

Några frekvensbaserade skillnader mellan svenskt och danskt talspråk

*Jens Allwood, Elisabeth Ahlsén, Peter Juel Henriksen, Magnus Gunnarsson, Kaarlo
Voionmaa, Hans Vappula & Leif Grönqvist*
Department of Linguistics
Göteborg University
&
CMOL
Copenhagen Business School

1 Syfte

I denna uppsats presenterar vi några första preliminära resultat av att undersöka 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. Det primära syftet med uppsatsen är att visa på metoden och sättet att arbeta. Vi hoppas att inom kort något mer ingående presentera de resultat vårt arbete lett fram till.

2 Bakgrund

2.1 Den danska korpusen BySoc

Den danska talspråkskorpusen BySoc består av transkriberad dialog av ca 80 så kallade Labovianska intervjuer, gjorda 1986-1990. 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'*). Korpusen beskrivs i Gregersen (1991).

2.2 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 inspelningar 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 1). Transkriptionerna är gjorda enligt Göteborg Transcription Standard (GTS, se Nivre et al. 2003), och den ortografiska standarden är MSO6, vilket i princip är standardskriftspråk modifierat för att fungera för talspråk (se Nivre 1999) för detaljer). Korpusen beskrivs i Allwood (1999), Allwood et al (2000), Allwood (2001) och Allwood et al (2002).

2.3 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 jämförelse-syften 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 någpn gång att utnyttja delkorpus. Transkriptionerna i GSLC kan återges i flera format, med olika stor närhet till standardortografi. För att underlätta jämförelse använder denna studie en version av GSLC som ger transkriptionerna i svensk standard-ortografi (med utelämnande av interpunktion).

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

En av tankarna bakom denna artikel 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 100 mest frekventa orden i de flesta språk 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 strukturererna i två språk (i detta fall danska och svenska) med varandra.

2.5 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

4. Läroböcker

Vad som är gemensamt för alla fyra 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.

Det finns inte mycket litteratur för en bredare publik om förhållandet mellan modernt danskt och svenskt talspråk. Läroböcker i svenska för danskar och i danska för svenskar är vad som kommer närmast - men de flesta av dessa tar inte alls eller nästan inte alls upp de talspråkstypiska dragen.

Dansk Grammatik for Svenskere (Christensen 1999) är ett exempel. Boken omfattar nio grammatik-artiklar som är ordnade i den traditionella ordningen med innehållskategorierna först och funktionsorden sist. Substantiv, adjektiv och verb står således först – och fyller det största utrymmet. De talspråkstypiska kategorierna, adverb, interjektioner och personliga pronomen, behandlas bara på 10 sidor, som i stort sett endast går igenom böjningsformerna. Det vanligaste ordet i danskt och svenskt talspråk, *det*, omnämns bara för sin (mindre vanliga) funktion som formellt subjekt, inte i sin (vanliga) funktion som allmänt diskurspronomen. Talspråkets kanske allra mest typiska kategori, interjektionerna, behandlas i bokens absolut kortaste avsnitt, på bara sju rader. Interjektioner som spelar en avgörande roll i talspråkets grammatik – *ja* och *nej*, återkopplingsord, diskurs-tagging (t ex danska *ik* och svenska *va*) – nämns inte alls. Med andra ord: Christensen 1999 och de övriga böckerna i samma genre vänder sig till den läsande och skrivande allmänheten, inte till den talande och lyssnande.

Det finns dock åtminstone en god kandidat till dansk-svensk framställning som tillgodoser talspråket, nämligen boken *Dansk som Nabosprog* (Cramer et al. 2001). Boken inleds med ett

kapitel om internordisk språkförståelse, som diskuterar hur bra eller dåligt danskar och svenskar förstår varandra – inte minst när de själva skall bedöma detta. De flesta undersökningar visar en ganska dålig förståelse över landsgränsen, men Cramer framhäver att danskar och svenskar förbluffande fort lär sig förstå varandra, om de i eget intresse tvingas att göra ett allvarligt försök. Boken går igenom de senaste 1000 årens språkhistoria i Norden, som inledning till de tyngre kapitlen i mitten, där dansk och svensk fonologi, ortografi, morfologi och ordförråd. behandlas kontrastivt. Boken innehåller några små avslutande kommentarer om de gemensamma syntaktiska dragen i danska och svenska, men inget om skillnader i talspråkets särdrag på detta område.

