

KORRESPONDANCEANALYSE –ET DYNAMISK OG FLEKSIBELT SOCIOLINGVISTISK VÆRKTØJ¹

Randi Benedikte Brodersen
Islands Universitet/Universitetet i Bergen

1. Indledning

Korrespondanceanalyse er en statistisk analysemetode som jeg bruger i min undersøgelse af danskeres sproglig tilpasning til norsk i Norge. I denne artikel vil jeg give en indføring i korrespondanceanalysen og vise et konkret eksempel på brug af denne metode.

Korrespondanceanalysen er en forholdsvis ny metode i dansk og nordisk sociolingvistik. Metoden er yderst velegnet til analyse af et stort og kompleks datamateriale og derfor særlig interessant i sociolingvistiske undersøgelser med mange informanter og variabler. Fremgangsmåden ved brug af korrespondanceanalysen og tolkningen og forståelsen af dens specielle visuelle fremstilling af resultater som punkter i et diagram kræver imidlertid en ret fyldig forklaring. Et eksempel på et diagram med en analyse af data fra mit pilotprojekt kan ses i figur 1 i slutningen af artiklen.

Artiklen indeholder efter indledningen disse 7 afsnit: 1) en præsentation af min undersøgelse og nogle af dens resultater, 2) en beskrivelse af en social faktor som er en af sociolingvistikens traditionelle variabler og en af mine variabler: nemlig køn, 3) en redegørelse for hvad korrespondanceanalyse er, og en begrundelse for hvorfor metoden er så attraktiv i sociolingvistisk sammenhæng, 4)

¹ Teksten er en bearbejdet version af mit indlæg på Det 12. Møde om Udforskningen af Dansk Sprog (MUDS), Århus Universitet, 9.–10. oktober 2008. Hjertelig tak til Peter Widell for teknisk og praktisk hjælp og til Daniel Apollon, Universitetet i Bergen, for at have bibragt mig en vis korrespondanceanalytisk indsigt. Tak til Det nordiske bibliotek i Athen for fine arbejdsforhold en uge i maj 2009 da jeg gennemarbejdede denne MUDS-artikel.

en forklaring på hvordan vi tolker et korrespondanceanalysediagram, 5) en omtale af den største udfordring ved korrespondanceanalysen, 6) en korrespondanceanalyse af de nævnte pilotprojektdata fra min undersøgelse og 7) til sidst en konklusion.

2. Danskeres sproglige akkommodation til norsk

Min undersøgelse består af et lille pilotprojekt² (med 11 informanter) og en større hovedundersøgelse (med 60 informanter). Informanterne er danske statsborgere, i alderen 21-77 år, og 65 % er kvinder og 35 % mænd. De kommer fra Bergen, Kristiansand og Oslo og har boet i Norge i mindst 3-4 år. Mange er norsk gift, nogle er dansk gift, og nogle er enlige. I pilotprojektet er informanterne ”håndplukket” af mig. I hovedundersøgelsen er informanterne et statistisk tilfældigt udvalg af danske statsborgere som er registreret i *Det Sentrale Personregister* i Norge (om udvælgelseskriterier se Brodersen 2003). I begge undersøgelser er det en vigtig og metodologisk pointe at de danske informanter er interviewet af norsk interviewer. Ved at bruge en norsk interviewer – og ikke en dansk – har jeg fået et datamateriale som egner sig til at belyse danskeres akkommodation til norsk.

Datamaterialet er indsamlet, bearbejdet og analyseret ved brug af kvantitative og kvalitative metoder og teknikker: halvstruktureret (telefon)interview, spørgeskema, transskribering (af bandede interview), ekscerpering (af lydlig træk), leksikalsk frekvensanalyse, korrespondanceanalyse og indholdsanalyse. I arbejdet med transskribering og frekvensanalyser har jeg brugt programmet CHILDES (Child Language Data Exchange System): <http://childes.psy.cmu.edu/> (Brodersen 2003).

Sproglig akkommodation drejer sig i min undersøgelse om de danske informanternes leksikalske og udtalemæssige akkommodation eller tilpasning til norsk. Leksikalsk akkommodation omfatter den norske og norskprægede del af informanternes ordforråd (i et interview). Fx norske talord, norske hverdagsord og udtryk som fx *bra*, *grei* og *være lei seg* (på dansk: *god(t)*, *i orden/OK/god/sød* og *være ked af det*)

² Betegnelsen pilotprojekt er et levn fra en tidlig fase i min undersøgelse. Dette projekt har udviklet sig til at blive en selvstændig delundersøgelse og har derfor en anden status og funktion end et pilotprojekt normalt har.

og dansk-norske hybridord. Udtalemæssig akkommodation er fx informanternes udskiftning af det lyse, danske *a* med det mørke, norske *a*, som i henholdsvis dansk *mand* og norsk *mann*. Flere eksempler på begge akkommodationsformer står i Brodersen 2003.

