

**Presentation av bok om svenskt och danskt
talspråk – skillnader och likheter utifrån
korpusdata
Projekt SweDanes**

Jens Allwood, Peter Juel Henriksen, Elisabeth Ahlsén, Magnus Gunnarsson,
Leif Grönqvist, Jan Katlev, Hans VappulaEllen

Syfte

Denna artikel är en kort presentation av en bok som ges ut av SweDanes-projektet. Projektet har haft stöd från NorFA:s språkteknologisatsning och har inneburit språkteknologisk metodutveckling för språkjämförande studier, baserade på två stora talspråkskorpusar, en svensk och en dansk. En del av projektets publikationer finns i referenslistan till denna presentation.

Boken riktar sig till alla som är intresserade av talspråk och jämförelser mellan danska och svenska. Vi har försökt att skriva den så att den inte ska kräva

specialistkunskaper i lingvistik. Vi vänder oss alltså till den språkligt intresserade allmänheten.

Boken bygger på en undersökning av skillnader och likheter mellan svenskt och danskt talspråk. Analysen kommer att grundas dels på korpusen Bysociale studier - BySoc (1,4 miljoner ord), som samlats in i Köpenhamn, och dels Göteborg Spoken Language Corpus (GSLC, 1,3 miljoner ord). I vår undersökning kommer vi att använda ordfrekvens som en indikator för att hitta intressanta lexikala och strukturella skillnader och likheter mellan svenskt och danskt talspråk..

Bakgrund

Den danska korpusen BySoc

Den danska talspråkskorpusen BySoc består av transkriberad dialog av ca 80 intervjuer med människor om deras liv. Inspelningarna gjordes i stadsdelen Nyboder i Köpenhamn, och omfattar således i stort sett bara en enda sociolekt. Könsfördelningen är tämligen jämn, och åldrarna varierar mellan ca 14 och 80. Samtalen är ofta över två timmar långa, och har inga förutbestämda ämnen, utan förloppet styrs av informanterna, som ofta väljer att berätta om lokalmiljön, sin barndom, skolgång, vänner, arbete eller studier. Många informanter som är över 25 år talar om sina barn. Många informanter under 20 år berättar om sina föräldrar. BySoc innehåller sammanlagt ca 120 timmars tal. Transkriptionerna är ortografiskt kontrollerade, dvs. antingen finns orden i den danska retskrivningsordbog eller i en kontrollerad undantagslista (t ex namn och konventionaliserade talspråksord som *ik'* och *vel'*). Information om korpusen finner man i Gregersen et al. (1981), Henrichsen 1997, 1998a, 1998b, Bilgram & Keson (1998) och Henrichsen (2002b).

Den svenska korpusen GSLC

GSLC-korpusen (The Göteborg Spoken Language Corpus) har i huvudsak inspelats och transkriberats under perioden 1978-2000 som del av ett antal olika projekt. (Den innehåller dock även några få inspelningar från 1960-talet.) Huvudkriteriet för uppbyggande av korpusen har varit att representera ett brett utbud av sociala verksamheter, och den innehåller inspelningarna från vitt skilda verksamheter som kyrkopredikningar, rättegångar, auktioner, middagssamtal, patient-läkarsamtal och shopping. Totalt innehåller korpusen drygt 360 inspelningar fördelade på 25 olika verksamhetstyper (för en översikt

över verksamhetstyperna, se bilaga 3). Transkriptionerna är gjorda enligt Göteborg Transcription Standard (GTS, se Nivre 1999b), och den ortografiska standarden är MSO6, vilket i princip är standardskriftspråk modifierat för att fungera för talspråk (se Nivre 1999a för detaljer). Korpusen beskrivs i Allwood (1998, 1999), Allwood et al (2000), Allwood (2001) och Allwood et al (2002). Information om taggning av korpusen finns i Nivre & Grönqvist (2001).

