

Utbildningssociologi som fysik och molekylärbiologi

Sverker Lundin

22 mars 2003

1 Två sätt att forska

I sin bok “Epistemic Cultures” berättar Karin Knorr-Cetina om sina erfarenheter från två platser där vetenskapsmän arbetar¹. Hon har observerat några högenergifysiker och molekylärbiologer och tycker sig se en väsentlig skillnad i hur dessa grupper av människor tänker om sin forskning och förstår sin omvärld. Knorr-Cetina vill beskriva de “kunskapsproducerande maskiner” som är verk samma i olika vetenskapliga discipliner. Med detta menar hon både hur arbetet i praktiken går till, hur man talar om det man håller på med och forskningsresultatens struktur, men dessutom och inte minst hur forskarna förhåller sig till varandra och sig själva. I sin redogörelse får man intryck av att de två platser där hon gjort sina deltagande observationer på flera plan står i motsatsförhållande till varandra.

Den fysikaliska forskningen innefattar ett stort antal människor som arbetar med ett gemensamt projekt. Forskningen är abstrakt och man ägnar stor möda åt att med hjälp av avancerad teknik och raffinerade metoder få fram “tecken” som står i ett dynamiskt förhållande till olika teorier med vars hjälp man försöker bestämma deras betydelse; Vad är det vi ser? Är det en störning eller en partikel? I denna värld, impregnerad av tekniska hjälpmedel och teorier som ger en känsla av att forskningsobjektet — partiklar som existerar bråkdelar av en sekund, bara som en effekt av ett gigantiskt maskineri — befinner sig oändligt långt från vardagsverkligheten och bara kan behandlas i form av undflyende och svårtolkade “spår” genererade av flera lager av förmedlande tekniker och statistiska teorier, uppstår ett språk där gränserna mellan det sociala, organiska och det materiella tekniska tenderar att suddas ut. Detektorn, den fantastiskt komplexa maskin som all verksamhet kretsar kring under det tiotal år som ett visst projekt pågår, tar i språket formen av en mänsklig individ, en vän eller ovän som föds, lever, blir sjuk, bråkar, är snäll, dum, åldras och slutligen dör. Tolkningsprocesserna, behandlingen av de data detektorn genererar, blir en kamp mellan det goda och det onda; störningarna är fiender som man med hjälp av olika tekniker lyckas döda. Tekniken och teorierna blir genom språket levande så att forskarna i kraft av de metaforer de använder tycks leva i en inte alls steril värld, utan i en organisk värld, där människorna bara utgör en bråkdel av alla

¹Knorr-Cetina (1999)

de aktörer som måste fås att fungera tillsammans, eller eventuellt bekämpas, för att nå projektets gemensamma mål.

Molekylärbiologerna behåller i alla avseenden sin individualitet. Här är forskningen strukturerad kring flera, relativt åtskiljda projekt, där varje individ har ett personligt ansvar och tar personliga risker i förhållande till sin egen forskning. Den politiska strukturen är hierarkisk och ledaren för laboratoriet har personligen makt över de forskningprojekt, av varierande antal, som pågår i hans laboratorium. Om forskning inom högenergifysik är långsamt kumulativ, kontinuerlig och impregnerad av abstrakta teorier för tolkning, är molekylärbiologin konkret — de mycket komplicerade experimentens utfall är fullständigt beroende av forskarens kroppsliga tekniska kompetens — vissa får det att fungera; andra inte. Den praktiska verksamheten är inte på samma intima sätt strukturerad med utgångspunkt från teorier för tolkning av resultat. Knorr–Cetina observerar också en motsatt tendens vad gäller språket, relativt högenergifysikerna. Molekylärbiologerna arbetar ju faktiskt med levande ting, bland annat möss, men dessa transformeras både genom en teknisk industriell process och genom dennas språkliga korrelat till “ting” som hanteras, språkligt och praktiskt, på samma sätt som vilken laboratorieutrustning som helst.

