme analyse

- giver os statistisk og sammenfattende helhedsbillede af et komplekst materiale og et visuelt overblik over alle mønstre og relationer i et materiale
- er i stand til at udforske samspillet mellem mange eller udvalgte variabler samtidig (og evt. også enkeltvariablers betydning i en helhed, gradvis, én efter én) og sammenhænge mellem mange sproglige og sociale variabler og enkeltvariablers betydning i en helhed i én enkelt analyse, og på den måde er det muligt at undersøge og se hvilke variabler der har størst eller mindst betydning, og få indblik i hvordan forskellige variabler samvirker
- kortlægger samspil, sammenhænge og modsætninger mellem variabler
- afdækker alle underliggende mønstre i et materiale
- er dynamisk og fleksibel og til en vis grad intuitiv tolkbar, til dels fordi et korrespondanceanalysekort bl.a. viser modsætninger, som fx i min analyse mellem de mindst og de mest akkommoderende danskere – og kort har ofte indtegnede linjer og figurer i form af ellipser eller polygoner som viser mønstre og relationer, som i mine tre kort nedenfor, og den slags markeringer letter som nævnt læsningen af kortet og tolkningen af analysen
- er hypotese- og teorigenererende. Korrespondanceanalysens teoretiske potentiale gør at analysen kan afdække ”sammenhænger mellem kategorier og variabler som reiser nye spørgsmål og problemstillinger. Ved hjælp af korrespondanceanalysen kan dataanalysen skape nye hypoteser, isteden for utelukkende å bekrefte eller avkrefte slike” (Rosenlund 1995: 74–75).

I sociolingvistiske undersøgelser er det ikke bare en stor fordel at kunne udforske mange variabler og relationer mellem variabler samtidig, men det er nødvendigt i undersøgelser hvor vi ved at mange variabler samvirker. Og som Bourdieu og mange andre sociologer har vist, har variabler ofte forskellig betydning alt efter hvilke andre variabler de optræder sammen med. Vi kan derfor ikke nøjes med at undersøge samspillet mellem variabler to og to, eller tre og tre, men har brug for at se på samspillet mellem flere variabler samtidig. Her åbner korrespondanceanalysen for nye analyse- og tolkningsmuligheder af sociolingvistiske data i forhold til traditionelle metoders mere begrænsede muligheder.

Korrespondanceanalysens dynamik og fleksibilitet er fordele som sammen med den grafiske præsentation af analyseresultaterne og repræsentationen af data i et socialt rum, gør korrespondanceanalyse særlig aktuel og attraktiv i sociolingvistiske undersøgelser.

Tolkning af korrespondanceanalysekort

Resultatet af korrespondanceanalysen er dels et korrespondanceanalysekort, eller ofte flere kort, dels en rapport med resultattabeller. Et korrespondanceanalysekort har to akser, og x-aksen, den vandrette og vigtigste akse, er første dimension. Første dimension afdækker, eller dekomponerer, den største del af inertien. Her finder vi de stærkeste modsætninger i materialet. Næst vigtigst er y-aksen, den vertikale akse, anden dimension, med de modsætninger der er koblet til den. Tredjevigtigst er z-aksen og tredje dimension, hvis der er tale om et tredimensionalt kort, osv.

Kortlæsning og resultattolkning kræver at vi ved a) hvad et korrespondanceanalysekort viser, b) hvad der er vigtigst at se efter på kortet og desuden c) hvordan vi tolker en akse eller en dimension og hvor mange akser vi tolker (Le Roux og Rouanet 2010: 10) og d) hvilke oplysninger i de tilhørende resultattabeller vi har brug for. Til hjælp i en mere detaljeret tolkning af analysen refererer den norske sociolog Johannes Hjellbrekke til en række regler som den franske statistiker Michel Jambu har opstillet (Hjellbrekke 1999: 44–47).

Med hensyn til tolkning af en akse bringer Brigitte Le Roux og Henri Rouanet et nyttigt citat af Jean-Paul Benzécri som både helt enkelt lærer os hvad vi skal se efter til højre og venstre for origo på et korrespondanceanalysekort, og samtidig i en koncentreret form gør opmærksom på at det gælder om kort og præcist at udtrykke kontrasten mellem de to ekstremer:

Interpreting an axis amounts to finding out what is similar, on the one hand, between all the elements figuring on the right side of the origin and, on the other hand, between all that is written on the left; and expressing with conciseness and precision the contrast (or opposition) between the two extremes. (Benzécri 1992: 405, efter Le Roux og Rouanet 2010: 10).

I resultattabellerne, med bl.a. række- og kolonnetabeller, som er nødvendige for at kunne foretage en præcis tolkning af punkternes placering og bidrag og betydning i analysen, har vi især brug for 1) alle punkters bidrag til de første to akser (og punkternes bidrag til inertien, dvs. punktskyens spredning i rummet) og 2) værdier som fortæller hvor godt punkterne er repræsenteret. Vi har brug for at læse alle punkters bidragsværdier og sammenligne dem indbyrdes og at kontrollere den visuelle tolkning, så vi undgår at fejltolke punkternes placering i rummet. Bidragsværdierne er udgangspunktet for punkternes placering og

angiver punkternes bidrag til de forskellige dimensioner.

Efter at have registreret de vigtigste punkters bidragsværdier og repræsentationskvalitet foretager vi en visuel inspektion af kortet. Vi ser da på fordelingen og spredningen af de to punktsæt på kortet og på ét punktsæt ad gangen, dvs. først alle variabelpunkters placering i forhold til hinanden og de indbyrdes afstande mellem punkterne i det punktsæt, og så alle informantpunkters placering osv. (eller omvendt). Punkter i kortets yderkant er de mest interessante, fordi de bidrager mest til opbygningen af en akse, mens punkter der ligger i og nær origo, x- og y-aksernes krydsningspunkt, repræsenterer gennemsnitsværdier i materialet. Disse punkter har lave bidrag. Målet for punkternes afstand fra origo er det kendte χ^2 . Jo større afstand, jo større afvigelse fra en helt tilfældig fordeling. De beregnede positive eller negative afvigelser som afgør punkters placering, kommer til udtryk grafisk i en repræsentation af *relationerne* mellem punkterne. Vi lægger mærke til store og små *afstande* mellem punkter og punkters *oppositioner og relationer* i forhold til andre punkter. Vi tolker helst klynger af punkter sammen snarere end enkeltpunkter i forhold til hinanden. Det afgørende er ikke punkternes (variabelkategoriernes og individernes) absolutte placering, men deres placering i forhold til hinanden. Punkter der ligger langt fra hinanden, har relativt mindre til fælles end punkter der ligger tættere på hinanden som har relativt mere til fælles. Punktpar der ligger i hver sin ende af en akse, til højre og venstre for origo, er som nævnt hinandens modsætninger.