Jämförelse BySoc - GSLC

Skillnaden i bakgrundssyfte mellan BySoc och GSLC innebär att medan BySoc innehåller en systematisk variation av ålder, kön och socialgrupp hos de intervjuade talarna, är verksamhetstypen i huvudsak konstant, nämligen sociolingvistisk intervju eller informellt samtal. I de flesta fall betyder detta ganska långa samtal mellan två personer. GSLC innehåller istället systematisk variation mellan olika sociala verksamheter, antalet talare är mycket större och deltagarnas egenskaper är inte primära kriterier för urval, utan snarare en följd av valet av verksamhet, dvs de är varierade och mindre kontrollerade än i BySoc. Transkriptionerna är också mer varierade i längd. För vissa jämförelsesyften är det därför lämpligt att använda en delkorpus ur GSLC som innehåller informella samtal, som är mer lika dem i BySoc. Nedan kommer vi att utnyttja denna möjlighet och göra jämförelser dels med denna delkorpus, dels med GSLC som helhet. Transkriptionerna i GSLC kan återges i flera format, med olika stor närhet till standardortografi. För att underlätta jämförelse med BySoc använder denna studie en version av GSLC som ger transkriptionerna i svensk standard-ortografi (med utelämnande av interpunktion).

Val av språk i boken

När vi började skriva denna bok hade vi tre språk att välja mellan som huvudspråk i boken – engelska, svenska och danska. Först tänkte vi att vi skulle slippa välja mellan svenska och danska genom att välja engelska. Senare insåg vi att de tänkta läsarna av boken i huvudsak är svenskar och danskar och att engelska därför vore olämpligt. Det återstod då att välja mellan danska och svenska. Eftersom vissa av författarna till boken är primärt svenskspråkiga och andra är primärt danskspråkiga, gick inte detta heller så lätt. Vi löste det genom att alla skrev på sitt språk och att vi sedan översatte alla svenska artiklar till danska och alla danska artiklar till svenska. Vi fick på så sätt två versioner av boken – en dansk och en svensk. Man har därför som läsare ett val om man vill läsa boken på svenska eller danska.

Vår metod

Nu några ord om den metod vi använt i boken.

Ordfrekvens som hjälpmedel för att hitta skillnader och likheter mellan svenskt och danskt talspråk

En av tankarna bakom boken är att undersöka möjligheterna för att använda ordfrekvens som hjälpmedel för att hitta intressanta skillnader och likheter mellan svenskt och danskt talspråk. Utgångspunkten är att de frekventa orden i de flesta språk ofta är synkategorematiska ord eller s k funktionsord, t ex prepositioner, konjunktioner eller artiklar. Dessa ord bidrar till att ge språkliga uttryck deras struktur genom att deras betydelse är relationell och kan användas för att relatera betydelsen hos substantiv, verb och adjektiv till varandra eller till olika faktorer i talsituationen. Genom att studera dessa ord kan man uppnå det dubbla syftet att se vilka ord som är vanligast i ett språk och vilka strukturer dessa ord är förenade med. I ett nästa steg kan sedan denna information utnyttjas för att jämföra de vanligaste orden och strukturerna i danska och svenska med varandra.

När vi gjort jämförelsen har vi grundat oss på en ordlista där de 1 000 mest frekventa orden i svenska översatts till danska och tvärtom. Vi har sedan jämfört frekvenserna för översättningsekvivalenter i danska och svenska. Om vi t ex översätter det svenska ordet *jag* till danska, blir det *jeg*. Det visar sig sedan att *jeg* är mycket vanligare i BySoc än *jag* är i GSLC. Vårt nästa steg blir då att försöka förklara varför skillnaden finns.

I det första steget har vi därför upprättat listor på alla ord som relativt sett är vanligare på danska än på svenska och alla ord som relativt sett är vanligare på svenska än på danska. Vi har sedan tagit de ordpar som uppvisar störst skillnader och försökt förklara varför skillnaden finns. T ex kan skillnaden i frekvens mellan *jeg* och *jag* förklaras av att man i den danska korpussen ber de personer som intervjuas att berätta om sig själva, vilket gör det mycket naturligt att använda pronomenet *jeg*. Några intervjuer av denna typ finns inte i det svenska materialet, vilket alltså inte på samma sätt motiverar användningen av pronomenet *jag*.

När vi i det andra steget försökt hitta förklaringar har vi utnyttjat våra kunskaper i svenska och danska (våra intuitioner), men vi har också tnyttjat

vårt inspelade och transkriberade material (de två korpusarna). Vi har försökt hitta exempel i korpusen och försökt understödja våra försök till förklaring genom att se på det material som finns i korpusen.

Förklaringstyper

De förklaringar vi funnit är av följande typer:

(1) Skillnader i inspelat material.

