

## Personidentifikation

E. A. WRIGLEY (ed.): IDENTIFYING PEOPLE IN THE PAST. *Edward Arnold. London 1973. 159 sider. Ca. kr. 78,00.*

I maj 1970 afholdt det berømte *Institute for Advanced Studies* i Princeton, USA, en konference for historikere, demografer og videnskabsmænd fra mange andre fag over emnet *Nominal Record Linkage in History*. Herved forstår man sammenkædning af oplysninger på grundlag af de i historiske kilder anførte personnavne – altså en af de allerhyppigste fremgangsmåder i mange typer af historisk forskning. På grund af sine metodiske problemer er det et emne, der må antages at være af betydning såvel for historikere inden for de fleste af fagets discipliner som for forskere, der ud fra en samfundsvidenskabelig indfaldsvinkel ønsker at udnytte kildemateriale, der er afsat af tidligere tiders samfundsprocesser.

Et af konferencens afkast er den foreliggende samling af afhandlinger under redaktion af E. A. Wrigley, der er et fremtrædende medlem af en betydningsfuld gruppe social- og befolkningshistorikere ved universitetet i Cambridge. Gruppen har inden for den sidste halve snes år på en række punkter bidraget til øget viden om og klarere forståelse af udviklingen på områder som befolkningsstruktur, socialstruktur og familie- og husstandsstruktur igennem de sidste 3–4 århundreder. Hovedinteressen har naturligt nok været engelske forhold, men man har dog også beskæftiget sig med andre samfund. Hertil kommer, at gruppen selvfølgelig også har arbejdet med de involverede metodiske og forskningstekniske problemer, og under Wrigleys redaktion har man publiceret et par samleværker, der fokuserede på sådanne problemer.

Det fælles tema for de 6 afhandlinger, som samlingen består af, er de forskellige problemer, man stilles overfor, dersom man i en historisk undersøgelse ønsker at sammenstille oplysninger fra én kilde om en bestemt person med oplysninger om den samme person hentet enten fra andre kilder eller fra andre steder i den samme kilde.

I mange historiske undersøgelser er man ikke i tvivl om, hvorvidt den statsmand eller storkøbmand, man møder i en kilde, nu også er ham, man er interesseret i. I sådanne sammenhænge er disse *identifikationsproblemer* kun af triviel art, og de volder normalt ingen vanskeligheder. Men jo længere bort fra samfundets forskellige eliter man bevæger sig, jo større bliver usikkerheden i identifikationen, både principielt og reelt. Hvad skal man forlange, før man tør hævde, at de 2 Søren Jensen'er, man har for sig i materialet, var én og samme person, dengang de/han levede? Det er umiddelbart indlysende, at jo mere information, man har om hver af de 2 personer, jo flere sammenligninger af enkeltoplysninger kan foretages. Mange fuldstændige overensstemmelser øger selvfølgelig tilliden til, at man har foretaget en korrekt sammenstilling.

Ian Winchester, der tidligere blandt andet har skrevet en udmærket intro-

duktionsartikel til emnet (*The Journal of Interdisciplinary History*, Vol. 1, 1970–71, s. 107–124), bruger i sine to bidrag udtrykket *identifikatorer* (*identifying items* er det engelske udtryk), om den slags oplysninger, man bruger til at identificere enkeltpersoner med. Da cpr-numrene jo hører en senere tid til, er personnavnet den mest almindeligt brugte og den mest værdifulde enkeltidentifikator; men oplysninger som køn, alder/fødselsår, fødested, bopæl, indtægt, erhverv og stilling, familieforhold etc. kan fuldt så vel – men normalt mindre éntydigt end navnet – bruges til at karakterisere de personer, man interesserer sig for.

De kilder eller kildesteder, man ønsker at sammenstille, indeholder hver for sig normalt kun et begrænset antal identifikatorer. Sammenstillingen, hvis formål er at øge vor viden om enkeltpersoner eller grupper af personer, må derfor ske på grundlag af det beskedne antal identifikatorer, som er fælles for de 2 kilder. Hertil kommer, at disse ofte er anført i en ufuldstændig og/eller fejlagtig (herunder upræcis) form. Man behøver blot at tænke på varierende stavemåder for person- og stednavne eller på manglende præcision med hensyn til angivelse af fødested eller alder/fødselsår.