1.1 Liminal kunskap

Ett område där de två sätten att forska skiljer sig åt är i hur de hanterar metodiska problem, det vill säga hur de reagerar när verkligheten bjuder motstånd och ett visst (del)mål inte kan nås. För fysikernas del kan detta innebära att en viss förväntad partikel inte går att “se” på grund av, antar man, störningar orsakade av imperfektioner i detektorn. För molekylärbiologerna består ofta problemen av att ett experiment, som rent praktiskt är komplicerat att genomföra, misslyckas. Fysikernas sätt att hantera dylikt motstånd kallar Knorr–Cetina för “Care of the Self”. Hon använder begreppet *liminal kunskap* för att beskriva hur fysiker riktar uppmärksamheten mot de instrument genom vilka de får kunskap om sitt objekt. Man tar fram matematiska modeller och gör statistiska analyser av tekniken (den materiella och konceptuella) som producerar de tecken man har att tolka. Molekylärbiologerna utnyttjar istället “Blind variation and Natural Selection”, för att modifiera de experimentella metoderna. Detta innebär i princip, om jag förstått Knorr–Cetina rätt, att man mer eller mindre slumpmässigt, med utgångspunkt från sin personliga erfarenhet (som givetvis blir enormt viktig i detta sammanhang), ändrar förutsättningarna i experimentet tills man får det att fungera (eller ger upp). Sedan noterar man (i bästa fall) hur det var man gjorde. Om kunskap i högenergifysikernas värld ackumuleras och förmedlas på ett teoretiskt formellt sätt i form av rapporter och regelbundet återkommande seminarier, får molekylärbiologerna snarare på ett individuellt och informellt sätt, genom personliga samtal och inte minst genom komiska anekdoter, kunskap om hur man på ett bra sätt genomför olika procedurer i laboratorier (fast i de komiska anekdoterna ofta motsatsen, det vill säga hur man absolut *inte* skall göra).

2 En liten undersökning av korrespondensanalysen

Som blivande utbildningssociolog känns det naturligt att relatera dessa beskrivningar till mitt eget sätt att forska — reflexivitet är ju en dygd. Det jag tänker försöka undersöka är mitt förhållande till det register över gymnasieelever jag arbetar med. De tecken jag får fram på skärmen, om det så är siffror (till exempel 24%) eller visualiseringar av korrespondensanalyser, har de stora likheter med högenergifysikernas med hjälp av avancerade statistiska metoder och bildbehandling framtagna partikelspår. Verkligheten, de fysiska, eller kanske i mitt fall snarare sociala, eleverna finns inte närvarande i mina tankar. Begreppsligt arbetar jag med kvantiteter som genererats, inte av en partikeldetektor, men av en annan sorts detektor, byråkratisk, institutionell och mycket avancerad. Liksom fysikerna har jag som sociolog teorier för att tolka de tecken denna detektor genererar; jag förväntar mig vissa spår på min skärm (en ekonomisk pol, en kulturell pol, etcetera). Mina experiment “misslyckas” aldrig, som biologernas, utan jag är konstant involverad i processer av tolkning i relation till vissa relativt konstanta teorier. Den fråga som nu osökt infinner sig är om jag har samma kontroll som fysikerna över produktionen av de tecken jag studerar; i vilken mån skaffar jag mig liminal kunskap? Hur hanterar jag situationer där resultaten inte stämmer överens med mina teoretiskt grundade förväntningar?

De visualiseringar man kan göra med hjälp av korrespondensanalys är fascinerande eftersom de skapar en materiell koppling mellan vissa av sociologins metaforer och en visuell verklighet. I bildernas höger och vänster kan vi se våra (politiska) teoriers höger (den ekonomiska polen) och vänster (den kulturella polen), liksom vårt upp (samhällets toppar) och ner (invandrarnas i botten av samhället). Genom att dessa bilder är så lätta att tolka, är de, när det stämmer överens med teorins prediktioner, starka argument för teorins giltighet. Detta gör dem ytterst lämpliga som objekt för närmare granskning.