Väldigt många av skillnaderna, till exempel att inte bara *jeg* utan att de flesta personliga pronomen är vanligare i det danska materialet, beror på att det danska materialet består av personliga intervjuer med människor om deras familj och livsöden. Det blir då naturligt att använda personliga pronomen.

(2) Skillnader mellan danska och svenska.

Dessa kan vara av följande typer:

Skillnader i polysemistruktur, dvs skillnader i de betydelser eller funktioner ord har.

Historiskt besläktade ord betyder ofta i stort sett samma sak, men skillnaden mellan dem leder till skillnader i frekvens. Ett exempel på detta är det vanligaste ordet i både svenskt och danskt talspråk nämligen ordet *det*. Detta ord är relativt sett något vanligare på svenska. En förklaring till detta är att några av ordet *det:s* funktioner på svenska tas över av *der* på danska. Till exempel heter *det finns* typiskt *der er*. Svenska *det* har alltså en vidare användning än danska *det* och därför också en annan polysemistruktur. Det har åtminstone en funktion eller betydelse utöver danska *det*.

Skillnader i vanliga ord och uttryck.

Ibland har av orsaker som inte är så lätta att analysera vissa flerordsuttryck blivit vanligare på det ena språket, som inte är det på det andra språket. Detta medför ofta att de ord som ingår i uttrycket också blir vanligare. Genom att uttrycket *et eller andet* är vanligt på danska, men inte på svenska, blir *andet* också relativt sett vanligare.

Besläktat med detta är att det kan finnas mer svårförklarade preferenser i vad som sägs. Man tycker helt enkelt mer om att använda ett ord på det ena språket än på det andra

Tidigare arbeten

När det gäller tidigare arbeten om skillnader och likheter mellan svenska och danska finns ingenting som primärt utgår från inspelat talspråk. Vad som finns kan möjligen sorteras i följande huvudkategorier.

1. Reseparlörer
2. Översikter ifrån ett traditionellt grammatiskt perspektiv
3. Bidrag till översättningsteori

Vad som är gemensamt för alla tre kategorierna är att de haft ett kvalitativt, normativt oftast skriftspråkligt perspektiv. Vår utgångspunkt i denna uppsats är i kontrast till detta snarare empiriskt, kvalitativt och korpusbaserat. En annan skillnad är att tidigare studier negligerat verksamhets (eller aktivitets)skillnader och antagit ett generellt standardspråk som inte varierar med aktivitet.

Disposition

Boken består av följande delar.

1. Introduktion – där bokens syfte och bakgrund förklaras.

2. Korpusbaserade jämförelser.

Detta är bokens huvuddel och här görs en mängd olika jämförelser baserade på ord som är märkbart vanligare på svenska eller danska.

3. Aktivitetsinflytande

Här diskuterar vi kortfattat några av de effekter som skillnaderna i inspelat material haft för våra korpusar.

Några kända egenskaper hos svenska och danska

Svenska och danska har mycket stora likheter, men vi skall här helt kort nämna några typer av skillnader som finns och som beskrivs mer utförligt i boken.

Uttal

Den allra största upplevda skillnaden mellan svenska och danska brukar vara uttalet.

När det gäller enskilda ljud är det så att vokaluttalet av samma eller motsvarande grafem (skriftbokstäver) i de två språken ofta skiljer sig åt. Till

exempel är flera danska vokaluttal mer lika tyskans än vad som är fallet för svenska. Några exempel är att svenskans speciella vokaluttal för grafemen *u* och *y* med olika typer av läpprundning inte alls motsvaras av danska *u* och *y* och att uttalet av danskans grafem *a* i ord som *gade* inte är särskilt likt svenskans uttal av *a* i *gata*, utan artikuleras längre fram med lite mer sluten mun och läpparna mer dragna åt sidorna.

Konsonantuttalet skiljer sig också på flera punkter, t ex. att svenskans grafem *b*, *d* och *g* uttalas ”hårt” genom att luftströmmen helt spärras av, medan danskans *b*, *d*, och *g* uttalas mycket ”mjukare” utan total avspärrning av luftströmmen och ibland helt faller bort.

Det danska ”stødet”, som artikuleras som ett ”stöt ljud” med en snabb avspärrning av luftflödet i svalget, är ett välkänt fenomen, som svenskan saknar och som kan vara svårt att få grepp om för en svensk. För en dansk kan däremot svenskans två ordaccenter, som gör att ord med samma stavelser kan uttalas olika och få olika betydelse, t.ex. sv. *stégen* – akut (da. trinnene) och *stégen* – grav (da. stigen), vålla problem. I själva verket har dessa två fenomen ett samband, så att akut accent i svenskan ofta motsvaras av *stød* i danskan.