Informanternes leksikalske akkommodation til norsk varierer fra ca. 1 % til ca. 6 % norske eller norskprægede ord og udtryk af deres samlede ordforråd i et interview. De lave tal hænger sammen med at dansk og norsk har et stort fælles ordforråd. De leksikalske forskelle er med andre ord ret små, og de opleves også af nogle informanter som små. Informanternes udtalemæssige akkommodation varierer fra 4 til 67 % af deres *a*-lyde i en del af interviewet.³ De høje tal afspejler udtaleforskelle mellem dansk og norsk. Mange af informanterne har mange norske *a*-lyde og oplever tydeligvis et stort behov for at fornorske deres danske *a*'er.

Informanternes akkommodation til norsk er kendetegnet ved betydelig variation; nogle akkommoderer næsten udelukkende leksikalsk, i form af norske ord og udtryk, mens andre akkommoderer på flere sproglige niveauer samtidig: leksikalsk og udtalemæssig og evt. også morfologisk og syntaktisk. Nogle informanter akkommoderer meget, nogle meget lidt. Efter omfanget af deres akkommodation hører informanterne til i en af tre grupper: 1) De fleste taler et dansk-norsk blandingssprog, 2) nogle taler (overvejende) dansk, og 3) nogle taler (overvejende) norsk.

Der er en klar sammenhæng mellem leksikalsk og udtalemæssig akkommodation på den måde at mange af de danske informanter udtaler norske ord og udtryk på norsk, og de informanter som har mest leksikalsk akkommodation, har også omfattende udtalemæssig akkommodation, og de akkommoderer også ofte både morfologisk og syntaktisk. Leksikalsk og udtalemæssig akkommodation må derfor ses i sammenhæng, og i sammenhæng med sociale variabler. Og netop når vi ønsker at udforske sammenhænge og relationer og få dem fremstillet visuelt, er korrespondanceanalysen en metode med særlige muligheder som andre analysemetoder ikke har.

At få kortlagt variationen i danskeres sproglige akkommodation til norsk er især interessant, når vi kobler den til sociale makrofaktorer

³ Tallene for leksikalsk og udtalemæssig akkommodation stammer fra analyser af materialet fra hovedundersøgelsen.

og variabler som bl.a. køn, alder, civilstand og antal år i Norge. Her kan korrespondanceanalyse afdække og vise sammenhænge og relationer mellem variabler og udforske hvordan variabler interagerer, i én samlet analyse. Men først kort om køn.

3. Køn – en interagerende variabel

I min undersøgelse er køn en interagerende eller samvirkende variabel, dvs. den får betydning sammen med og i forhold til en række andre sociale variabler. Det viser sig fordi jeg har analyseret flere sociale variabler samtidig ved brug af korrespondanceanalyse.

Grunden til mit fokus på køn er at jeg vil understrege betydningen af at behandle køn i sammenhæng og samspil med de øvrige sociale variabler. Det er nemlig først muligt at finde ud af hvordan disse variabler virker sammen og har forskellige betydninger for sproglig variation (og akkommodation) når vi undersøger dem i samspil med hinanden. I sociolingvistiske tradition derimod bliver sociale variabler ofte undersøgt isoleret, dvs. hvad de betyder enkeltvis eller to og to i forhold til sproglig variation.

Køn er i et klassisk sociolingvistisk paradigme en vigtig, fast social makrofaktor og ofte en statisk, biologisk kategori: Informanter *er* kvinder eller mænd. Køn er blevet og bliver ofte stadig håndteret som en selvstændig og uafhængig social variabel, ligesom alder og klasse. I et nyere og konstruktivistisk paradigme, især inden for antropologisk og etnografisk orienteret sociolingvistik, er opfattelsen at køn er socialt konstrueret, og dermed bliver kønsbegrebet mere dynamisk: mennesker *konstruerer*, praktiserer, gør og forhandler sig frem deres kønsidentitet, eller en feminin eller maskulin identitet. Køn er her også knyttet til såkaldte praksisfællesskaber og identitet (se fx Eckert 1998 og Quist 2005). I nyere studier kombinerer og integrerer sociolingvister flere teoretiske ståsteder og indsigter. Det betyder at køn kan være både en social variabel som interagerer med fx klasse eller alder, og performans, i samspil med sociale faktorer som fx arbejde, uddannelse og civilstatus.

Den skitserede udvikling – og paradigmeskiftet – fra essentialisme til konstruktivisme er omtalt i en lang række udgivelser om