2.1 Sociologi som molekylärbiologi

Det instrument jag använder för att ta fram dessa bilder är ett dataprogram som heter SPAD. Jag ger SPAD data från mina register, ställer in ett antal parametrar, och får sedan nya data som kan visualiseras, antingen i SPADs grafiska modul, eller med hjälp av något annat program. Den del i denna process som jag (av nyfikenhet och misstanke) vill fokusera på är inställningen av den parameter som anger hur stor andel av den totala populationen som måste ingå i en viss modalitet för att denna skall vara aktiv i analysen. Den plats i programmets användargränssnitt där denna parameter ställs in är ganska anonym — känner man inte till hur programmet fungerar är sannolikheten stor att man inte upptäcker den. Vid den ruta där ett värde skall anges finns texten “Ventilation des modalités actives d’effectif inférieur à (en %)”, och värdet 2.0% är förinställt som ett förslag på lämpligt värde.

Trots att jag använder programmet rutinemässigt känner jag inte till någon praxis för vilket värde man bör ange, vilket antyder att parametern inte har någon

större betydelse. De tillfällen då parametern gör sig påmind är om situationen inträffar att ingen av modaliteterna (eller endast en) i en viss variabel innehåller mer än 2.0% av den totala populationen. Då fungerar inte programmet, det vill säga man får inte ut några data som kan användas för att göra en bild. Lösningen är att sänka parameterens värde från 2.0 till något lägre, exempelvis 1.0%, och sedan försöka igen.

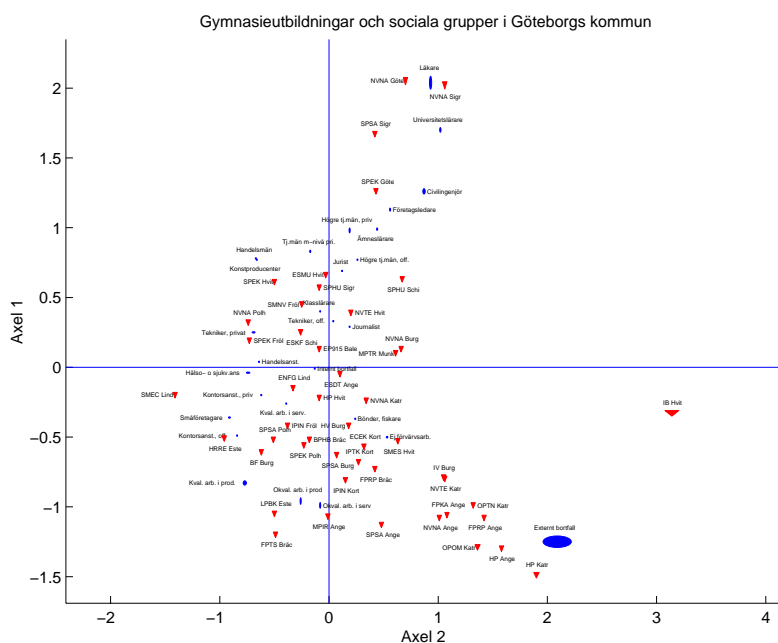
I ljuset av Knorr–Cetinas bok ter sig denna lösning på problemet mycket lik den molekylärbiologerna använde då de successivt, genom att med utgångspunkt från sin personliga erfarenhet variera vissa parametrar, lyckades få en viss experimentell procedur att fungera. Denna metod ger en uppsättning riktmärken för hur man får experiment att fungera, dock ingen exakt kunskap om vilken roll de olika komponenterna i ett experiment bidrar till att det “lyckas”, och i vilken mån de påverkar experimentets resultat.

Jag har två förslag till förklaringar av varför min sociologiska “experimentella” praktik påminner mer om molekylärbiologernas än högenergifysikernas. Dels är de resultat jag får fram — bilderna — inte utsatta för den sortens kritik som skulle motivera en närmare granskning av villkoren för dess tillkomst; jag är, mig veterligen, den enda på jordens yta som undersöker Göteborgs gymnasieskolor med hjälp av korrespondensanalys. En andra orsak är att det är besvärligt att ta fram liminal kunskap: jag är inte speciellt duktig på att använda SPAD — programmet är dokumenterat på franska, som jag inte behärskar (än); det är inte alls uppenbart hur man kan undersöka programmets funktion på ett systematiskt sätt.