Ibland skiljer sig även ordbetoningen (eller tryckaccenten). Några exempel är sv. *'erkänna* (tryck på första stavelsen)– da. *er'kende* (tryck på andra stavelsen) och motsatsen i sv. *för'bud* – da. *'forbud*, ordpar som är klara fall av ”falska vänner” när det gäller just uttalet.

Ordböjning och grammatik

De svenska och danska ordböjningssystemen för substantiv skiljer sig åt, även om båda språken har två genus (*häst-en*, men *hus-et* på svenska). Till exempel kan de ord som motsvarar varandra på de två språken ha olika genus (sv. *en människa* – da. *et menneske*; da. *en bi* – sv. *ett bi*). Plural böjning och bestämdhetsböjning kan också skilja sig, som när svenskans pluralform *hus* motsvaras av danskans pluralform *huse*. I svenskan heter det i bestämd form plural *husen*, i danskan *husene*.

Det finns även skillnader i verbböjning. Några exempel är sv. *hette* – da. *hed*; sv. *högg* – da *huggede* och sv. *träffade* – da. *traf*, samt hjälpverbsskillnaden i sv. *mannen har åkt hem* – da. *manden er kjørt hjem*.

Andra skillnader som man lätt lägger märke till är pronomenskillnader, som i da. *det er mig* – sv. *det är jag* och ordföljdsskillnader, som i da. *børnene tog deres bøger frem* – sv. *barnen tog fram sina böcker*.

I vissa fall kan skillnader som vi finner mellan den svenska korpusen, som i huvudsak består av västsvenska talare, och den danska korpusen, som består av Köpenhamns-dialekt, bli mindre när man jämför med andra dialekter, vilket kan ge oss en liten inblick i språkens gemensamma historia. I sydsvenska dialekter kan man säga *där är* i stället för *det finns*, liksom på danska (*der er*). Ett annat exempel är att betoningsskillnaden mellan sv. *den 'här* och da. *'den her* inte finns mellan svenska och jylländska där man säger *den 'her*.

Dessa är alla skillnader som gör att det inte är helt lätt att gå mellan svenska och danska trots att språken är mycket lika. Vad vi har försökt göra som grund för beskrivningen är att finna relativt klara och förhoppningsvis intressanta skillnader mellan svenskt och danskt talspråk, varav en del inte tidigare beskrivits, genom att utgå från frekvensskillnader för enskilda ord i de två stora talspråkskorpusar som boken bygger på. I följande avsnitt ger vi några exempel på vad vi funnit.

Några exempel på korpusbaserade iakttagelser om skillnader mellan svenska och danska

I detta avsnitt ges ett par exempel från boken på vad en frekvensbaserad jämförelse kan ge för information om svenskt och danskt talspråk. De två exempel vi har valt är pronomen och vaghetsmarkörer, vilka båda har en hög frekvens i talspråk jämfört med skriftspråk och därför kan antas ha talspråksspecifika användningar.

Pronomen

Vi kunde se ovan att vissa pronomenskillnader är kända och lätta att lägga märke till. Vad kan då frekvensskillnader mellan den svenska och den danska korpusen ge oss för information?

För det första kan vi konstatera att det också finns stora likheter mellan användningen av personliga pronomen i danska och svenska.

Men letar vi efter språktypiska skillnader i jämförbara aktiviteter (informella samtal) finner vi tre sådana typer:

1) Kopula-satser av typ ”*det är PRONOMEN (...)*” har på danska ett pronomen i ackusativ

det är mig der sidder og klovner i det

altså det er ham der er instruktør

Men på svenska ett pronomen i nominativ

det är jag som tackar

det är han som säljer det

Denna typ av konstruktioner är vanlig i båda språken och ger en frekvensskillnad mellan nominativ- och ackusativformer. En annan skillnad vi här kan se är att svenskan saknar motsvarighet till danskans *der*, vilket gör att *som* används mer i svenskan än i danskan.

2) Artighetsformer är nästan utdöda i båda språken, men danskans *de/dem/deres* förekommer i korpusen, medan svenskans titlar, *han, hon* och *ni* som tilltal inte förekommer.