Avslutningsvis vill vi nämna de två största och mest renommerade översättningsordböckerna mellan danska och svenska, nämligen

Molde, B. (2000) Nordstedts Dansk-Svenska Ordbok, Nordstedts Ordbok.

Munch-Petersen, V. P.; E. Harmann (2001) Svensk-Dansk Ordbog, Gyldendal.

Dessa två – som båda finns i förkortade utgåvor – är på 726 respektive 1218 sidor. Munch-Peteren är den mest djupgående och täcker t.ex. även finlandssvenska. Trots dessa böckers oomtvistliga auktoritet måste båda sägas vara delvis föråldrade, eftersom de fortfarande i de senaste utgåvorna har med sig en del icke aktuellt material från den tid då de skapades, strax efter andra världskriget.

3 Jämförelse

I tabell 1 nedan visas de 100 vanligaste orden i BySoc och GSLC.

Tabell 1. De 100 vanligaste orden i BySoc och GSLC

BySoc				GSLC		
Rang	Ord	Antal	Ppm	Ord	Antal	Ppm
1	det	74130	55530	det	79782	63074
2	ja	47126	35301	är	37942	29996
3	og	41532	31111	och	35856	28347
4	jeg	39362	29485	ja	33394	26400
5	er	38307	28695	att	30998	24506
6	så	36188	27108	jag	29174	23064
7	der	32300	24195	så	26773	21166
8	ikke	24855	18618	som	19466	15389
9	var	23458	17572	inte	19152	15141
10	i	23341	17484	har	18800	14862
11	har	18996	14229	vi	18670	14760
12	at	18921	14173	på	18038	14260
13	mm	18187	13623	man	17779	14055
14	de	17564	13157	då	17557	13880
15	ik'	17099	12808	i	17244	13632
16	men	16943	12691	du	16539	13075
17	jo	15776	11817	en	15789	12482
18	du	15553	11650	ju	15618	12347
19	en	15406	11540	men	14934	11806
20	altså	14239	10666	dom	14766	11673
21	på	13343	9995	den	14505	11467
22	vi	13326	9982	för	14389	11375
23	han	13314	9973	här	14128	11169
24	man	13095	9809	om	12956	10242
25	nej	13033	9762	m	12267	9698
26	sådan	12837	9616	eh	11606	9175
27	med	11384	8527	kan	11157	8820
28	til	10628	7961	med	10466	8274
29	den	10590	7932	där	9797	7745
30	også	10390	7783	var	8373	6619
31	kan	10032	7514	ska	8258	6528
32	nå	9891	7409	till	8254	6525
33	noget	9648	7227	nä	8060	6372

34	for	9099	6816	vad	7730	6111
35	af	7511	5626	eller	6537	5168
36	da	7018	5257	va	6440	5091
37	eller	6789	5085	ett	6079	4805
38	hvad	6624	4962	äh	5532	4373
39	som	6611	4952	nu	5247	4148
40	hun	6609	4950	han	5173	4089
41	et	6456	4836	av	4813	3805
42	havde	6220	4659	också	4706	3720
43	om	6162	4615	ha	4668	3690
44	meg- et	5726	4289	får	4512	3567
45	skal	5643	4227	när	4378	3461
46	ved	5591	4188	skulle	4305	3403
47	nu	5265	3943	väl	4089	3232
48	fordi	5232	3919	bara	3951	3123
49	godt	4622	3462	tycker	3904	3086
50	kun- ne	4553	3410	säga	3816	3016
51	hvor	4359	3265	sen	3780	2988
52	nog- en	4154	3111	vara	3748	2963
53	når	4113	3081	finns	3635	2873
54	været	3851	2884	alltså	3560	2814
55	skulle	3757	2814	myc- ket	3480	2751
56	være	3705	2775	hur	3459	2734
57	lige	3614	2707	kom- mer	3425	2707
58	her	3591	2690	något	3412	2697
59	dem	3390	2539	lite	3412	2697
60	fra	3346	2506	tror	3369	2663
61	ud	3345	2505	sej	3350	2648
62	hvis	3342	2503	vet	3301	2609
63	tror	3230	2419	någon	3262	2578
64	vil	3227	2417	hade	3093	2445
65	nok	3197	2394	kan- ske	3054	2414
66	min	3147	2357	öh	2965	2344
67	år	3095	2318	göra	2877	2274