2.2 Sociologi som högenergifysik

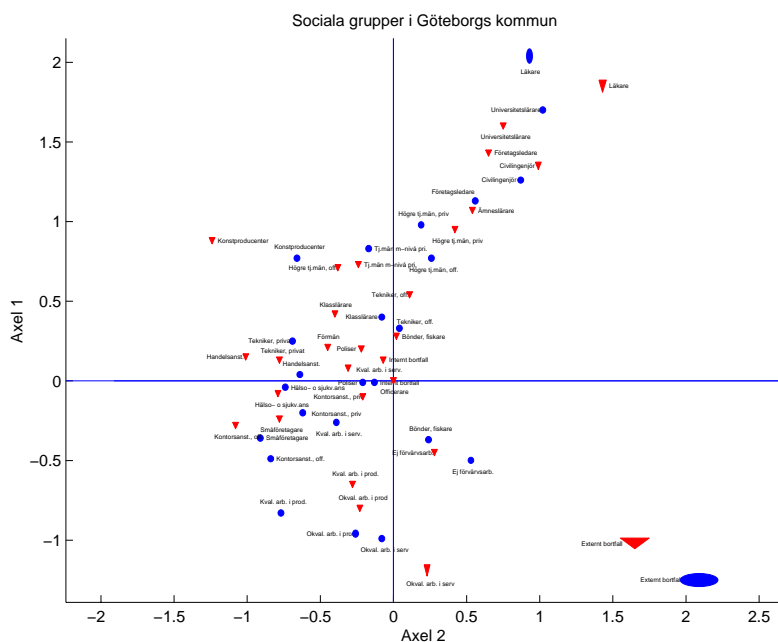
Låt mig kalla parametern jag är intresserad av att studera, det vill säga den parameter som avgör hur stor andel av totala populationen som måste ingå i en modalitet för att denna modalitet skall vara aktiv i analysen, för p . I mina analyser av Göteborg brukar jag låta $p = 1.0\%$. I figur 1 ser vi en typisk korrespondensanalys där utbildningar (program + skola) relateras till de yrken föräldrarna till eleverna på utbildningen har, det vill säga utbildningar och sociala grupper är *aktiva variabler* i analysen. Det sätt jag ämnar “analysera” data från SPAD är genom att titta var de sociala grupperna hamnar, givet olika inställningar i SPAD, och främst då olika värden för p . Eftersom vi generellt sett har fler individer i de sociala grupperna än på de respektive utbildningarna (som är många fler) antar jag att om de sociala grupperna hamnar på olika platser för olika värden för p så gör även utbildningarna det (något som visar sig vara sant vid en närmare undersökning). För att göra det möjligt att enkelt jämföra resultat från olika korrespondensanalyser har jag skrivit ett program för MATLAB där man ganska enkelt kan experimentera med de resultat SPAD genererar — jag har byggt ett litet laboratorium.

En inställning i SPAD som jag inte nämnt tidigare avgör om man skall ha möjlighet att för hand välja vilka modaliteter som skall vara aktiva (utöver detta val filtreras sedan modaliteterna med utgångspunkt från värdet på p). För att begränsa undersökningens komplexitet vill jag här, i motsats till vad man av praktiska skäl och vana alltid väljer annars, *inte* ha möjlighet att välja vilka