Både selve analysen og tolkningen af kortene er iterativ, det vil sige gentagelses- og venden-tilbagepræget. Denne iterative analyse- og tolkningsproces har at gøre med teoretiske antagelser og forventninger knyttet til valget af variabler, og det er som nævnt ofte nødvendigt at køre flere analyser om for at nå frem til en tilfredsstillende analyse.

Data og variabler i korrespondanceanalysen af danskeres sproglige tilpasning til norsk

I min korrespondanceanalyse er der 49 danske informanter og 15 variabler (12 sociale og 3 sproglige variabler) med i alt 51 kategorier. Informanterne er et statistisk tilfældigt udvalg af danske statsborgere over 18 år som er registreret i *Det Sentrale Personregister* i Norge. Af informanterne bor 20 i Bergen, 14 i Kristiansand og 15 i Oslo. 27 er kvinder, og 22 mænd. Den yngste er 26 år, den ældste 67. Deres alder ved flytning til Norge varierer fra 17 til 50 år, og antal år i Norge fra 2 til 40 år. Med hensyn til civilstatus er 16

dansk gift, 27 norsk gift, 3 er enlige, og 3 har en partner som hverken er dansk eller norsk. Disse og flere andre oplysninger kommer fra registreringsskemaer som informanterne har udfyldt.

De 12 baggrundsvariabler og deres 40 kategorier er disse:

- *Køn* (med 2 kategorier): Kvinde; Mand
- *Informantby* (med 3 kategorier): Bergen; Kristiansand; Oslo
- *Alder* (med 3 kategorier): 26–42 år; 46–57 år; 59–67 år (disse alderskategorier ligger i datamaterialet på den måde at informanternes alder fordeler sig i de tre kategorier, dvs. ingen informanter er under 26 år, ingen er 43, 44 og 45 og ingen 58)
- *Alder ved flytning til Norge* (med 4 kategorier): Til N i en alder af < 22 år; Til N i en alder af 23–27 år; Til N i en alder af 30–38 år; Til N i en alder af > 40 år (disse kategorier ligger i datamaterialet på den måde at informanternes alder ved flytning fordeler sig i de nævnte grupper, og ingen informanter flyttede til Norge i en alder af 28 og 29 år og 39 år)
- *Antal år i Norge* (med 3 kategorier): < 13 år i N; 14–20 år i N; > 21 år i N
- *Civilstand* (med 4 kategorier): Norsk gift; Dansk gift; Enlig; Anden civilstand
- *Uddannelsesland* (med 3 kategorier): Uddannet i Danmark (Udd. i DK); Uddannet i Norge (Udd. i N); Uddannet i Danmark og Norge (Udd. i DK & N)
- *Børn* (med 5 kategorier): Børn i norsk skole (hvis et af flere børn i en familie er 5 år eller ældre kategoriseres informanten som havende børn i norsk skole); Ingen børn; Voksne børn i Norge; Voksne børn i Danmark; Små børn
- *Motiv for at flytte til Norge* (med 4 kategorier): Flyttemotiv arbejde (eget eller partners arbejde); Flyttemotiv familie; Flyttemotiv andet (tilfældigheder, nye udfordringer, udfærdstrang, kompromis Færøerne og DK); Flyttemotiv ej oplyst
- *Uddannelse* (med 2 kategorier): Boglig uddannelse (dvs. højere uddannelse efter gymnasium, men ikke tekniske og håndværksmæssige uddannelser); Ikke-boglig uddannelse (dvs. praktiske og tekniske uddannelser)
- *Erhvervmæssig brug af skriftsprog* (med 3 kategorier): Skriver meget på arbejde; Skriver lidt på arbejde (skemaer, rapporter ol.); Skriver ikke (eller kun af og til)
- *National identitet* (med 4 kategorier): Dansk id.; Norsk id.; Dansk & norsk id. (inkl. ”mest dansk”

og ”mest norsk”); Anden id. (hverken dansk eller norsk; skandinav; europæer; ved ikke; ikke opgivet).

De 3 sproglige variabler og deres 11 kategorier er:

- *Leksikalsk tilpasning* (med 3 kategorier): Ringe leksikalsk tilpasning (= 0,9–1,6 % norske ord); Middel leksikalsk tilpasning (= 2–4,9 % norske ord); Fuld leksikalsk tilpasning (= norsk ordforråd)
- *Udtaletilpasning* (med 5 kategorier): Ingen udtaletilpasning; Ringe udtaletilpasning (= under 21 % mørke norske *a*-lyde); Moderat udtaletilpasning (= 31–40 %); Stærk udtaletilpasning (= 62–67 %); Fuld udtaletilpasning (= 100 %)
- *Talesprog* (med 3 kategorier): Talesprog DK = Dansk (svarer til Ringe leksikalsk tilpasning og Ingen udtaletilpasning); Talesprog N = Norsk (svarer til stærk leksikalsk og udtalemæssig tilpasning); Talesprog Bl. = Blandingssprog (rummer Middel leksikalsk tilpasning og udtalemæssig tilpasning varierende fra ringe til stærk).

Den leksikalske variabel er andelen af norske ord i det telefoninterview hvor de danske informanter taler med den norske interviewer som omtalt i indledningen. Udtalevariablen drejer sig som nævnt om de danske informanternes tilpasning af deres lyse, danske *a*-lyde til mørke, norske *a*-lyde. Variablen Talesprog falder indholdsmæssig sammen med de to tilpasningsvariabler og fungerer på den måde som kontrolvariabel. De to tilpasningsvariabler stammer fra to analyser af transskriberede telefoninterview med de 49 informanter. De to variabelers kategorier er procentangivelser fra en leksikalsk frekvensanalyse og en variationsanalyse.

Frekvensanalysen er en analyse af de 49 informanternes *samlede leksikon* i telefoninterviewet (i alt 191.100 ord). Variationsanalysen er en analyse af 41 informanternes *udtale af a-lyde*, dvs. kun ét udtaletræk i et uddrag af interviewet. Der er 41 variationsanalyser, og ikke 49, fordi 8 interview er transskriberet på norsk eftersom de 8 informanter taler norsk, dvs. de har 100 % tilpasning og bruger kun norske *a*-lyde. Der er tale om ét udtaletræk, nemlig vokalen *a*, som er et meget hyppigt tilpasningstræk blandt alle informanter der tilpasser deres udtale til norsk, og det rammer et stort antal almindelige ord. Derfor er det egnet til at repræsentere udtalemæssig tilpasning. (Endnu en grund til at fokusere på ét enkelt træk er at en fonetisk analyse er meget tidkrævende).

materiales omfang. I statistisk sammenhæng er 49 danske informanter trods alt ikke alverden, når vi tager i betragtning at der er i alt 21 354 danske statsborgere i Norge (per 1. januar 2012) og 87 % af dem er 18 år og ældre (e-post 2. okt. 2012 fra seniorrådgiver Minja Tea Dzamarija, Statistisk Sentralbyrå). Men som nævnt er korrespondanceanalysen en stabil teknik selv i undersøgelser med få informanter.