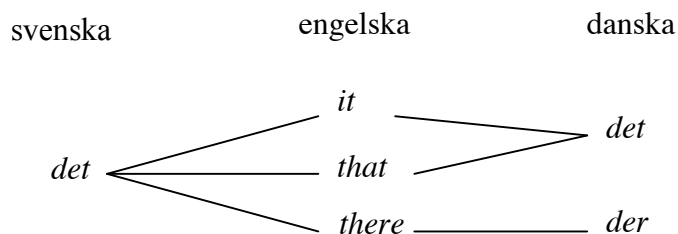
68	mig	2927	2192	gör	2721	2151
69	have	2910	2179	blir	2699	2133
70	lidt	2840	2127	går	2665	2106
71	i	2822	2113	vill	2613	2065
72	bare	2814	2107	just	2595	2051
73	kom	2762	2069	ut	2467	1950
74	sige	2750	2060	ni	2401	1898
75	op	2573	1927	måste	2355	1861
76	ham	2521	1888	lik- som	2341	1850
77	synes	2510	1880	bra	2321	1834
78	mang -e	2457	1840	ta	2306	1823
79	blev	2300	1722	alla	2269	1793
80	kom mer	2287	1713	in	2193	1733
81	ind	2278	1706	jo	2137	1689
82	mere	2148	1609	upp	2133	1686
83	gik	2115	1584	så- dant	2124	1679
84	sagde	2059	1542	hon	2123	1678

85	to	2047	1533	menar	2086	1649
86	over	2045	1531	någo- nting	2082	1645
87	bliver	2019	1512	nej	2001	1581
88	må	2010	1505	säger	1894	1497
89	gör	2008	1504	väld- igt	1890	1494
90	sig	1998	1496	mer	1883	1488
91	siger	1964	1471	natu- ren	1862	1472
92	fik	1961	1468	få	1850	1462
93	går	1931	1446	mej	1819	1438
94	ville	1884	1411	utan	1775	1403
95	ja- men	1871	1401	två	1771	1400
96	helt	1871	1401	se	1741	1376
97	alle	1851	1386	sätt	1694	1339
98	sam- men	1837	1376	andra	1687	1333
99	s-	1739	1302	tar	1661	1313
100	andet	1737	1301	från	1653	1306

3.2 Likheter

Mest värd att notera är den stora likheten mellan språken. De hundra mest frekventa orden i båda språken har alla direkt historisk släktskap med ord i det andra språket. Exakt tre av orden har samma frekvensrang på danska och svenska, 19% har samma rang inom ett intervall på fem och bland de 100 mest frekventa på både danska och svenska är 67% av orden gemensamma (dvs historiskt och etymologiskt besläktade och ofta med samma betydelse). Eftersom de mest frekventa orden är synkategorematiska (funktionsord), uppfattar vi detta som ett direkt stöd för att de två språken strukturellt sett är mycket lika varandra. Trots att orden är historiskt besläktade och att frekvensrangordningen är lika kan det dock finnas mindre skillnader. Ett exempel är ordet *det* (i talad svenska normalt *de*). Detta är det vanligaste ordet i såväl danskt som svenskt talspråk, men relativt sett är det vanligare i svenska än i danska. Detta beror förmodligen på att polysemin eller flerfunktionaliteten hos det svenska *det* är större jämfört med det danska *det*. En jämförelse med engelska illustrerar skillnaderna: det

svenska *det* kan bl a översättas med *it*, *that* eller *there*, medan det danska *det* kan översättas med *it* och *that* men inte *there*, vilket istället motsvarar det danska *der*.