3) Opersonlig agent, dvs *man*, som i *man vi ju inte vara till besvär* är vanligare i svenskan.

Reflexiva pronomen (*sig/sej*) är betydligt mer vanliga i den svenska än i den danska korpusen och detta har flera förklaringar.

- Svenskan har fler reflexiva verb eller konstruktioner som inte har danska motsvarigheter. Några exempel är att *lära sig* på svenska, heter *lære* på danska, *ta sig till X* på svenska, *komme til X* på danska, *ta på sej* på svenska, *tage noget på* på danska, *ha med sig* på svenska, *have med* på danska, och *söka sej in till stan* på svenska heter *tage ind til byen* på danska.

- Svenskan har flera syntaktiska konstruktioner där *sig/sej* + infinitiv (objekt med infinitiv) motsvarar *att han/hon/den* + finit verb, som i exemplen *som dom anser sej ha* (da. *som de mener at have*), *vad de tror sig ha hört eller sett* (da. *hvad de tror at have hørt eller sett*). Danskan har enbart i ”högre skriftspråk” uttryck som ”mener sig att have”.

- I svenskan används konstruktionen *i och för sig* flitigt och påverkar frekvensskillnaden.

Vaghetsmarkörer

Ganska och rätt

Det finns tre svenska ord för vaghet som vållar problem vid översättning til danska, *ganska*, *ungefär* och *något*. *Ganska* har en direkt släkting i det danska ordet *ganske*, men dels är det ordet inte alls lika vanligt i danska som i svenska, dels används det danska ordet ofta i sin ”tyska” betydelse, ’helt, fullständigt’:

og du sidder bare ganske stille...

Det svenska uttrycket *och du sitter bara ganska stilla* skulle snarare innebära att man inte satt fullständigt stilla.

De två ord som oftast följer direkt efter *ganska* i GSLC är *mycket* och *bra*, och dessa ords danska motsvarigheter *meget* och *godt* bestäms oftast av *ret*:
det er ret godt betalt...

På svenska kan detta emellertid också uttryckas med *rätt*; jämför följande översättningar av de danska exemplen:

det är ganska/rätt bra betalt

Man kan också notera att det svenska *ganska* och *rätt* kan följas av *så*:

det är ganska/rätt så bra betalt

Det är dock förhållandevis sällsynt i båda språken.

Både *rätt* och *ret* betyder också ’korrekt’, vilket gör det svårt att tolka frekvenssiffrorna, men man kan uppskatta att 57% i GSLC har betydelsen. Det danska *ret* används i *ganska*-betydelsen i ca 90% av fallen. En annan komplikation är dock att *ret* även kan användas i negerade satser, där man på svenska byter ut *ganska* mot *väldigt*, *särskilt* eller liknande:

ja fordi der skal ikke ret meget til før man kommer til at...

och detta gav inte särskilt stor utdelning...

Två andra danska ord med liknande betydelse som *ganska* är *rimelig* och *temmelig*, som är betydligt vanligare än de svenska orden *rimlig* och *tämligen*. Exempel ur BySoc:

mine forældre er også rimelig gamle...

der var temmelig varmt...

Ungefär

Den mest direkta motsvarigheten till det svenska *ungefär* är danskans *omtrent*, men det senare är inte alls lika vanligt. Två näraliggande synonymer, *omkring* och *cirka*, finns i båda språken med ungefär samma frekvenser. Det danska *nogenlunde* är vanligare än det svenska *någorlunda*, men frekvenserna är alltför låga för att väga upp *ungefärs* övervikt.

De danska orden *omtrent*, *cirka* och *omkring* används alla för att modifiera kvantifierande uttryck, vilket det svenska *ungefär* också kan göra:

*så det var jo **omtrent** to månedslønninger...*
*ja mm // men er det ikke nu her **omkring** din alder hvor...*
*...de sparer jo **cirka** tohundrede tusinde om året ik'...*

Ungefär kan också modifiera något som inte är direkt kvantifierande:

*// kan man säga så **ungefär***
*jag skulle kunna /// kunna säga **ungefär** som så här att jag / jag läste en BOK*

I sådana fall motsvaras ordet av flera olika danska uttryck, som *noget i retning* eller *sådan cirka*:

*ja altså: **noget i retning** af hvad: hvad dronningen taler ik'*
*Grønnedal / det ligger: syd på / **sådan cirka** på højde med Bergen*

En annan vanlig användning av *ungefär* är vid liknelser:

*...nä men det är **ungefär** som att skilja på form och innehåll...*

På danska motsvaras *ungefär* som möjligen av *ligesom*, men det ordet förmedlar inte vaghet otvetydigt:

*altså det er **ligesom** at spille musik for køer ik'...*

Studerar man de svenska förekomsterna av *ungefär* ser man att ordet ofta används som ett mycket generellt vaghetsuttryck:

*...hon kom ju i i sandaletter och bikini **ungefär** hit...*
*för den **ungefär** vänder borta hos oss...*
*...att vi kallar det för huvudet då **ungefär** och så har vi där kroppen då...*
*...hur man ska kunna planera för / framtiden och inte bara / tuta och köra **ungefär**...*
*...och sedan drar man **ungefär** medelvärdet av det...*

De danska orden *omtrent*, *omkring* eller *cirka* används inte på detta sätt. Det verkar i själva verket vara ett drag i svenskan som inte alls finns i danskan. Det

svenska *typ* används oftast på samma sätt, och har en bedövande övervikt jämfört med det danska *type*.

...i andra länder som det är liksom **typ** USA och så där...

...men det är inte så att det finns i **typ** alla religioner eller så...

...och då hade han **typ** burit hem honom då...

Något

En liknande vaghet, om än inte exakt den samma, kan danska talare föra in i sina yttranden genom t ex *sådan noget*:

...det hedder vist Astor Burger eller **sådan noget**...

...hvad havde jeg været halvandet år eller **sådan noget**...

... vi går mere til / private fester og **sådan noget** ik'...

Sådan noget förekommer 3 264 gånger i BySoc, medan den direkta svenska motsvarigheten *något sådant (nåt sånt)* bara förekommer 290 gånger i GSLC. Man bör dock notera att det danska uttrycket så att säga inte sprider vagheten över hela yttrandet på samma sätt som det svenska *typ/ungefär*, utan endast modifierar det närmast föregående uttrycket.

Orden *något/någon/några* är betydligt vanligare i BySoc (15 390 förekomster) än i GSLC (8 035 förekomster).

På svenska har de negerade varianterna *inte något/någon/några* alternativen *inget/ingen/inga*, vilket inte alls är lika vanligt på danska. *Ingen* är den enda formen som är representerad i BySoc och den används för alla de tre svenska alternativen:

... de har **ingen** interesse i de...

... det var **ingen** gode råd i hvert fald...

... det er der **ingen** af os der gør...

Detta förklarar en del av underskottet för *något/någon/några* i svenskan, men alls inte hela.

Övriga vaghetsuttryck

Det finns ett antal andra vaghetsuttryck som är värda att nämnas.

- *Så att säga/så at sige*

Uttrycket är mycket vanligare på svenska än på danska, 993 förekomster mot 33. Detta kan delvis förklaras med skillnad i aktivitetstyp, eftersom uttrycket har ett lite högre stilvärde och GSLC har en mycket mindre andel informella samtal än BySoc.

- *Liksom/ligesom*

Uttrycket är betydligt vanligare på svenska än danska.

- *Runt/rundt*

Ordet är vanligare i BySoc än i GSLC och dessutom är det mindre vanligt att ordet används i cirka-betydelsen i BySoc än i GSLC. Man kan också notera att det på svenska var betydligt vanligare att *runt* fick modifiera en nominalfras direkt, oftast spatialt (t.ex. *runt ögonen*), medan det på danska oftare användes för att modifiera ett verb (t.ex. *han behøver ikke rende rundt med brødkniv*).

- *En eller annan/en eller anden*

Uttrycket förekommer bara en enda gång i GSLC, men 263 gånger i BySoc. Det motsvaras oftast av *något/någon* på svenska.

det var några försäljare som ville sälja nåt

det var en eller anden seger som ville selge et eller andet

Sammanfattningsvis uttrycks vaghet på lite olika sätt på danska och svenska.

När man på svenska föredrar *ganska* och *ungefär* används på danska ofta *noget/nogen/nogle*, ofta i kombinationen *sådan noget*. Särskilt värt att notera är att den mycket generaliserade vaghetsanvändningen av *ungefär* och *typ* som finns på svenska (t.ex. *och då hade han typ burit hem honom då*) inte alls hittades i det danska materialet.