sprog og køn⁴ inden for de sidste 10-15 år. Paradigmeskiftet er også nævnt i et indlæg på en konference om Sprog og køn i København i 2008. Her peger Tove Bull på at vi ved for lidt om ”hjernen vår til å insistere på sosialkonstruktivismen som einaste teoretiske ståstad”, og vi er fanget af ”hegemoniske ideologiar som framstiller kjønn som essens” (Bull 2008:2). Bulls konklusion er at vi har brug for flere store undersøgelser med køn som social variabel, også selv om det ikke længere er politisk korrekt i sociolingvistikken (Bull 2008:9). Ja, vi har brug for flere af den slags undersøgelser, men kun hvis vi undersøger køn som interagerende og relationel variabel. Men her slår traditionelle sociolingvistiske korrelationsmålinger ikke til. Det vi opnår med den slags målinger, er nok bare at holde liv i det såkaldte kønsparadoks som er blevet citeret, forklaret og fortolket flittigt lige siden William Labov præsenterede det i 1970’erne (Labov 2001:292). For en præsentation og en diskussion af paradokset henviser jeg af pladshensyn til Mæhlum, Akselberg, Røyneland & Sandøy 2008:120. Her finder vi i øvrigt også en omtale af en vigtig undersøgelse af unges sprogbrug med fokus på bl.a. identitet og køn, analyseret ved brug af korrespondanceanalyse (Røyneland 2005). Køn har her relativ ringe indvirkning på de undersøgte sprogræk sammenlignet med en faktor som gruppeidentitet, som var ”mest utslagsgivende”, men til sammen er også køn og gruppeidentitet ”utslagsgivende” selv om den vigtigste forskel alligevel ikke er at finde mellem piger og drenge, men mellem to forskellige drengegrupper (Mæhlum, Akselberg, Røyneland & Sandøy 2008:120). Den nævnte undersøgelse er vigtig fordi den tydeligt viser at vi skal være varsomme med at generalisere med hensyn til sproglige kønsforskelle (– og generaliseringer kan bidrage til at forstærke en uheldig kønpolarisering), for der kan være større forskelle mellem forskellige grupper af drenge (eller piger) end mellem drenge og piger. Det er med andre ord især interessant at

⁴ Sprog-og-køn-forskningen har længe især fokuseret på forskelle mellem kvinders og mænds sprog, men der er i stigende grad også fokus på at problematisere og kontekstualisere køn og kønsforskelle, på køn som feminin og maskulin performans og på kønsforskelles betydning (jf. Pedersen 1998: 22). At der er stor interesse for emnet sprog og køn, især blandt kvindelige forskere, ser vi ved bare at kaste et blik på litteraturlistens forfatternavne ved publikationer om sprog og køn som bl.a. de seneste års håndbøger, antologier og artikler (fx Coates 1998, Ehrlich 2008, Eckert & McConnell-Ginet 2003, Holmes & Meyerhoff 2003 og Pedersen 1998).

fokusere på hvad køn betyder i samspil med andre variabler. Dette brede og samlende blik på køn kan korrespondanceanalyse give os, fordi metoden gør det muligt at udforske køn og andre variabler samtidig i én analyse, og da kan vi se hvilke variabler der betyder mest og mindst sammen med køn.

4. Hvad er korrespondanceanalyse, og hvorfor er korrespondanceanalyse særlig attraktiv i sociolingvistikken?

Min præsentation af korrespondanceanalysen bygger dels på materiale og indsigter fra et intensivt kursus i praktisk korrespondanceanalyse ved Daniel Apollon, Universitetet i Bergen, i efteråret 2007, dels på indføringsværker og konkrete eksempler på brug af korrespondanceanalyse (Apollon 2007, Hjelbrekke 1999, Hovden 1997, Høyen 2004 og Jæger 2003).

Korrespondanceanalyse er især blevet brugt i sociologisk forskning og er blevet internationalt kendt i løbet af 1980'erne på grund af sociologen Pierre Bourdieus forskning. Interessen er stadig stigende, også i Norden. I nordisk sociolingvistik er metoden dog mig bekendt hidtil kun blevet brugt i to norske ph.d.-undersøgelser (den ene er Røyneland 2005, den anden har jeg forgæves forsøgt at få fat i en henvisning til).

Korrespondanceanalyse er en deskriptiv, ikke-kausal og eksplorerende statistisk metode, som findes i forskellige varianter, både bivariate og multivariate. Bivariat korrespondanceanalyse kan, som navnet antyder, kun udforske sammenhænge og relationer mellem to variabler ad gangen. Multivariat korrespondanceanalyse derimod giver mulighed for at udforske sammenhænge og relationer mellem flere variabler samtidig. Det er kun denne variant jeg har brugt i min undersøgelse, og kun den jeg omtaler i det følgende.

Korrespondanceanalysens formål er at afdække de underliggende strukturer eller mønstre i et stort og kompleks materiale og vise relationer og sammenhænge mellem informanter og sproglige og sociale variabelkategorier, ved hjælp af et korrespondanceanalysediagram og ved brug af færrest mulige dimensioner. Og det er ofte to dimensioner (se figur 1 i slutningen af artiklen). Korrespondanceanalysen kan bruges på alle typer data og variabler der er homogene og kategoriske

Korrespondanceanalyse

(som fx variabelen køn med sine to kategorier, og de variabler med variabelkategorier som står i tabel 1 nedenfor).

Daniel Apollon karakteriserer korrespondanceanalyse sådan: ”CA [correspondance analysis, RBB] is still the only unique ready to use all-in-one multidimensional method allowing to study ”individuals” and their ”characteristics” with one unified representation” (Apollon 2007:19). Bordieu skriver at det er den metode ”som synes best egnet til å gripe alle de innsamlede observasjonene samtidig og til å avdekke de underliggende strukturene uten å innføre ulike former for forutinntatthet” (Bourdieu 2002: 68), dvs. korrespondanceanalyse utforsker sammenhænge i et materiale uden at tage stilling til det på forhånd, uden hypoteseopstilling, og giver derfor mulighed for at opdage nye og måske helt uventede sammenhænge. Metodens induktive og eksplorerende karakter kommer også tydeligt frem i en ofte citeret beskrivelse af Jean-Paul Benzécri, fransk lingvist og statistiker, som regnes som metodens grundlægger⁵:

Modellen må følge data, og ikke omvendt. Dette er [...] et leitt trekk ved tilpassinga av matematikken til samfunnsvitskapane; ein overflod av a priori spesifiserte modellar, som deretter vert konfronterte med data gjennom det ein kallar testar. Snart tener testen til å rettferdiggjere ein modell på eit felt der det finst fleire parametar 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 at trekkje ut strukturer med utgangspunkt i data (Benzécri 1973II: 6, citeret efter og oversat af Hjellbrekke 1999: 8).