3.2 Skillnader

Vi ska nu titta på några skillnader mellan danskt och svenskt talspråk så som de förekommer bland de 100 vanligaste orden. I tabell 2 nedan ger vi en översikt över ord som är vanligare i danska än i svenska, och i tabell 3 presenteras ord som är vanligare i svenska. Orden har rangordnats efter frekvens och skillnader mellan svenska och danska beräknas efter korpusandel mätt i promille.

4 Ord bland de 100 mest frekventa som är vanligare på danska än på svenska.

Tabell 2 och tabell 8 nedan visar alla ord bland de 100 vanligaste som har minst 10% övervikt i den ena korpusen gentemot den andra korpusen. För att räkna ut *övervikt* dividerar vi ppm (*parts per million*) för det ord som har störst frekvens med ppm för det ord som har minst frekvens. Så t.ex. har svenska *det* i GSLC 63074 ppm, medan danska *det* har 55530 ppm. Övervikten för svenska *det* blir då $63074/55530 = 113\%$. Om vi ser det från dansk synpunkt, blir resultatet $55530/63074 = 88\%$, dvs att förekomsten för danska *det* utgör 88% av förekomsterna för svenska *det*. (Siffrorna är avrundade.)

Tabell 2. Ord som är vanligare på danska än på svenska

Rang i da.	Ord	Antal	Sv. övers.	Da. överv. %
76	ham	2521	honom	547
98	sammen	1837	ihop	422
79	blev	2300	blev	414
66	min	3147	min	412
20	altså	14239	alltså	379
50	kunne	4553	kunna	373
83	gik	2115	gick	336
54	været	3851	varit	325
73	kom	2762	kom	297
40	hun	6609	hon	294
33	noget	9648	något	267
9	var	23458	var	265
65	nok	3197	nog	249
23	han	13314	han	243
92	fik	1961	fick	230
78	mange	2457	många	215
30	også	10390	också	209
86	over	2045	över	196
60	fra	3346	från	191
42	havde	6220	hade	190
49	godt	4622	bra	188
100	andet	1737	annat	173

96	helt	1871	helt	164
46	ved	5591	vet	160
7	der	32300	som	157
44	meget	5726	mycket	155
25	nej	13033	nä	153
68	mig	2927	mej	152
35	af	7511	av	147
13	mm	18187	m	140
2	ja	47126	ja	133
57	lige	3614	just	131
10	i	23341	i	128
6	så	36188	så	128
61	ud	3345	ut	128
4	jeg	39362	jag	127
84	sagde	2059	sade	125
8	ikke	24855	inte	122
28	til	10628	till	122
52	nogen	4154	någon	120
51	hvor	4359	hur	119
64	vil	3227	vill	117
75	op	2573	upp	114
14	de	17564	dom	112
71	I	2822	ni	111

4. 1 Danska ord som är vanligare än motsvarande svenska ord

Nedan diskuterar vi nu 12 danska ord bland de 100 mest frekventa som har mer än 200% övervikt över motsvarande svenska ord. I avsnitt 5 diskuterar vi sedan 12 svenska ord som har mer än 200% övervikt över motsvarande danska ord.

Orden diskteras i ordningsföljd efter hur mycket vanligare de är i danska respektive svenska. Undantag görs när en hel grupp av ord diskteras. Se exempel 1 nedan där

samtliga 3:e persons pronomen förts ihop med ham som är det ord bland de hundra mest frekventa som relativt sett har störst övervikt på danska jämfört med svenska.