Jeg kommenterer ikke andre mønstre end de tre nævnte og heller ikke informantpunkterne enkeltvis, men disse punkter er knyttet til de viste mønstre på den måde at de indrammer hvert mønster, og fremgangsmåden for tolkning af dem er den samme som for variabelpunkterne. Jeg gør ikke nærmere rede for resultattabellernes indhold og tager ikke tabellerne med, men de er tilgængelige via min hjemmeside: <https://uni.hi.is/rbb/>. De indeholder i min analyse oplysninger for bidrag og værdier knyttet til 10 akser.

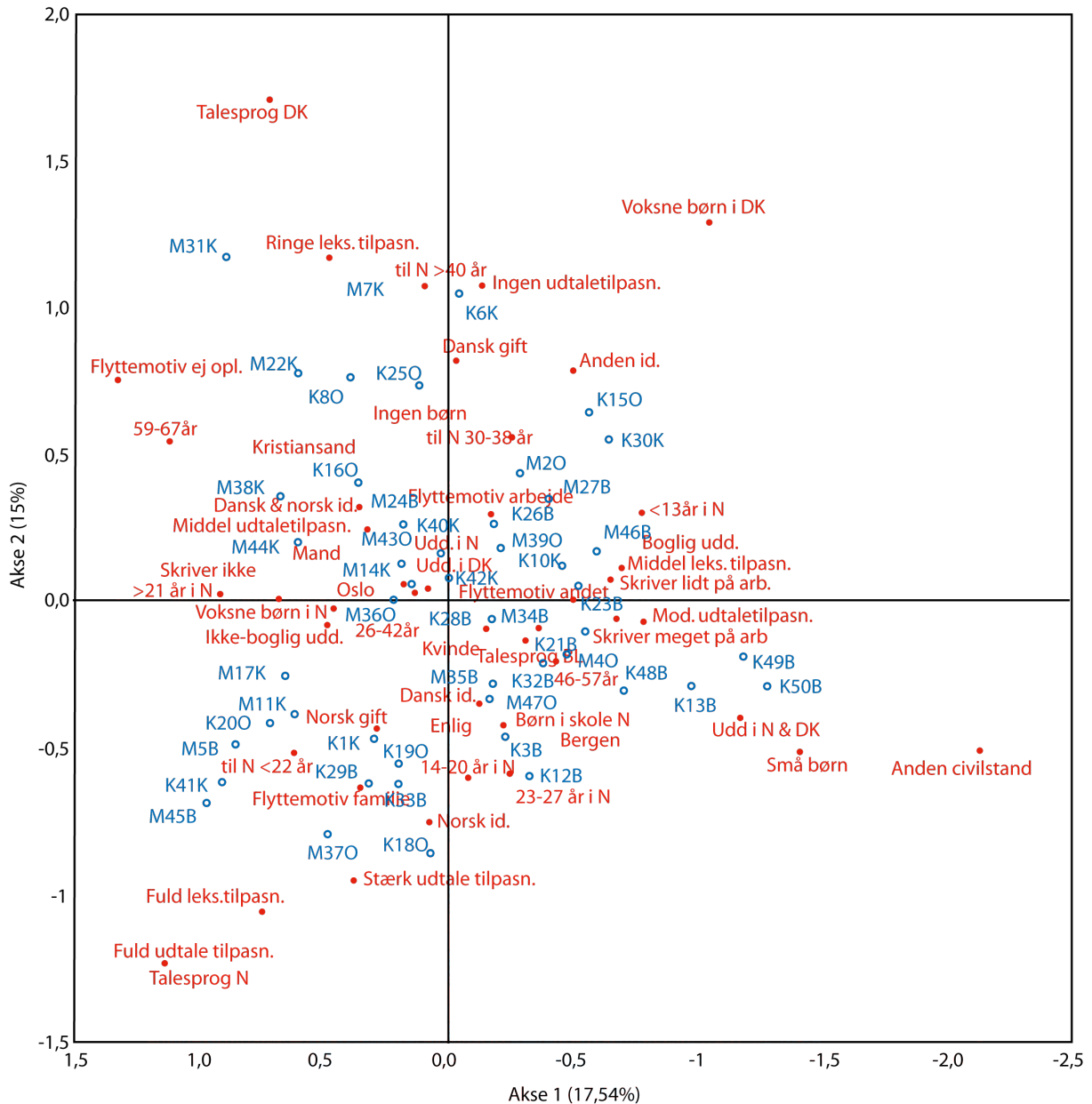
De fire kort i figur 4, 5, 6 og 7 nedenfor indeholder alle de samme oplysninger, dvs. alle informanter og variabelkategorier. Kortet i figur 4 er et baggrundskort med alle informanter og variabler, men uden indtegnede mønstre. Det er det samme kort som i de tre figurer nedenfor, men uden indtegnede mønstre. De tre kort er også alle de samme, dvs. de rummer de samme variabel- og informantpunkter og de samme mønstre. Forskellen mellem kortene er kun, at der på hvert kort er indtegnet et eget mønster som analysen får frem, dvs. det leksikalske tilpasningsmønster, antal-år-i-Norge-mønsteret og bostedsmønsteret. De tre mønstre kunne i princippet være tegnet ind på ét kort, men det ville give et kort med overordentlig mange detaljer, og det ville derfor være vanskeligt at beskrive og læse.

På kortene ser vi et todimensionalt plan, som består af et koordinatsystem med en x- og en y-akse som deler kortet i fire felter og vandret og lodret i to halvdele. På kortene er akse 1 og 2 med første og anden dimension vist, og alle 49 informanter og alle 51 variabelkategorier, dvs. informanternes egenskaber, er repræsenteret som punkter. De værdier der er knyttet

til hver akse og de i alt 10 dimensioner som er analyseret, står i de nævnte resultattabeller. På kortene ser vi ved begge akser i parentes den procentdel som akserne afdækker af inertien eller variansen i materialet. Den første akse, x-aksen, den vandrette akse ("CA axis 1"), afdækker 17,54 %. Den anden akse, y-aksen, den lodrette akse ("CA axis 2"), 15 %. Til sammen afdækker de to akser 32,54 %.