Korrespondanceanalyse har altså en datadreven og dynamisk tilgang, afdækker strukturer og trækker den vigtigste information ud af et materiale: Det sker ved at analysen fjerner mindre vigtig information og på den måde reducerer et stort og kompleks datasæt til et mindre datasæt, *samtidig* med at data beholder deres kvalitative karakter. Vi kan sige at korrespondanceanalyse er en metode og en teknik som kan behandle kvantitative data på en kvalitativ måde, og den kan undersøge samspillet mellem mange variabler samtidig eller gradvis en efter en.

⁵ Benzécri udviklede korrespondanceanalysen i 1960'erne i protest mod den deduktionistiske lingvistik, som især Noam Chomsky repræsenterede (Hovden 1997:1 (vedlæg G)). Korrespondanceanalysens teoretiske rødder går dog længere tilbage i tid og er knyttet til statistikere som bl.a. Karl Pearson og Louis Gutman (Apollon 2007:19).

Korrespondanceanalysen har en relationel måde at tænke data på og er i stand til at udforske både enkeltvariablers betydning i en helhed og samspillet mellem mange variabelkategorier samtidig, og vi kan eksperimentere med at udelade og tilføje variabler i en analyse og se hvilke variabler der har størst eller mindst betydning, og hvordan forskellige variabler interagerer. Denne dynamik og fleksibilitet er en af de særlige fordele ved korrespondanceanalysen og gør metoden særlig aktuell og interessant i sociolingvistiske undersøgelser.

Korrespondanceanalysen giver et samlet og visuelt overblik over de relationer der eksisterer imellem alle eller udvalgte informanter og variabler i et materiale. Dette åbner for nye analyse- og tolkningsmuligheder af sociolingvistiske data, i forhold til de begrænsede muligheder der ligger i sociolingvistikens klassiske korrelationsanalyse.

Stærkt forenklet fremstillet består den korrespondanceanalytiske proces af fire trin. *Første trin* er omkodningen af data til binære størrelser, fx ved brug af Excel. Det betyder at alle informanter – eller andre undersøgelsesenheder – og alle variablers kategorier skal have koden 1 eller 0, hvor 1 betyder at besidde en egenskab, og 0 fravær af en egenskab. Dette gør vi enten manuelt eller ved hjælp af et program som koder informanter og variabelkategorier til et binært, disjunktivt (gensidig udelukkende) format. Analysen skelner ikke mellem informanter og variabler og heller ikke mellem afhængige og uafhængige variabler. Det betyder at informanter og variabelværdier alle bliver kodet og behandlet på samme måde, som egenskaber. Resultater af omkodningen er en tabel med rådata i rækker og kolonner, med nuller og ettaller, dvs. en binær datamatrice – eller datamatrix – som i tabel 1 nedenfor. *Andet trin* er en matematisk oversættelse af de binære data med alle række- og kolonneprofilers egenskaber, ved brug af et statistikprogram⁶ (i tabel 1 er en rækkeprofil en række med en bestemt kombination af ettaller og nuller, og en kolonneprofil er en kolonne med en bestemt kombination af ettaller og nuller). Resultatet er bl.a. to nye tabeller; den ene viser alle rækkeprofilers individuelle bidrag, den anden alle kolonneprofilers bidrag, dvs. de enkelte variabelkategoriers og informanters bidrag i den samlede

⁶ Som fx SPSS eller Analytica for Mac, Analytica Toolbox for Matlab, udviklet af Daniel Apollon, Universitetet i Bergen

Korrespondanceanalyse

variation. Disse tabeller er analysens råmateriale. *Tredje trin* er konverteringen og visualiseringen af hver af de to tabellers indhold til to reducerede repræsentationer i diagramform. *Fjerde trin* er så en samlet repræsentation af de to repræsentationer i ét fælles rum, som i diagrammet i figur 1 sidst i artiklen. Den grafiske gengivelse af dette rum er et korrespondanceanalysediagram med to eller evt. tre dimensioner. I denne proces bliver informanter og alle sociale og sproglige variabelkategorier omformet til såkaldte vektorer (som er profilerne som bliver omdannet til geometriske objekter med en størrelse eller en masse og en retning i et to- eller flerdimensionelt rum) og repræsenteret visuelt i diagrammet ved punkter.