1. Personliga pronomen - Tredje person

Danska *han* (övervikt 243 % över svenska *han* i GSLC)

Danska *hun* (övervikt 294 % över svenska *hon* i GSLC),

Danska *ham* (övervikt 547 % över svenska *honom* i GSLC)

Tabell 1 och den mer fullständiga frekvenslistan visar att maskulina pronomen är vanligare än feminina. Tabell 2 visar i tillägg att BySoc innehåller betydligt fler 3:e persons pronomen än GSLC. Några möjliga förklaringar till denna skillnad är dels att den svenska korpusen ibland innehåller formen *han* istället för *honom* i objektsposition, dels att man i danska kan använda objektsformen *ham* t ex i svar på *vem*-frågor, där svenska skulle använda *han*.

A: *Vem kom?* B: *Ham*

2. Sammen

Danska *sammen*: (övervikt 422 % över svenska *ihop* i GSLC)

Danskans *sammen* kan översättas på åtminstone tre sätt till svenska. Siffrorna inom parentes anger frekvens.

Danska		Svenska
<i>sammen</i> (1837)	—	<i>ihop</i> (673)
	—	<i>samma</i> (683)
<i>tillsammen</i> (4)	—	<i>tillsammans</i> (96)

Den främsta anledningen till den danska övervikten är därför troligen att ett enda ord *sammen* (de 4 förekomsterna för *tillsammen* får betraktas som marginella) svarar mot tre svenska ord. Danska har inget ord som historiskt motsvarar *ihop*. Om man slår ihop frekvenserna för de tre svenska orden (1452) ser man att *sammen* fortfarande har en viss övervikt (1837).

En av anledningarna till detta är att *sammen* tycks förekomma som konventionaliserad kollokation med vissa verb (i vårt urval 47 stycken) oftare än t.ex. svenska *samman* (i samma urval 18). Se exemplen i tabell 3 nedan. Troligen får vi liknande resultat även då vi tar hänsyn till kollokationer med *ihop* och *tillsammans* i de båda korpusarna.

Tabell 3. Verb som förekommer tillsammans med *sammen* och *samman*

Spalt 1	Spalt 2	Spalt 3	Spalt 4
Danska verb	Danska verb	Danska verb	Svenska verb
arbejde	holde	spille	braka
blande	hænge	spise	fläta
bliv e	høre	stuve	foga
bo	kæde	svejse	gå
brage	klappe	tage	hänga
bringe	klaske	tale	höra
cykle	klumpe sig	træffe	knyta
fælde	knække	være	komma
finde	knyta	vokse op	koppla
flytte	komme		möta
fløjte	krumme sig		prata
foga	køre		räkna
fungere	lave		sluta sig
gro	lege		ställa
gå	lære		sätta
gøre	læse		träffa
ha	opholde sig		väga
harmonere	spare		väva

I BySoc talar man i alla intervjuer om informanternas sociala förhållanden. Det är då mycket naturligt att tala om vad man gör *tillsammans* (*samman, ihop*). I GSLC är både aktiviteter och samtalsämnen mycket mer varierande. Här kan således även skillnader i samtalsämne spela en viss roll.

3. Blev

Danska *blev* (övervikt 414 % över svenska *blev* i GSLC)

Blev är både på danska och svenska aktiv preteritumform av danska *blive* och svenska *bli / bliva*. Verbet används både som vanligt verb med ungefär samma betydelse i bägge språken och som hjälpverb.

Den viktigaste anledningen till den danska formen *blevs* övervikt är att ungefär 43% av alla danska *blev* används i passiv-konstruktioner (i ett slumpmässigt urval på 110 förekomster av *blev*). Motsvarande siffra för svenska är ungefär 14%. I övriga fall används på svenska s-passiv.

En annan anledning till den danska övervikten för *blev* är att danska *blev* har två svenska översättningar, *blev* och *förblev*

En tredje anledning är aktivitet. Man bör även här uppmärksamma den roll som samtalsämne spelar. BySoc innehåller intervjuer om händelser i det förflutna medan samtalsämnena i GSLC är mer varierade.

4. Min

Danska *min*: (övervikt 412 % över svenska *min*)

Danska *jeg*: (övervikt 127% över svenska *jag*)

Eftersom danska och svenska har samma böjningsmönster (se nedan), beror skillnaden mellan danska och svenska knappast skillnaden på olikheter i grammatiskt system.