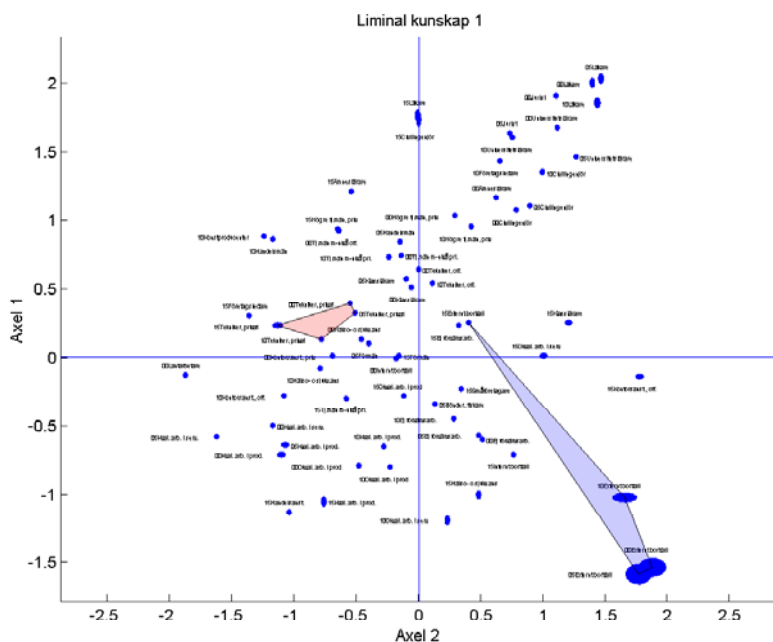


Figur 1: En korrespondensanalys av gymnasieutbildningar i Göteborg. Av pedagogiska och estetiska skäl visas inte alla utbildningar. Speciellt nära mitten är det många utbildningar som inte visas.

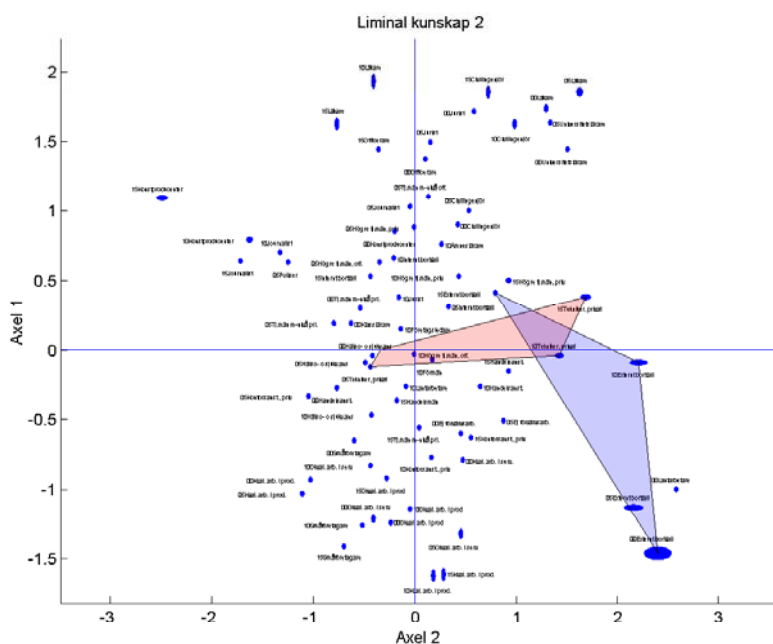
modaliteter som skall vara aktiva. Det finns två anledningar till detta; dels finns det ingen anledning att välja modaliteter för hand när jag gör denna undersökning av hur parametern p påverkar analysen, dels *måste* man, av någon okänd anledning välja bort minst en modalitet om man har denna möjlighet. Då jag gör mina analyser av Göteborg väljer jag slentrianmässigt bort den sociala gruppen Officierare, eftersom denna grupp har så få individer att den på grund av p ändå aldrig är aktiv i analysen. Jag blev förvånad när jag jämförde de sociala gruppernas placering i en analys då vi inte har möjlighet att välja vilka modaliteter som skall vara aktiva, och deras placering i figur 1. I figur 2 har elipserna exakt samma koordinater som de sociala grupperna i figur 1. Trianglarna visar de sociala gruppernas placering i den nya analysen, där den enda skillnaden (trodde jag) var att man där inte hade möjlighet att välja vilka modaliteter som skulle vara aktiva. Uppenbarligen hanterar SPAD (av någon outgrundlig anledning) data olika beroende på denna inställning. Skillnaden är ju som synes inte speciellt stor, men den är fullt synlig. En liten liminal kunskapsbit. Jag går vidare och gör, på fysikerns systematiska vis, analyser för $p = 0, p = 0.5, p = 1.0$ samt $p = 1.5$. I figur 3 visas de sociala gruppernas placering i alla dessa analyser; före varje modalitets namn syns det värde för p som symbolen motsvarar. Jag har gjort polygoner runt de positioner som Privatanställda tekniker respektive Externt bortfall hamnar på, för att man tydligare skall se hur de hamnar på olika platser beroende på värdet på p . Slutsatsen jag drar är att värdet på p har betydelse, speciellt för vissa mindre sociala grupper, till exempel lantarbetare. Skillnaden för $p = 0$ jämfört med $p = 1.5$ är stor för alla grupper. Vilken av bilderna är mest rättvisande? (Om vi använder standardvärdet $p = 2.0$ får vi