VAR\INF	K1	K2	K3	K4	K5	M1	M2	M3	M4	M5	M6	Forklaring:
Kk	1	1	1	1	1	0	0	0	0	0	0	VAR \ INF: Variabel \ Informant
Km	0	0	0	0	0	1	1	1	1	1	1	Køn/Kvinde
Ald<37	1	0	0	1	0	1	0	0	0	0	0	Køn/Mand
Ald43-57	0	0	1	0	1	0	1	1	1	1	0	Alder under 37 år
Ald64+	0	1	0	0	0	0	0	0	0	0	1	Alder 43-57 år
AldFlyt<27	1	0	1	1	1	1	1	0	0	1	1	Alder over 64 år
AldFlyt>38	0	1	0	0	0	0	0	1	1	0	0	Alder Flyttet til Norge under 27 år
CivstN	1	0	1	1	1	1	1	0	1	1	1	Alder Flyttet til Norge over 38
CivstD	0	1	0	0	0	0	0	1	0	0	0	Norsk partner
ÅrN<11	0	0	0	0	0	1	0	1	1	0	0	Dansk partner
ÅrN16-19	1	0	0	1	0	0	1	0	0	0	0	År i Norge under 11 år
ÅrN>27	0	1	1	0	1	0	0	0	0	1	1	År i Norge 16-19 år
UddIkkeBog	0	0	1	1	0	0	0	0	1	1	1	År i Norge over 27 år
UddBog	1	1	0	0	1	1	1	1	1	0	0	Uddannelse Ikke boglig
FLYtMotFam	0	0	0	1	1	0	1	0	0	0	0	Uddannelse Boglig
FLYtMotArb	0	1	0	0	0	1	0	1	0	1	0	Motiv flytning til Norge: Familie
FLYtMotAnd	1	0	1	0	0	0	0	1	0	0	1	Motiv flytning til Norge: Arbejde
IdD	0	0	1	0	0	1	1	1	1	0	0	Motiv flytning til Norge: Andet
IdN	1	1	0	1	0	0	0	0	0	1	1	National identitet: Dansk
IdD&N	0	0	0	0	1	0	0	0	0	0	1	National identitet: Norsk
TsprD	0	1	0	0	1	1	0	1	1	0	0	National identitet: Dansk & Norsk
TsprN	1	0	1	1	0	0	1	0	0	1	1	Talesprog Dansk
SkspD	0	0	0	0	1	1	0	1	1	0	0	Talesprog Norsk
SkspN	1	0	0	1	0	0	1	0	0	0	0	Skriftsprog Dansk
SkspBI	0	1	1	0	0	0	0	0	0	1	1	Skriftsprog Norsk
Arbskr+	1	1	0	0	1	1	1	1	0	0	0	Skriftsprog Blandingsprog
Arbskr-	0	0	1	1	0	0	0	0	1	1	1	Arbejde med brug af skriftsprog
BørnsprN	1	1	1	1	1	1	1	0	0	1	1	Arbejde uden brug af skriftsprog
BørnsprD	0	0	0	0	1	0	0	1	1	0	0	Børns sprog Norsk
AkkFon++	1	0	1	1	0	0	1	0	0	1	1	Børns sprog Dansk
AkkFon-	0	1	0	0	1	1	0	1	1	0	0	Fon akkommodation: Meget omfattende
												Fon akkommodation: Ingen/tå enkeltord

Tabel 1. Binær datamatrice (mine pilotprojektdata)

I tabel 1 ser vi kolonner med pilotprojektets 11 informanter og rækker med 13 variabels kategorier, i en 11x31-matrice, med i alt 321 celler. Her har hver række – med hver variabelkategori (fx har køn to kategorier) – og hver kolonne, med hver informant, en egen *profil*. En rækkeprofil er den kombination af ettaller og nuller som står i hver række i tabel 1, og en kolonneprofil er den kombination af ettaller og

nuller som står i hver kolonne i tabel 1. Fx har informant K1, i første kolonne, følgende profil eller karakteristik: 1-0-1-0-0-1-0-1-0-0-1-0-0-1-0-0-1-0-1-0-0-1-0-1-0-1-0-1-0. Dette er de egenskaber, som kendetegner K1 og som svarer til alle aktuelle variabelkategorier. I tabellen ser vi fx at K1 for de to første variabler har række-egenskaberne 1-0-1-0-0. Det første 1-tal betyder at K1 er kvinde (Kk), det første 0 betyder ikke mand (Km). Det næste 1-tal betyder ”under 37 år”, og de to nuller til sidst betyder ikke ”43-57 år” og ikke ”over 64 år”. Alle egenskaber er som nævnt gensidigt udelukkende.

5. Hvordan tolker vi et korrespondanceanalysediagram?

Et korrespondanceanalysediagram, som i figur 1 i slutningen af artiklen, viser et todimensionalt plan som består af et koordinatsystem med en X- og en Y-akse. Akserne deler diagrammet i fire felter og vandret og lodret i to halvdele, og i disse felter er der punkter som repræsenterer informanter og variabelkategorier. Diagrammet giver et visuelt overblik over et materiale i form af en grafisk repræsentation. For at kunne tolke resultaterne i diagrammet, har vi brug for nogle tolkningsprincipper.

Punkter som er placeret i diagrammets yderkant, er de mest interessante. De giver som regel den betydeligste og tydeligste variation i materialet og bidrager dermed mest til opbygningen af en akse, mens punkter der ligger omkring x- og y-aksernes krydsningspunkt, det såkaldte barycenter, repræsenterer gennemsnitsværdier i materialet, og disse punkters bidrag er lave.