X-aksen får sine vigtigste bidrag fra: a) *År-i-Norge*-punkterne *Op til 13 år i N* og *Over 20 år i N*, b) *Boglig uddannelse* og *Ikke-boglig uddannelse*, *Middel leksikalsk tilpasning*, *Skriver ikke*, *Talesprog Blanding*, *Ringe udtaletilpasning* og alderskategorien *59–67 år*. Denne dimension rummer to interessante modsætninger, nemlig a) mellem kort opholdstid (*Op til 13 år i N*) og lang (*Over 20 år i N*) opholdstid og b) mellem boglig og ikke-boglig uddannelse. Kort og lang opholdstid, som er den vigtigste modsætning, vender jeg tilbage til nedenfor ved figur 6. Med hensyn til modsætningsparret *Boglig uddannelse* og *Ikke-boglig uddannelse* ser der ud til at være en sammenhæng mellem på den ene side *Boglig uddannelse* og *Middel leksikalsk tilpasning*, *Ringe udtaletilpasning* og *Op til 13 år i N* og på den anden side *Ikke-boglig uddannelse*, *Skriver ikke* og *Over 20 år i N*. Det tyder på at uddannelse er en kraft der giver energi til moderat tilpasning, at boglig uddannelse og sproglig tilpasning ofte hænger sammen, og det samme gælder ofte ikke-boglig uddannelse, ingen/begrænset brug af (dansk) skriftsprog og lang opholdstid.

Y-aksens vigtigste bidrag kommer fra a) *Fuld leksikalsk tilpasning* og *Ringe leksikalsk tilpasning*, b) *Ingen udtaletilpasning* og *Stærk udtaletilpasning*, *Til Norge i en alder af over 40 år*, c) *Talesprog dansk* og *Talesprog Norsk*, d) *Dansk gift* og *Norsk gift*. Her har vi en dobbelt modsætning: mellem a) høj og lav grad af leksikalsk tilpasning og b) ditto udtalemæssig tilpasning. Det betyder at de to tilpasningsformer går hånd i hånd. Den fjerde modsætning er mellem c) *Dansk gift* og *Norsk gift*, som man kunne forvente.

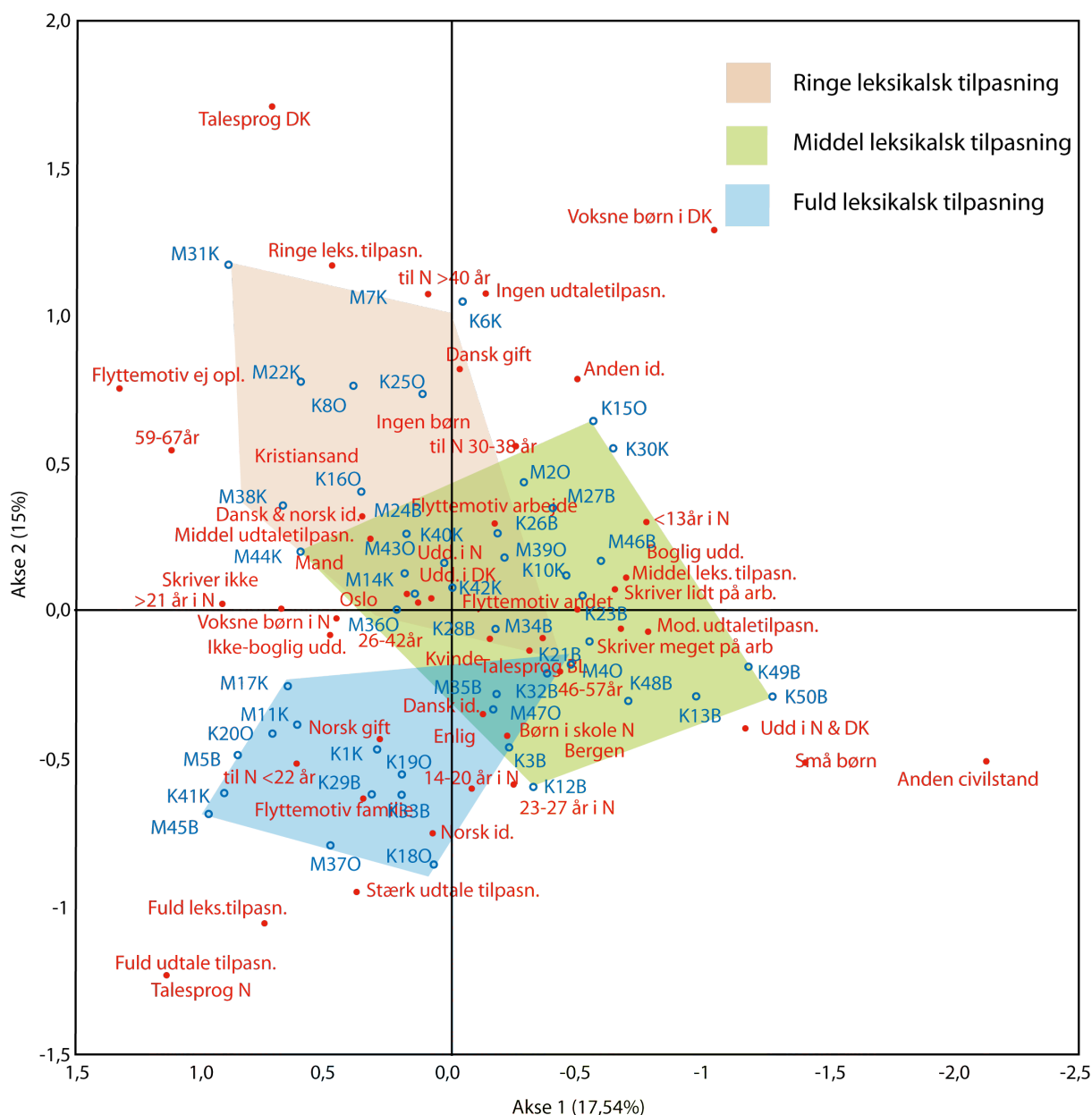


Figur 4. Korrespondanceanalyse af sproglig tilpasning med 49 danske informanter i Norge og 51 variabelkategorier. Akse 1 og 2

På kortet i figur 5 nedenfor ser vi det leksikalske tilpasningsmønster, dvs. omfanget af danskeres leksikalske tilpasning til norsk. Kortets tre indtegnede mangekanter viser: 1) de tre førnævnte gruppers tilpasning med gruppernes størrelse, spredning og placering i forhold til hinanden, 2) modsætningen mellem gruppen med mindst og gruppen med mest tilpasning, markeret ved henholdsvis den røde og den blå mangekant som ligger langs med y-aksen (de to har modsat fortegn på y-aksen), og 3) hvordan informanterne fordeler sig i de tre grupper.