Figur 2: En jämförelse av de sociala gruppernas placering mellan två olika inställningar i SPAD. Den enda skillnaden är huruvida man får eller inte får välja aktiva modaliteter för hand. Samma modaliteter är i praktiken aktiva, enligt de textdata SPAD genererar.



Figur 3:



Figur 4:

ytterligare en annan bild.)

En naturlig fråga är varför man över huvud taget skall filtrera bort vissa modaliteter. Varför inte helt enkelt ha med alla? Anledningen är att modaliteter med få, men udda, individer kan påverka bildens utseende så mycket att den inte längre säger något väsentligt om det stora flertalet individer. Man måste därför välja vilken typ av information man vill att bilderna skall visa. Konstruktörerna av SPAD föreslår som sagt att man skall låta $p = 2.0\%$.

2.3 Göteborg jämfört med Stockholm

När jag nu tagit mig samman och gjort lite program för att jämföra bilder som skapats med olika inställningar i SPAD vill jag gärna titta lite på Stockholms kommun. Eftersom det redan gjorts en undersökning av Stockholm, och en bild, liknande de jag tar fram här, finns med i en publikation, tycker jag det vore intressant att se om jag kan få samma bild, och i så fall för vilket värde på p . I figur 4 visas motsvarande analys för Stockholm som figur 3 visar för Göteborg, bortsett från att jag här använt inställningen i SPAD som medger att man själv väljer vilka modaliteter som skall vara aktiva, för att kunna jämföra med den publicerade bilden (som jag gissar är gjort med denna inställning). Det verkar, om man jämför med sidan 92 i "Välfärd och skola" som att inget av de värden på p som jag valt ger riktigt samma bild. (För att bedöma likhet tittar jag på förhållandet mellan Externt bortfall, Kvalificerade arbetare i produktion och Okvalificerade arbetare i Service. Med tanke på att Externt bortfall tycks flyttas högre upp för högre p , samtidigt som Kvalificerade arbetare i produktion flyttas

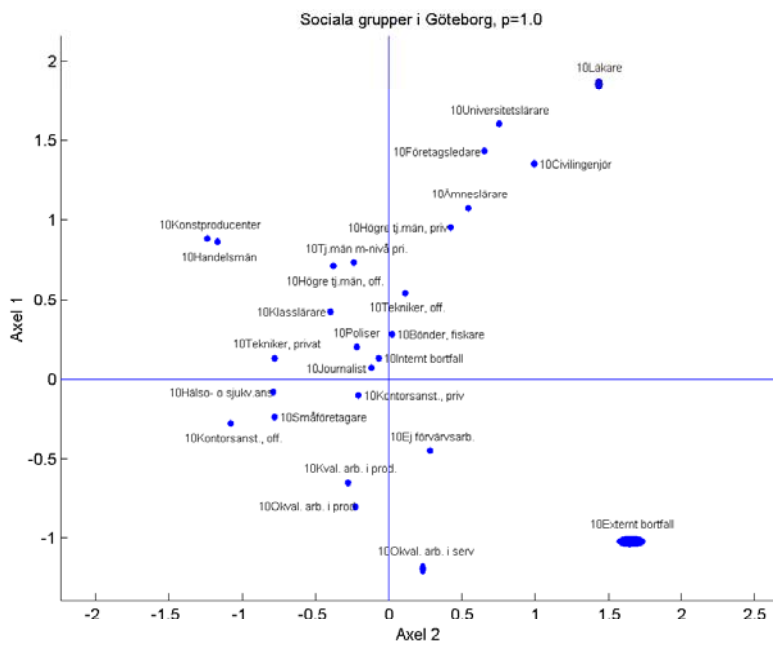
nedåt för höga p verkar det troligt att vi har $0 \leq p \leq 1$.) Om man betraktar varje korrespondensanalys som ett experiment är det inte på något vis märkligt att jag inte lyckas återskapa den tryckta bilden exakt. Tvärt om tycks just detta att det inte är så enkelt att replicera experiment vara typiskt för vetenskaplig verksamhet.