Vi tolker helst klynger af punkter sammen snarere end enkeltpunkter i forhold til hinanden og i første runde et punktsæt ad gangen, dvs. først de sociale variabelkategoriers placering og indbyrdes afstande og så informantpunkternes placering, eller omvendt. Og vi registrerer store og små *afstande* mellem punkter og punkters *relationer* og *oppositioner* i forhold til andre punkter; punkter der ligger langt fra hinanden har relativt mindre til fælles end punkter der ligger tættere hinanden som har relativt mere til fælles. Punkters afstand fra diagrammets midtpunkt er deres afstand fra gennemsnitsværdier.

Vi finder de vigtigste punkters bidragsværdier i de tilhørende række- og kolonnetabeller, som af pladshensyn desværre ikke er med i artiklen. Det er disse værdier der er udgangspunktet for punkternes

Korrespondanceanalyse

placering, og som angiver punkters bidrag til første eller anden dimension. De punkter der bidrager mest, ligger over gennemsnittet, eller mere præcist, indtager ekstreme positioner, dvs. ligger langt fra gennemsnittet. Når vi har disse værdier, kan vi undgå at fejltolke punkters placering.

Langs med X-aksen, som er den vigtigste akse, har vi hoveddimensionen, første dimension. Denne akse angiver den største reduktion af variansen (angivet ved hjælp af χ^2 -distance). Her finder vi de vigtigste modsætninger i materialet. Næstvigtigst er Y-aksen og anden dimension og de relationer der er koblet til den. Akserne hedder i figur 2 henholdsvis "CA axis 1" og "CA axis 2". I figur 2 dækker de to akser henholdsvis 32,99 og 18,01 og til sammen 51 % af variansen.⁷ Vi kan navngive diagrammets dimensioner efter hvilken struktur de får frem. Fx kan en tydelig struktur i figur 2 hedde tidsdimensionen.

Eftersom analysen indebærer en reduktion, vil ikke al variation i materialet blive forklaret i de første to eller tre dimensioner, som er dem vi kan se i en grafisk gengivelse. Der skal mange flere dimensioner til for at dække eller forklare den samlede variation i et materiale, og derfor kan vi også tale om et multidimensionalt rum. Det aktuelle og nødvendige antal dimensioner for at kunne forklare al variationen i et materiale er det mindste antal rækker eller kolonner minus 1 (Hjelbrekke 1999:197), dvs. i mit aktuelle tilfælde 10 dimensioner (fordi der er 11 informanter i pilotprojektet). Disse dimensioner kan vi få frem trinvis i flere diagrammer.

6 En udfordring ved korrespondanceanalysen

Måske er du en kritisk læser som undervejs i min lovprisende præsentation af korrespondanceanalysen stiller samme spørgsmål som min venlige fagfællelæser: "Er der slet ingen diskutabile punkter i metoden?" Jo. Korrespondanceanalysen har som alle andre metoder både fordele og ulemper. Det er som nævnt en deskriptiv og udforskende metode, og den kan ikke afdække årsagssammenhænge og teste hypoteser.

Den mest åbenlyse og største udfordring ved metoden er dog nok faren for fejltolkning. For selv om et korrespondanceanalysediagram til en vis grad er intuitivt fortolkbart, kan det være vanskeligt at

⁷ Hvis to akser til sammen når op på at forklare omkring 50% af variationen, har vi en god analyse.

overskue og tolke et diagram med mange punkter (som vi får når der er mange informanter og mange variabelkategorier i en analyse), især hvis diagrammet er tredimensionalt. , og for at kunne foretage en præcis vurdering og tolkning af punkternes placering og bidrag, har vi som nævnt altid brug for punkternes bidragsværdier i de tabeller som er grundlaget for punkternes placering og som hører sammen med korrespondanceanalysediagrammet, de statistiske mellemregninger så at sige.

7. Korrespondanceanalyse af sammenhænge mellem sproglig akkommodation og sociale faktorer

Nu analyseglimt af de pilotdata vi ser i diagrammet i figur 1 (i slutningen af artiklen). Glimt fordi jeg kun kan kommentere nogle få sammenhænge og resultater. Det er oplagt først at se på køn og variabelkategorier nær de to kønskategorier (Kk og Mk) i figuren. Det vigtigste at bemærke er at køn er en kategori som bidrager uhyre lidt til opbygningen af både akse 1 og akse 2. Det fremgår ikke af diagrammet, men af de tabeller som diagrammet bygger på og som viser alle variabelkategoriers bidrag til de to akser. Vi ved alligevel at der er forskel på kvinders og mænds identitet, og i diagrammet ser vi en klar forskel på de kategorier som kan siges at indgå i en kvindeidentitet, og dem der er en del af en mandeidentitet. Vi ser at kategorierne kvinde (Kk), norsktalende børn (BørnSprN), alder ved flytning under 27 år (AldFlyt <27), norsk gift (CivstN) og familie som flyttemotiv (FlytMotFam) alle ligger nær hinanden i venstre side af diagrammet, mens kategorien mand (Mk) ligger lidt mere isoleret i højre side, men dog nær kategorierne alder 43–57 år, boglig uddannelse (UddBog) og dansk identitet (IdD). Det betyder dels den sammenhæng at kvinderne er kommet til Norge i yngre alder end mændene og at deres flyttemotiv er familie snarere end arbejde, dels en sammenhæng mellem det at være mand, at have boglig uddannelse, dansk identitet og arbejde med brug af skriftsprog. De variabler og kategorier der bidrager mest til opbygningen af akse 1, er talesprog (mest TsprD og så TsprN), skriftsprog (kun SksprD) og de tidsrelaterede variabler som år i Norge (især ÅrN <11) og alder ved flytning til Norge (AldFlyt <27 og AldFlyt >38). I opbygningen af akse 2 er det også tidsrelaterede kategorier der er vigtigst: nemlig Ald64+ og ÅrN >27.