Den stadig voksende anvendelse af moderne databehandling i historisk forskning gør det naturligt at spørge, om der eventuelt skulle være hjælp at hente ved at anvende edb i det ganske omfattende søgningsarbejde, som i en række undersøgelsessituationer er nødvendigt for at løse identifikationsproblemer af denne type; det forekommer specielt rimeligt at stille dette spørgsmål i forbindelse med undersøgelser, der bygger på sammenstilling af oplysninger om mange enkeltpersoner fra kirkebøger, folketællinger, valg- og stemmelister, lægdsruller, udvandringsprotokoller og tilsvarende typer af massedata.

Problemet med at opstille kriterier for, hvornår man vil hævde, at en sammenstilling er korrekt – at der vitterlig er tale om én og kun én person – er principielt det samme, hvad enten man arbejder med et beskedent antal personer, eller man orienterer sin forskning mod forholdene inden for bestemte større grupper af personer, ud fra for eksempel en demografisk, en socialhistorisk eller en politisk-historisk indfaldsvinkel. Men i det sidste tilfælde vil formuleringen af sådanne kriterier være et mere påtrængende problem, fordi der er flere tilfælde, der skal bearbejdes, og fordi der normalt er færre identifikatorer, der kan bringes i anvendelse.

Hovedprincippet bag afhandlingerne er, at det må være muligt at foretage sådanne sammenstillinger primært på grundlag af efternavn og fornavn(e) – med et skarpt øje på kronologien. Andre identifikatorer kommer først ind i billedet på et senere tidspunkt. Fremgangsmåden er selvfølgelig særligt tillokkende ved den type demografiske undersøgelser, hvor man især på grundlag af kirkebogsoplysninger søger at rekonstruere familier og deres udviklingsforløb over bryllup, barnefødsler og ægtefællernes død. Særligt i samfund, hvor brugen af slægtsnavne allerede tidligt blev almindelig (for eksempel England og Frankrig) skulle der være gode muligheder for sådanne familie-rekonstruktionsstudier. I den

største af afhandlingerne gennemgår Wrigley da også sammen med R. S. Schofield principperne i familie-rekonstruktion på grundlag af datamatisk sammenstilling af kirkebogsoplysninger. Fremstillingen vil i sine hovedlinjer være af værdi for alle, der interesserer sig for denne type forskning; men mange historikere vil dog nok finde den vel avanceret sine steder – således forudsættes kendskab til matrix-algebra – og samtidig savner man flere steder en redegørelse for de foreslåede metoders fortrin frem for andre løsninger. De samme indvendinger kan rettes mod Mark Skolnicks bidrag om et specielt edb-program til løsning af disse matchningsproblemer under givne betingelser. At der kan være forskellige opfattelser af, hvad den bedste fremgangsmåde er, fremgår af de forskellige fremgangsmåder, som henholdsvis Wrigley og Schofield på den ene side og Skolnick på den anden side vælger, når de blandt flere mulige sammenstillinger skal vælge »den bedste«.

Begge disse mere teknisk orienterede bidrag er båret af en tilsyneladende sikker overbevisning om metodens fortræffelighed, både når det gælder rekonstruktion af familier, og når det drejer sig om sammenstilling af oplysninger om enkeltindivider. Det er imidlertid værd at mærke sig, at det i begge bidragene understreges, at man endnu ikke er kommet dertil, at man kan undvære menneskelig dømmekraft til at afgøre, om de særligt vanskelige tilfælde skal sammenstilles eller ej.