Mangekanternes betydning på kortet i figur 5 (nedenfor):

- Rød mangelkant = De 11 informanter med ringe leksikalsk tilpasning, dvs. (0,9–1,6 %)
- Grøn mangelkant = De 24 informanter med middel leksikalsk tilpasning (2–4,9 %)
- Blå mangelkant = De 14 informanter med fuld leksikalsk tilpasning (norsk ordforråd).



Figur 5. Korrespondanceanalyse af 49 danskeres leksikalske tilpasning til norsk. Akse 1 og 2

I figur 5 (ovenfor) har vi *den røde gruppe* med informanter med mindst tilpasning i nord og *den blå gruppe* med informanter med mest tilpasning i syd, i hver sin halvdel af kortet, som hinandens modsætninger, og i midten *den grønne gruppe* med informanter med middel og mest ”gennemsnitlig” tilpasning. I den røde mangekant ligger også punktet *Dansk gift*, mod nord, næsten på y-aksen, og i nærheden også punktet *59-67 år*, i vest. Begge trækker i retning af mindst tilpasning, i dansk retning. I den blå mangekant har vi også punktet *Norsk gift*, i syd. Det betyder at civilstand er en kraft i tilpasningen. Den faktor trækker i to modsatte retninger: De dansk gifte informanter tilpasser

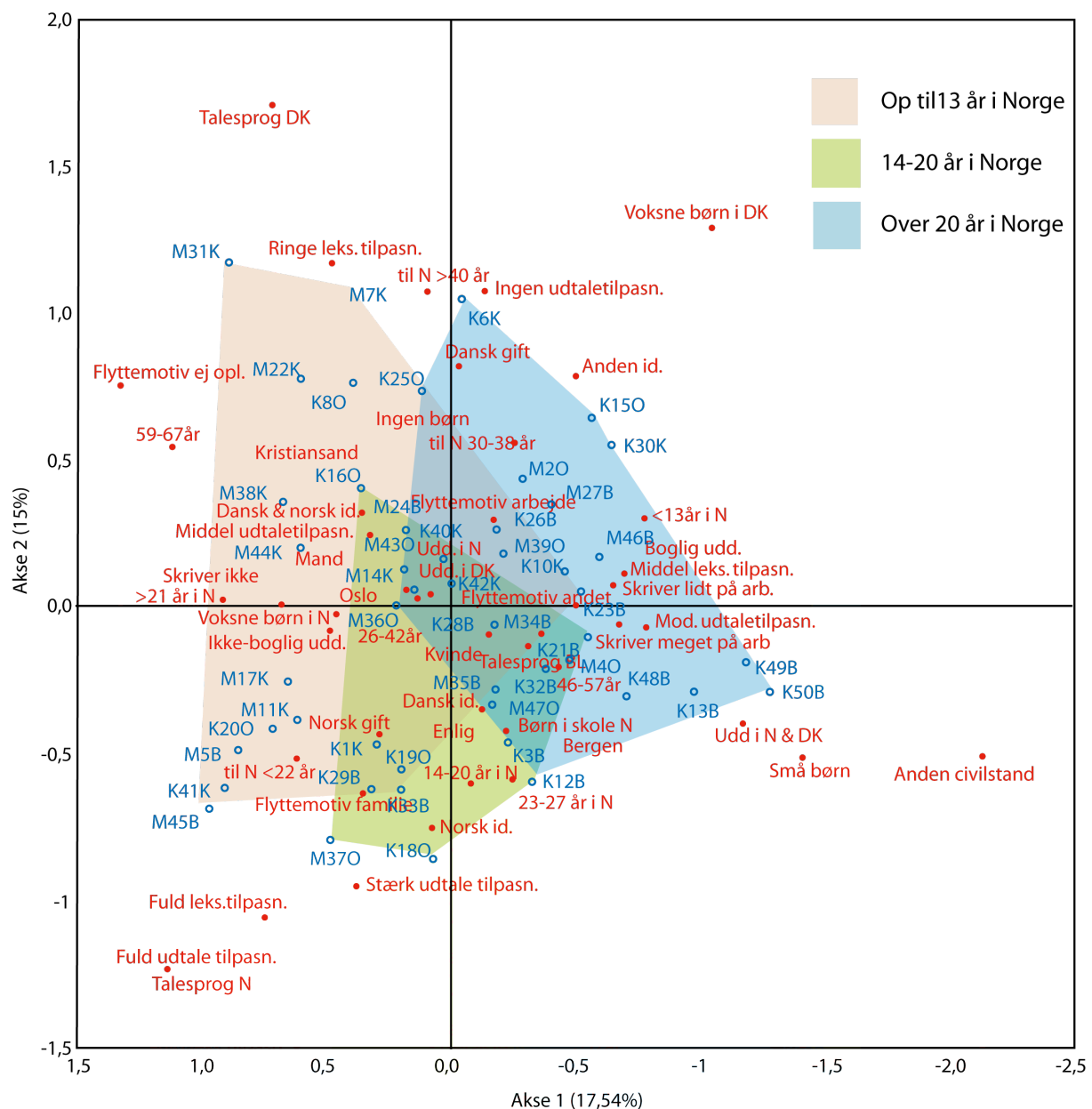
sig relativt mindst, og de norsk gifte informanter relativt mest, som forventet, men at være dansk gift betyder mere end at være norsk gift. Desuden ser vi at nær punktet *Ringe leksikalsk tilpasning* i nord ligger punktet *Ingen udtaletilpasning*, og nær punktet *Fuld udtaletilpasning* i sydvest ligger både *Stærk udtaletilpasning* og *Fuld udtaletilpasning*. Det betyder at leksikalsk og udtalemæssig tilpasning til dels går hånd i hånd (udtaletilpasningsmønsteret er ikke indtegnet). I den grønne mangelkant hvor vi har informanter med mest "gennemsnitlig" tilpasning, har vi også punktet *Talesprog Blandingssprog* som kan siges at sammenfatte den typisk "gennemsnitlige" tilpasning. I den

røde mangelkantar øverst er der en samanheng mellom ringe sproglig tilpassing, det at vere dansk gift og at komme til Norge i velvoksen alder, dvs. over 40 år.

De punkter der præger anden dimension mest er *leksikalsk tilpasning* sammen med *Ingen udtaletilpasning*. De to tilpasningsformer hører med andre ord sammen som hånd i handske.