Korrespondanceanalyse

En interessant og tydelig modsætning ser vi mellem norsk identitet (IdN), ikke-boglig uddannelse (UddIkkeBog)), arbejde uden brug af skriftsprog (ArbSkr-) og blandingssprog som skriftsprog (SksprBl) over for dansk identitet (IdD), boglig uddannelse (UddBog) og arbejde med brug af skriftsprog (Arbskr+).

Så til de tre indtegnede linjer med pile i diagrammet. De viser en udvikling i tid og modsætninger relateret til de tre tidsvariabler: antal år i Norge, alder og alder ved flytning til Norge i forhold til variabler relateret til sprog og national identitet.

- Den fuldt optrukne linje – med variablen antal år i Norge – peger på modsætningen mellem kort og lang opholdstid i Norge: relationen mellem på den ene side, i diagrammets højre halvdel: under 11 år i Norge (ÅrN<11), ingen eller ringe sproglig akkommodation (AkkFon-), dansk talesprog (TsprD), dansk skriftsprog (SksprD), og på den anden side, i diagrammets venstre halvdel: 16-19 år i Norge (ÅrN16-19) og over 27 år i Norge (ÅrN>27) og omfattende sproglig akkommodation (AkkFon+) og norsk talesprog (TsprN) og norsk skriftsprog (SksprN).
- Den stiplede linje – med variablen alder – understreger modsætninger mellem alder under 37 år (ALD<38), alder 43-57 år (ALD43-57), nær diagrammets midtpunkt, og alder over 64 år (ALD64+).
- Den punkterede linje – alder ved flytning til Norge – viser modsætningen mellem flytning til Norge i en alder under 27 år (AldFlyt<27) og i en alder over 38 år (AldFlyt>38) og samtidig sammenhængen mellem den første og yngste aldersgruppe, det at være norsk gift (CivstN) og have norsktalende børn (BørnSprN) over for sammenhængen mellem den ældste aldersgruppe og det at være dansk gift (CivstD).

8. Konklusion og afsluttende bemærkninger

Med udgangspunkt i min undersøgelse af sproglig akkommodation har jeg forsøgt at vise at korrespondanceanalyse har egenskaber der gør den særlig velegnet som sociolingvistisk analysemetode. Det interessante ved korrespondanceanalyse i min undersøgelse er dels dens relationelle tilgang til data og dens visuelle fremstilling

af et kompleks materiale med mange informanter og variabler og deres indbyrdes relationer, dels dens empiriske fleksibilitet og dynamik, som er unik sammenlignet med andre statistiske metoder. Korrespondanceanalysen giver et godt billede af hvordan sociale faktorer interagerer og hvilken betydning de har enkeltvis og i forhold til hinanden i den helhed de er en del af.