2 bidrag af Yves Blayo og David Herlihy nuancerer billedet af de muligheder, som denne teknik åbner for, ved at redegøre for konkrete undersøgelser. Flere af bidragene beskæftiger sig med mulighederne for at bruge navne-koder, der bedst muligt kan tage højde for ændringer i skrivemåde og fejlagtige stavemåder. Blayo supplerer disse ved at fremlægge ganske lovende resultater af nogle forsøg hermed på grundlag af et fransk materiale, der tidligere manuelt var blevet sammenstillet til familier. Men han understreger selv, at den valgte landsby måske er et dårligt eksempel, fordi kildematerialet er af meget høj kvalitet, og derfor måske ikke repræsentativt. Herlihys arbejde med toscanske skattelister fra det 15. århundrede er derimod et eksempel på, at det ikke har været muligt at få stort andet end alfabetisk ordnede lister ud af computeren. Men det kan selvfølgelig også være en god hjælp, hvis materialet er stort. En del af forklaringen på de nedslående resultater er utvivlsomt de opstillede kriterier til afgørelse af, om en sammenstilling var acceptabel, kombineret med en vis geografisk mobilitet i området, hvad der naturligvis mindsker mulighederne for at »fange« folk.

Det er nærliggende at spørge, om sådanne fremgangsmåder vil kunne finde anvendelse i tilsvarende danske undersøgelser.

Her vil en række forhold spille ind. Brugen af patronymier – især på landet – til langt op i det 19. århundrede betyder, at det ikke på grundlag af folks efternavn vil være muligt at køre en eller flere slægter ud på grundlag af et fælles efternavn for derefter at gennemføre det egentlige søgnings- og rekonstruk-

tionsarbejde inden for denne gruppe personer. Den i disse afhandlinger omhandlede metode kan altså ikke uden væsentlige ændringer overføres til et dansk materiale. Hertil kommer, at dansk navneskik tidligere har betydet en koncentration på forholdsvis få fornavne (se for eksempel Hans Chr. Johansen i *Historie*, IX, s. 95; og Poul Thestrup i *The Scandinavian Economic History Review*, XX, s. 9), hvad der på grund af brugen af patronymier selvfølgelig også førte til en koncentration på relativt få efternavne (se for slutningen af det 19. århundrede F. Nielsen m. fl., *Dansk Navneskik*, Kbh., 1899, s. 132–136). Udfra fordelingen af for- og efternavne og evt. andre identifikatorer kan man beregne skøn over sandsynligheden for at finde en bestemt person eller for at en foretagen sammenstilling må antages at være korrekt; Thestrup har givet et opløftende eksempel herpå (*anf. værk*, s. 9f), og emnet behandles i øvrigt flere steder i de 6 afhandlinger.

Imidlertid har nogle danske undersøgelser (således Hans Chr. Johansens undersøgelse af befolkningsforholdene i Rovsø herred omkring år 1800, *anf. værk*, s. 94f; og anmelderens egen undersøgelse af de metodiske problemer ved databehandling af et folketællingsmateriale, *Folketællingen 1845*, Århus, 1970, især s. 198–204) peget på ikke ubetydelige vanskeligheder ved at foretage denne type af identifikationer på tværs af forskellige kilder. Samtidig har andre (Thestrup, *anf. værk*; Anna Thestrup i *Fortid og Nutid*, XXIV; og Erik Høgh, *Vælgeradfærd i Danmark 1849–1901*, Kbh., 1972) fremlagt resultater af analyser inden for primært én kildegruppe, der tyder på, at vanskelighederne i denne forbindelse havde været af et yderst overkommeligt omfang.

Udfra hidtidige erfaringer er det altså ikke muligt at hente noget entydigt svar på spørgsmålet om metodens anvendelighed i en dansk sammenhæng. Man må derfor med spænding se frem til fremlæggelsen af undersøgelser, der har benyttet edb til sammenkædning af større mængder personoplysninger i et dansk kildemateriale, og sådanne undersøgelser er da også på vej. Man må så håbe, at disse kommende arbejder i højere grad, end tilfældet er med de foreliggende afhandlinger om *Nominal Record Linkage in History*, vil indeholde veldokumenterede sammenligninger af de relative fordele og ulemper ved henholdsvis datamatisk og manuel sammenkædning af personoplysninger i historiske kildematerialer.

[April 1974]

Jørgen Elklit

## Demografiske mikrostudier

KURT ÅGREN *m. fl.*: ARISTOCRATS, FARMERS, PROLETARIANS, *Essays in Swedish Demographic History, Studia Historica Upsaliensa*, 1973, 119 s.