På kortet i figur 6 (nedenfor) ser vi opholdstidsmønstret, dvs. hvad antal år i Norge betyder. De tre manglekanter markerer: 1) de tre Antal-år-i-Norge-

grupper og deres placering i forhold til hinanden, 2) modsætningen mellem kort og lang opholdstid i Norge (*Op til 13 år i Norge* og *14–20 år i Norge*), 3) hvad opholdstid betyder i sammenhæng med sproglig tilpasning og visse andre faktorer og 4) udviklingen med hensyn til antal år i Norge, fra kortest opholdstid (op til 13 år) til mellemlang opholdstid (14–20 år) og videre til længst opholdstid (over 20 år). (De to ekstreme punkter *Voksne Børn i Danmark* og *Anden civilstand* er som nævnt tidligere såkaldte udeliggere og er trods deres placering ikke vigtige).



Figur 6. Korrespondanceanalyse af 49 danskeres sproglige tilpasning til norsk i forhold til antal år i Norge. Akse 1 og 2

Mangekanternes betydning på kortet i figur 6 (ovenfor):

- Blå mangekant = De 20 informanter med kortest opholdstid i Norge = under 13 år
- Gul mangekant = De 11 informanter med mellem-lang opholdstid i Norge = 14–20 år
- Rød mangekant = De 19 informanter med længst opholdstid i Norge = over 21 år.

I figur 6 (ovenfor) har vi *den blå gruppe* i øst og *den røde gruppe* med mest i vest, i hver sin halvdel af kortet, som hinandens modsætninger, og i mellem de to *den grønne gruppe* mest sydlig. Ved de tre grupper tilpasningsspredning er det interessant at den røde gruppe med længst opholdstid har størst spredning, lige fra de mest dansksprogede informanter i den danske ende i nord til de mest norske informanter i syd, mens den gule gruppe i midten strækker sig længst mod syd og den blå gruppe mod øst. Alle tre grupper har med andre ord stor tilpasningsvariabilitet. Videre er punktet *Under 13 år i Norge* interessant pga. nærheden til punktet *Middel leksikalsk tilpasning* og *Moderat udtaletilpasning*. Det kan betyde, at både leksikalsk og udtalemæssig tilpasning allerede sker inden for en opholdstid på *under 13 år*, at tid er den kraft der giver tilpasningsenergi til et vist punkt. Punktet *14–20 år i Norge* er interessant fordi det *ikke*

ligger nær højtilpasningspunkterne i sydvest, men nær punktet *Skriver ikke*. Det kan for det første tyde på, at lang opholdstid i Norge (over 21 år) kan være koblet til det ikke at have et skriftsprog eller begrænset brug af skriftsprog, hvad der måske kan opfattes sådan at en del af ”prisen” for lang opholdstid er problemer knyttet til at skrive (fx findes der i interviewmaterialet og de sprogbiografiske beretninger oplysninger fra informanter som fortæller, at de har problemer med at skrive og/eller at de har mistet deres danske skriftsprog). Det kan for det andet betyde, at lang opholdstid *ikke* er en kraft der giver tilpasningsenergi og trækker i retning af stærkere tilpasning med årene. Tilpasningen stopper med andre ord op eller når en grænse på et tidspunkt. Ellers ville lang opholdstid være koblet til stærkere tilpasning. Opholdstidsvariablen *Antal år i Norge* er den variabel, der præger x-aksen mest og giver kulør med modsætningen mellem kortest og længst opholdstid (mellem *Under 13 år* og *over 21 år i Norge*). Tid er den kraft der strækker første dimension.

På kortet i figur 7 (nedenfor) ser vi bostedsmønstret, dvs. hvilken betydning bosted har. De tre mangekanter fortæller os: 1) de tre bosted-grupper og deres placering i forhold til hinanden, 2) modsætningen mellem Kristiansand-gruppen og de to andre grupper, 3) de tre grupperes spredning i forhold til hinanden.

Mangekanternes betydning på kortet i figur 7 ovenfor:

- Blå mangelant = De 15 informanter i Kristiansand øverst på kortet
- Gul mangelant = De 14 informanter i Oslo mest i midten af kortet
- Rød mangelant = De 20 informanter i Bergen nederst på kortet.

21

manter i Kristiansand befinder sig i den sprogligt mest danske (eller mindst norske) ende af kortet. Det kan betyde, at den korte afstand til Danmark og den sproglige nærhed mellem kristiansandsk og dansk spiller en vis rolle og at det har betydning, at kristiansandere er mere vant til og derfor bedre til at forstå dansk end bergensere og osloseanere. Oslo-gruppen ligger i midten og er ”gennemsnitgruppen”.

Sammenfattende viser korrespondanceanalysen de mest og mindst betydningsfulde faktorer i tilpasnings-sammenhæng. Vi får med andre ord en slags ”opskrift” på, hvad der kan give mest og mindst tilpasning, og enkelte overraskelser:

- En dansker som er norsk gift, bor i Oslo, har forholdsvis lang opholdstid i Norge (op til 13 år) og en boglig uddannelse, kan have en moderat leksikalsk og udtalemæssig tilpasning.
- Nogle danskere i Kristiansand har typisk kun ringe leksikalsk tilpasning, mens andre kan også have enten moderat eller stærk tilpasning, både leksikalsk og udtalemæssig.
- En dansker, som er dansk gift og som er kommet til Norge i velvoksen alder (40+), eller som er midaldrende (59+) eller måske har boet i Norge i mange år (over 20 år), vil være tilbøjelig til at tale dansk og kun tilpasse sig lidt leksikalsk.
- Uden (nævneværdig) betydning i tilpasningssammenhæng er bl.a. køn og identitet.

Som afslutning opsummerer jeg nu punktvis de omtalte og de viste resultater i form af svar på det tredelte spørgsmål jeg stillede i indledningen, og supplerer med forklarende kommentarer som bygger på oplysninger informanterne har givet i det tidligere nævnte telefoninterview.

Hvordan tilpasser danskere i Norge sig til norsk?

Korrespondanceanalysen viser disse overordnede resultater:

- Mange danskere i Norge tilpasser sig leksikalsk og udtalemæssigt og kombinerer ofte de to former for tilpasning, dvs. danskere som tilpasser sig meget leksikalsk, tilpasser også deres udtale mere til norsk, men ingen danskere har kun udtalemæssig tilpasning. De to forskellige tilpasningsformer skal derfor ses i sammenhæng.
- Mange danskere i Oslo tilpasser sig leksikalsk mere ”gennemsnitligt” i forhold til mange danskere i Bergen og Kristiansand. Måske bl.a. pga. større tilpasningspres i Oslo. Måske fordi Oslo-norsk

har stor prestige som hovedstadssprog og er nær bokmål som igen er nær dansk. Måske er folk i Oslo mindre vant til eller mindre gode til at forstå dansk end folk i Bergen og Kristiansand er.