2.4 Lite förståelse och några problem

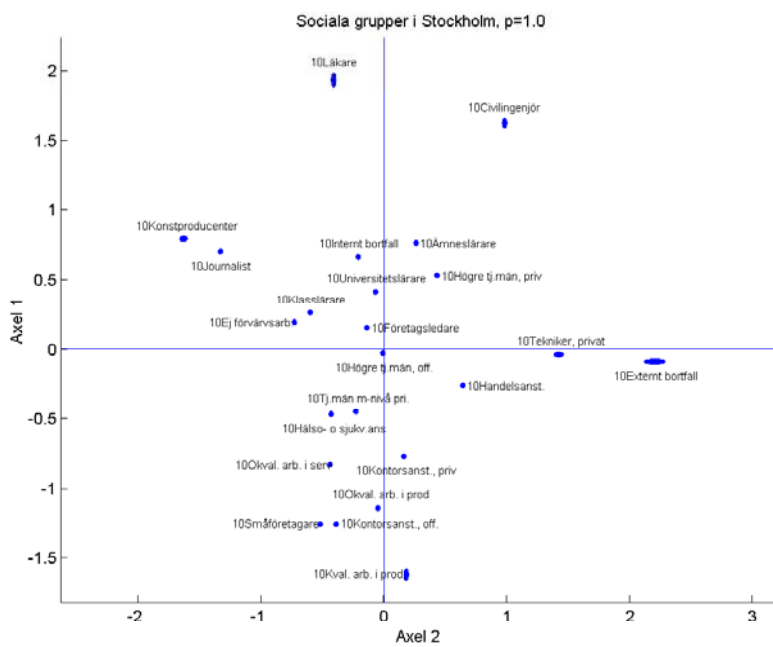
Om man jämför de bilder jag fått för Stockholm och Göteborg tycker jag mig se en möjlighet till konstruktiv användning av den nya informationen. Gruppen Externt bortfall har, beroende på p , två karaktäristiska positioner; en längst ner till höger och en till höger nära mitten. För höga p hamnar Externt bortfall nära mitten. Problemet är att tröskelvärde för när denna grupp flyttar sig är olika för Göteborg och Stockholm. För det värde jag — godtyckligt — valt i mina analyser, $p = 1.0$, tycker man sig se väsentligt skillnader mellan Göteborgs och Stockholms sociala strukturer (se figurerna 5 och 6). Skillnaden blir dock mycket mindre om vi för Stockholm låter $p = 0.5$ (se figur 7). Varför ger olika värden för p liknande bilder?