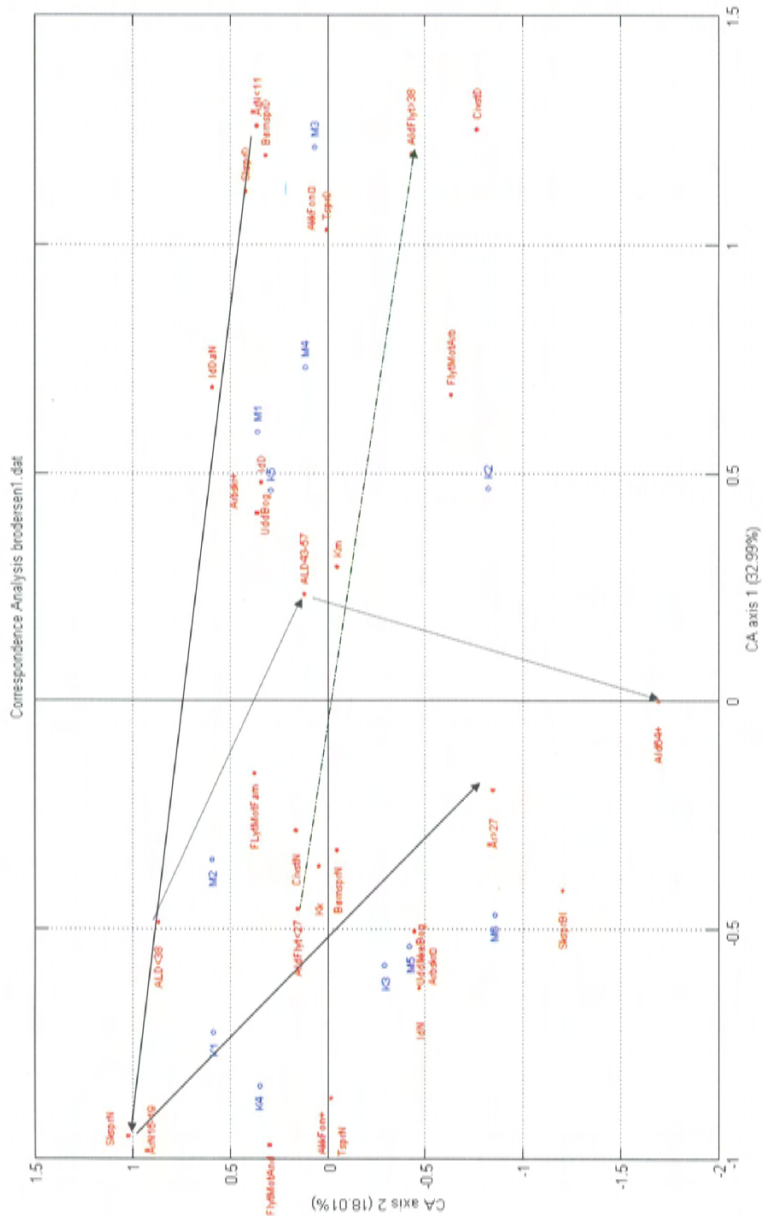
Min analyse) viser at de sproglige variabelkategorier og de tidsrelaterede variabelkategorier til sammen bidrager mest til de to hoveddimensioner (akse 1 og 2). De mest betydningsfulde sociale faktorer for informanternes akkommodation er: alder, antal år i Norge og alder ved flytning til Norge. Det er faktorer der indgår i tidsdimensionen. Analysen viser også at køn ikke spiller nogen særlig rolle. Køn leverer et nævneværdigt bidrag til opbygningen af de to hoveddimensioner.

Min konklusion er at korrespondanceanalysen beriger sociolingvistikken med en fleksibel og dynamisk metode og giver et nuanceret billede af samspillet mellem mange faktorer og kategorier samtidig. En ekstra pointe er at korrespondanceanalysen kan udforske og vise mange og nye sammenhænge i et kompleks materiale i én samlet, visuelt fremstilling.

Litteratur

- Apollon, Daniel (2007) *Korrespondance Analysis. An Introduction*. Ppt-præsentation på intensiv workshop efteråret 2007. Universitetet i Bergen.
- Brodersen, Randi Benedikte (2003) Leksikalsk og fonologisk akkommodation i ph.d.-projektet *Sproglig variation, sproglig akkommodation og sprogholdninger blandt danskere i Norge*. Nordisk dialektologi. Akselberg, Gunnstein, Anne Marit Bødal & Helge Sandøy (red.). Oslo: Novus:263–277.
- Bourdieu, Pierre (2002) *Distinksjonen. En sosiologisk kritikk av dømmekraften*. Oversat af Annick Prieur, med et indledende essay af Kjetil Jakobsen. Oslo: De norske bokklubbene.
- Bull, Tove (2008) Kva konsekvenser har teorier om kjønn for sociolingvistisk språk- og kjønnsforskning? *Den 7. Nordiske Konferanse om Sprog og Køn 24.-25. oktober 2008*. Københavns Universitet (upubliceret manuskript).

Korrespondanceanalyse



Figur 1. Korrespondanceanalysediagram: 11 informanter og 13 variabler (pilotprojektdata)

- Coates, Jennifer (red.) (1998) *Language and Gender. A Reader*. Oxford: Blackwell.
- Eckert & McConnell-Ginet (2003) *Language and Gender*. New York: Cambridge University Press.
- Ehrlich, Susan (red.) (2008) *Language and Gender*. London: Routledge.
- Gudiksen, Asgerd, Henrik Hovmark, Pia Quist, Jann Scheuer & Iben Stampe Sletten (red.) (2009) *Dialektforskning i 100 år*. København: Københavns Universitet.
- Hjelbrekke, Johannes (1999) *Innføring i korrespondanseanalyse*. Bergen: Fagbogforlaget.
- Holmes, Janet & Mirian Meyerhoff (red.) (2003) *Handbook of Language and Gender*. Oxford: Blackwell.
- Høyen, Marianne (2004) *Korrespondanceanalyse*. København: Akademisk Forlag.
- Jæger, Mads (2003) *Korrespondanceanalyse – et sociologisk værktøj*. Socialforskningsinstituttet. <http://www.sociology.ku.dk/kursfil/1143030642604.-pdf> (relokalisert 30.11.2009).
- Labov, William (2001) *Principles of Linguistic Change. Social Factors 2*. Mass.: Blackwell.
- Mæhlum, Brit, Gunnstein Akselberg, Unn Røyneland & Helge Sandøy (2008) *Språkmøte. Innføring i sosiolingvistikk*. Oslo: Cappelen.
- Pedersen, Inge Lise 1998. Sprog og køn. *NyS* 24:9-40.
- Quist, Pia (2005) *Stilistiske Praksisser i Storbyens Heterogene Skole. En etnografisk og sociolingvistisk undersøgelse af sproglig variation*. København: Københavns Universitet (ph.d.-afhandling).
- 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 (ph.d.-afhandling).