Det skal tilføjes, at sproglig tilpasning er én af de danske informanternes tilpasningsstrategier, og det er en strategi som alle informanter har et bevidst og reflekteret forhold til. Den forekommer ofte i kombination med andre tilpasningsstrategier, fx langsommere og tydeligere tale og brug af gentagelser og omformuleringer.

Hvor meget?

- Danskeres leksikalske tilpasning til norsk er præget af stor spredning og variabilitet, især i Kristiansand. Her forekommer ingen, moderat og stærk tilpasning. *Ingen* tilpasning måske fordi kristiansandsk ligner dansk mere end andre norske talesprog og fordi kristiansandere er vant til at høre dansk og særlig gode til at forstå dansk. ”Her forstår alle dansk!”, siger en informant fra Kristiansand. *Stærk* tilpasning måske fordi kristiansandsk er lettere at lære på grund af nærheden til dansk. Kristiansandere er måske også meget tolerante i forhold til danskeres sproglige tilpasning fordi de selv er ”reservedanskere”, som de kalder sig.
- Mange danskeres leksikalske og udtalemæssige tilpasning er meget variabel af omfang: Nogle har næsten ingen tilpasning, nogle har bare leksikalsk tilpasning, nogle tilpasser sig både leksikalsk og udtalemæssig, nogle har ringe leksikalsk og udtalemæssig tilpasning, andre har omfattende tilpasning på begge områder.
- Mange tilpasser sig moderat leksikalsk, en del tilpasser sig i ringe grad leksikalsk og en del er leksikalsk stærkt norskpræget. Moderat tilpasning er for mange tilstrækkeligt (for at blive forstået) og også en naturlig del af det at bo i Norge og være udsat for daglig norskpåvirkning. Ringe tilpasning kan hænge sammen med at der er mange leksikalske ligheder mellem dansk og norsk (eller få forskelle hvis det er forskelle vi fokuserer på), og derfor er behovet for tilpasning ikke særlig stort. Stærk og fuld tilpasning er for nogle udtryk for et bevidst valg, et ønske om at ville tale norsk, at høre hjemme i Norge.

Hvorfor (– dvs. hvilke faktorer påvirker og forklarer tilpasningen)?

- Tilpasningsfremmende faktorer er især at være norsk gift, at bo i Oslo, at have kort opholdstid i

Norge og boglig uddannelse. Modsat er tilpasningshæmmende faktorer at være dansk gift, at flytte til Norge efter at have rundet de 40 og at være over 60 år.

- Bosted er en vigtig faktor: Der er en klar tendens til at danskere i Kristiansand tilpasser sig både forholdsvis mindre og mere end danskere i Bergen og Oslo. Danskere i Oslo tilpasser sig ofte mere moderat og mere gennemsnitligt end danskere i de to andre byer.
- Køn er ikke – som i mange andre sociolingvistiske undersøgelser – en faktor, der spiller en vigtig rolle. Kvinder ser ikke ud til at tilpasse sig generelt hverken mere eller mindre end mænd, og mænd ikke mindre eller mere end kvinder. Der findes med andre ord ikke et udpræget kønsmønster (vi ser på kortene at kønsvariablens kategori Kvinde ligger forholdsvis nær origo, lidt under og nær x-aksen, og kategorien Mand ligger lidt længere væk fra origo, lidt over x-aksen).
- Som supplement til de oplysninger korrespondanceanalysen giver, vil jeg tilføje for det første mange informanternes rapporterede forklaringer på, hvorfor de tilpasser sig til norsk (fra det indledningsvis omtalte telefoninterview): a) forståelseshensyn, b) gradvis og langvarig norskpåvirkning, c) vane og afsmitning, d) et ønske om at høre til i Norge og e) resultatet af et bevidst valg, og for det andet en del informantforklaringer på ingen eller begrænset tilpasning: a) unødvendig, b) norsk og dansk er dialekter og c) brug af andre tilpasningsstrategier end sproglige.

Opsummering og konklusion

Med præsentationen af min undersøgelse af danskeres sproglige tilpasning til norsk i Norge har jeg vist et sociolingvistisk eksempel på brug af multivariat korrespondanceanalyse. Jeg har vist at denne geometriske metode er særlig velegnet til sociolingvistisk analyse, fordi den giver et sammenfattende og visuelt helhedsbillede og viser hvilke kræfter der har størst og mindst og ringe betydning for omfanget af danskeres sproglig tilpasning til norsk: Vigtige tilpasningsfaktorer er især kort og lang opholdstid, dvs. < 13 år og > 21 år i Norge, og civilstand, dvs. at være dansk og norsk gift.

Korrespondanceanalyse afdækker i min undersøgelse tre mønstre: 1) et leksikalsk tilpasningsmønster (med op til 5 % leksikalsk tilpasning) med civilstatus som en stærkere tilpasningshæmmende kraft for en dansk gift dansker end en tilpasningsfremmende kraft for en norsk gift dansker, 2) et bostedsmønster med bl.a. både mest og mindst tilpasning i Kristiansand og

mest gennemsnitlig tilpasning i Oslo og 3) et opholdstidsmønster, hvor lang opholdstid kan være koblet til ikke at have et skriftsprog, så måske er det prisen nogle danskere betaler for lang opholdstid, og tid er en kraft der giver en vis energi til tilpasning, for tilpasning sker allerede efter kort opholdstid i Norge, mens lang opholdstid ikke er en kraft der trækker i retning af stærkere tilpasning med årene.

Som konklusion lyder svarene på de indledende spørgsmål: Ja, der *er* forskel på hvordan danskere tilpasser sig i Bergen, Kristiansand og Oslo: Der er størst tilpasningsvariabilitet i Kristiansand, og typisk tilpasning i Oslo, hvad der kan skyldes et vist tilpasningspres i retning af norsk. Der *er* tilpasningsforskelle knyttet til antal år i Norge, civilstand og uddannelse: Såvel kort og lang opholdstid i Norge som det at være dansk og norsk gift er vigtige faktorer. Også uddannelse har en vis kraft og tilpasningsbetydning på den måde at der er en tendens til at moderat tilpasning og boglig uddannelse hører sammen, og det samme gælder ikke-boglig uddannelse, lang opholdstid i Norge og ingen eller begrænset brug af (dansk) skriftsprog. Alder og alder ved flytning til Norge har mindre betydning end vi kunne forvente. Køn og national identitet spiller ingen nævneværdig rolle i min undersøgelse.