Även övriga undersökta vaghetsuttryck skiljer sig märkbart åt i frekvens, så att de är vanligare på svenska än på danska (med undantag av *cirka* och *omkring*, vilka har ungefär samma frekvens). Det är svårt att dra slutsatser av detta eftersom vagheten uttrycks så olika, och eftersom orden används till annat än vaghet. En intressant fråga är om svenskar skulle vara mer försiktiga och föredra att uttrycka sig vagare.

Slutord

Om dessa och andra likheter och skillnader mellan svenskt och danskt talspråk kan man läsa mer i den bok som ges ut av SweDanes-projektet. Det är vår förhoppning att beskrivningen av dessa korpusbaserade iakttagelser skall bidra till vår kunskap om hur svenskar och danskar talar, dels till gagn för språkvetenskap i allmänhet och språkteknologi, dels som en hjälp till danskar och svenskar att förstå varandra och lyckas uttrycka sig bra i svensk-danska samtal.

Referenser

Allwood J. 1998. Some Frequency Based Differences between Spoken and Written Swedish, *Papers from the 16th Scandinavian Conference of*

Linguistics, Department of Finnish and General Linguistics, Turku University, pp 18-2

Allwood J. (Red) 1999. Talspråksfrekvenser - ny och utvidgad upplaga, *Gothenburg Papers in Theoretical Linguistics* (S 21), Göteborg University, Department of Linguistics.

Allwood J. 2001. Capturing Differences between Social Activities in Spoken Language. In Kenesei, I., & Harnish, R.M. (Eds.) *Perspectives on Semantics, Pragmatics and Discourse*. Amsterdam. John Benjamins, pp. 301-319

Allwood, J., Björnberg, M., Grönqvist, L., Ahlsén, E. & Ottesjö, C. 2000. The spoken language corpus at the department of linguistics, Göteborg University. *FQS – Forum Qualitative Social Research*, Volume 1, No. 3 – December 2000.

Allwood, J., Grönqvist, L. Ahlsén, E. & Gunnarsson, M. 2002. Annotations and tools for an activity based spoken language corpus. In van Kuppevelt, J. (ed.) *Current and New Directions in Discourse and Dialogue* (Proceedings from SIGDial workshop Aalborg Aug. 2002). Kluwer Academic Publishers.

Allwood, J., Henrichsen, P. J., Ahlsén, E., Grönqvist, L. & Gunnarsson, M. 2002. Transliteration Between Spoken Language Corpora – Moving between Danish BySoc and Swedish GSLC; *Gothenburg Pap. in Theoretical Ling.* 86. (Forthcoming in *Nordic Journal of Linguistics*.)

Bilgram, T. & Keson, B. 1998. The Construction of a Tagged Danish Corpus; NODALIDA-98

Gregersen, F. et al. 1991. *The Copenhagen Study in Urban Sociolinguistics*; 1 + 2. København: Reitzel.

Henrichsen, P. J. 1997. Talesprog med ansigtsløftning. Utilisering af et stort dansk talesprogs-korpus. *Instrumentalis* 10/1997, IAAS, Københavns Universitet.

Henrichsen, P. J. 1998a. Talesprog med netstrømper, Internet-adgang til et stort dansk talesprogs-korpus. *Instrumentalis* 12/1998, IAAS, Københavns Universitet.

Henrichsen, P. J. 1998b. Peeking into the Danish living room. In *Proceedings from NoDaLiDa* 1998.

Henrichsen, P. J. 2002a. *Fyrre Kilometer Kryds og Bolle - metoder til grammatisk opmærkning i største skala*; Akademisk Forlag: NyS 30

Henrichsen, P. J. 2002b. *Some Statistically based Differences Between Spoken and Written Danish*; Gothenburg Pap. in Theoretical Ling. 88.

Henrichsen, P. J. (in press) Siblings and Cousins – Statistical Methods for Spoken Language Analysis; Acta Linguistica Hafniensia.

Nivre J. 1999a. Modifierad Standardortografi.. Göteborg University, Department of Linguistics.

Nivre J. 1999b. Transcription Standard Version 6.2. Göteborg University, Department of Linguistics.

Nivre, J. & Grönqvist, L. 2001. *Tagging a Corpus of Spoken Swedish*; International Journal of Corpus Linguistics 6/1.