I analysen av Göteborg ingår 4770 individer, i Stockholm 6548 individer. Detta gör att om vi låter $p = 1.0\%$ så innebär detta i Göteborg att alla utbildningar med fler än 47 elever är aktiva, medan det för Stockholm innebär att bara utbildningar med fler än 65 elever är aktiva. Kontentan blir att man i Göteborg får 42 utbildningar av 119 aktiva, medan man i Stockholm får 22 av 179 utbildningar aktiva.² Den bild vi får speglar rekryteringen för de utbildningar som är aktiva, vilket gör bilderna beroende av p . (Intressant att notera kan ju vara att bilderna tycks påverkas på liknande sätt både för Göteborg och Stockholm, vilket måste betyda att skillnaden mellan de stora och de små utbildningarna är liknande på något sätt.) Min närmare granskning av konsekvenserna av p väcker, tycker jag, minst två frågor. 1) Vilket värde är lämpligt att använda? Frågan tycks ju motiverad, eftersom olika värden ger olika bilder. En lämplig procedur tycker jag verkar vara att bestämma sig för en *absolut* tröskel, i antal individer, som man anser lämpligt att ha med i analysen, och sedan, med utgångspunkt från antalet utbildningar, beräkna det motsvarande värdet för p . Alternativt skall man kanske låta $p = 0$, och om detta inte fungerar för hand ta bort de utbildningar som stör. 2) Hur skall man förhålla sig till korrespondensanalyser där vissa variabler har väldigt många modaliteter? Metodens stabilitet minskar då man låter p vara lågt. I praktiken kan detta leda till följande problem: Om p är för lågt kommer slumpmässiga variationer och mycket små utbildningar att störa analysen så att bilderna inte stämmer överens med teorin. Om p är för stort kommer inte tillräckligt många utbildningar med i analysen för att bilderna skall visa upp den variation i rekrytering som teorin förutsäger. Men om man väljer p enbart med utgångspunkt från bildernas överensstämmelse med sina teoretiska förväntningar, varför då gå omvägen över omfattande datainsamling snarare än att helt enkelt rita de bilder man vill ha för hand?

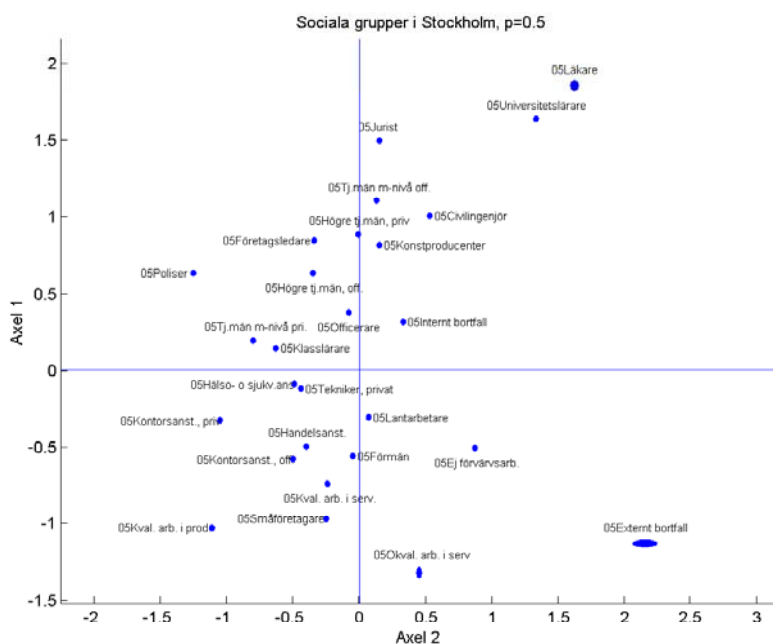
²Problemetets kärna är att konstruktörerna av SPAD antar att antalet individer per modalitet är proportionellt mot det totala antalet individer. Detta gäller ju inte för skolklasser, som tenderar att vara ungefär lika stora i alla kommuner, oberoende av hur många som bor där.



Figur 5:



Figur 6:



Figur 7:

3 Slutsatser

Karin Knorr–Cetina redogör i sin “Epistemic Cultures” för två sätt att forska. Utöver den inblick Knorr–Cetina ger i vissa högenergfysikers och molekylärbiologers värld konstruerar hon två idealtypiska forskare; fysikern och molekylärbiologen. Jag känner igen båda dessa idealtyper som sidor hos mig själv. Efter som jag i mitt sociologiska värv mest låter min inre molekylärbiolog styra (med trial–and–error och intuitiva beslut), blev jag inspirerad att som ett litet projekt angripa mina forskningsproblem som en sann fysiker, och skaffa lite liminal kunskap. Detta resulterade i att en rad problem med korrespondensanalysen blev synliga för mig. Även om jag tidigare kände till att dessa problem existerade har jag nu en betydligt djupare kunskap i vari de består, och vad man eventuellt kan göra åt dem.

Referenser

Knorr-Cetina, K. (1999). *Epistemic Cultures*. Harvard University Press.