□

Randi Benedikte Brodersen

Lektor i dansk sprog ved Danskafdelingen, Islands Universitet (Háskóli Íslands) (og på orlov fra en stilling ved Institutt for filosofi og førstesemesterstudier, Universitetet i Bergen). E-post: rbb@hi.is og randi.brodersen@fof.uib.no

Abstract

In my presentation of my study of the linguistic accommodation to Norwegian among Danes in Norway I have shown a sociolinguistic example of the use of multiple correspondence analysis. This geometric method is very suitable for sociolinguistic analyses, because it gives a visual and holistic picture. It is capable of handling all informants and variables at the same time, in one unified graphic. It shows some strong and weak forces in the accommodation among the Danes: Important factors are marital status (being married to a Dane or a Norwegian) and the number of years in Norway (< 13 years and > 21 years in Norway).

My correspondence analysis shows three patterns: 1) A lexical accommodation pattern with marital status as a stronger force for a Dane married to another Dane than for a Norwegian married to a Dane, 2) residence pattern with the strongest accommodation in Kristiansand and the strongest average accommodation in Oslo, and 3) a number of years in Norway pattern where time is a force giving energy to accommodation already within a small number of years even though Danes do not accommodate more as they get older.

My conclusion is that accommodation among Danes in Norway does differ in the three cities investigated. The accommodation variability is biggest in Kristiansand and most typical and average in Oslo. There are accommodation differences related to short as well as long time of residence in Norway and to being married to a another Dane or a Norwegian. Education as well is important. Age and migration age seem to be less important than one might expect. Gender and national identity do not seem to play a role in my study. However, it is important to stress that my correspondence analysis shows the complexity of the factors influencing the accommodation and the interrelation of several factors. My correspondence analysis also gives rise to new questions and theories which call upon a new correspondence analysis.

Keywords

Linguistic accommodation among Danes in Norway, sociolinguistics, multiple correspondence analysis (MCA), geometrical data analysis (GDA), holistic statistics.

Referencer

- Benzécri, Jean-Paul 1973. *L'analyse des données* I-II. Paris: Dunod.
- Benzécri, Jean-Paul 1992. *Correspondence Analysis Handbook*. New York/Basel/Hong Kong: Marcel Bakker.
- Bourdieu, Pierre 1984. *Distinction. A Social Critique of the Judgement of Taste*. London: Routledge & Kegan Paul.
- Bourdieu, Pierre 1988. *Homo Academicus*. Cambridge: Polity Press.
- Bourdieu, Pierre 1995. Distinktionen. *En sociologisk kritik af dommekraften*. Oversættelse til norsk ved Annick Prieur, oversættelse af efterord ved Theo Barth og indledning ved Dag Østerberg. Helsingør: Det lille forlag.
- Brinkkjær, Ulf 2007. Regression og geometrisk data analyse (1. del). *Praktiske Grunde* 3-4/2007: 5–23.
- Brinkkjær, Ulf 2010. Regression og geometrisk data analyse (2. del). *Praktiske Grunde* 2/2010: 27–35.
- Brodersen, Randi Benedikte 1998. Leksikalsk akkommodation blandt 11 danskere i Norge. In *Ord etter ord. Heidersskrift til Oddvar Nes*, red. Akselberg, Gunnstein & Jarle Bondevik, 57–78. Bergen: Norsk bokreidingslag.
- Brodersen, Randi Benedikte 2009. Korrespondanceanalyse – et dynamisk og fleksibelt sociolingvistisk værktøj. In *12. Møde om Udforskningen af Dansk Sprog 9.-10. oktober 2008*, red. Hansen, Inger Schoonderbeek & Peter Widell, 115–130. Århus: Aarhus Universitet.
- Brodersen, Randi Benedikte 2010. Hvordan, hvor meget og hvorfor tilpasser danskere i Norge sig sprogligt til norsk? Hovedresultater fra en sproglig undersøgelse vist ved brug af korrespondanceanalyse. In *Millimåla. Ársrit Stofnunar Vigdísar Finnbogadóttur í erlendum tungumálum*. red. Þráinsdóttir, Rebekka & Erla Erlendsdóttir, 247–280. Reykjavík: Vigdís Finnbogadóttir for Fremmedsprog, Islands Universitet.
- Clausen, Sten-Erik 1989. *Korrespondanseanalyse. Grafisk representasjon av frekvenstabeller*. Oslo: Norsk institutt for by- og regionsforskning.
- Clausen, Sten-Erik 1998. *Applied Correspondence Analysis. An Introduction*. Thousand Oaks, CA: Sage.
- Greenacre, Michael J. 1984. *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Greenacre, Michael J. 1993. *Correspondence Analysis in Practice*. London: Academic Press.
- Greenacre, Michael og Jörg Blasius (red.) 1998. *Visualization of Categorical Data*. New York: Academic Press.
- Gronmo, Sigmund 2004. *Samfunnsvitenskapelige metoder*. Bergen: Fagbokforlaget.
- Halvorsen, Knut 2008. *Å forske på samfunnet. En innføring i samfunnsvitenskapelig metode*. Oslo: Cappelen (5. udgave).
- Hjellbrekke, Johannes 1999. *Innføring i korrespondanseanalyse*. Bergen: Fagbokforlaget.
- Høyen, Marianne 2004. *Korrespondanceanalyse*. København: Akademisk Forlag.
- Høyen, Marianne 2010. *Introduktion til korrespondanceanalyse*. Odense: Syddansk Universitetsforlag.
- Karverket 2009. *Norgeskart*. <http://kart.statkart.no/adaptive2/default.aspx?gui=1&lang=2#> (28. sept. 2012).
- Knutsen, Anne Moseng 2007. *Variation du français a Abidjan (Côte d'Ivoire): étude d'un continuum linguistique et social*. Oslo: Universitetet i Oslo.
- Le Roux, Brigitte og Henry Rouanet 2010. *Multiple Correspondence Analysis*. London: Sage.
- Ringdal, Kristen 2004 *Enhet og mangfold*. Bergen: Fagbokforlaget.
- Rosenlund, Lennart 1995. Korrespondanseanalyse. Dataanalysens 'magiske øye'. *Sosiologisk tidsskrift* 1995/1, 55–78.
- Rosenlund, Lennart 2009. *Exploring the City with Bourdieu's. Applying Pierre Bourdieu's theories and methods to study the community*. Saarbrücken: VDM Verlag Dr. Müller.
- Røyneland, Unn (2005) *Dialektnivellering, ungdom og identitet: ein komparativ studie av språkleg variasjon og endring i to tilgrensande dialektområde, Røros og Tynset*. Oslo: Universitetet i Oslo.