Danska	Svenska
<i>min</i>	<i>min</i>
<i>mit</i>	<i>mitt</i>
<i>mine</i>	<i>mina</i>

I BySoc förekommer 77% av alla förekomster av *min* tillsammans med släkt- och vänskapsord, medan i GSLC motsvarande andel är 12,5% och för informella samtal 11,4%.

Den viktigaste anledningen till övervikten för danska *min* är därför troligen samtalsämne. Detta är sannolikt också förklaringen till övervikten för *jeg*.

5. Alltså

Danska *alltså*: (övervikt 379 % över svenska *alltså* i GSLC)

Det danska ordet *alltså* har rang 20 (frekvens 10666 ppm), medan det svenska ordet *alltså* har rang 54 (frekvens 2814 ppm). I både svenska och danska är ordet slutsatsmarkör, men har i danska utvidgat sin funktion till emfas och framkallande av reaktion:

- (1) de har det der hjemme ik/altså
 (2) det kan jeg altså også godt far
 (/ = paus)

På svenska har ordet mindre utvidgad funktion och ligger oftast närmare slutsats.

det finns alltså anledning

En annan trolig förklaring är skillnaden i samtalsämne och aktivitet. BySoc innehåller endast informella samtal medan GSLC är mera varierad. En anledning är därför troligen skillnaden i formalitet, *altså* är typiskt för informellt språk och detta finns det mer av i BySoc.

Om vi undersöker var i ett yttrande *altså* respektive *alltså* förekommer, finner vi att 1490 förekomster eller 10,5% av alla danska *altså* står yttrandeinitialt, medan 16 förekomster eller 0,4% av svenska *alltså* står i samma position. Detta tycks alltså vara en skillnad i hur ordet används i de två språken.

Tabell 4. Yttrande-position för *altså* och *alltså*

	Danska		Svenska	
	Frekv	%	Frekv	%
Yttrandeinitialt	1490	10,5%	16	0,4
Omedelbart före paus	1981	7,6%	117	3,3%
Omedelbart efter paus	2050	14,4%	68	1,9%

Detta intryck förstärks av att danska *altså* mycket oftare än svenska *alltså* förekommer som första ord efter paus. I båda språken är medial position den vanligaste, men denna position är relativt sett mycket vanligare i svenska än i danska.

6. Kunne

Danska *kunne* (övervikt 373 % över svenska kunna i GSLC)

Den främsta anledningen till den danska övervikten är att danska *kunne* motsvarar flera svenska former.

Danska

kunne

Svenska

kunna (infinitiv)

kunde (preteritum, konjunktiv)

kunnat (supinum)

En annan anledning till övervikten är att det finns speciella kollokationer på danska som inte finns på svenska, t.ex. danska *de kunne godt sammen* som motsvarar svenska *de kom överens*.

Eftersom enbart 2323 (51%) av de danska förekomsterna av *kunne* utgör infinitiv innebär detta att de resterande 49% torde motsvaras av svenska preteritum- och supinumformer. Frekvensen för svenska *kunna* är 3410, vilket tyder på att svenska gör större bruk av *kunna* i infinitiv än danska.

7. Gik vs. Gick

Danska *gik* (övervikt 336 % över svenska gick i GLSC)

Intressant nog gäller övervikten för danska *gik* över svenska *gick* inte danska *går* över svenska *går*. I det senare fallet är det svenska ordet relativt sett vanligare.

Den viktigaste anledningen till den danska övervikten för *gik* tycks alltså vara aktivitet och samtalsämne. BySoc innehåller samtal om personliga händelser i det förflutna, vilket leder till en övervikt för preteritum.

Samtalsämnets vikt reflekteras i följande tabell

Tabell 5. Tre olika kontexter för *gik/gick*

	Danska		Svenska	
	frekv	%	frekv	%
konstruktioner med personliga pronomen som subjekt (objekt) för <i>gik/gick</i>	1883	89%	244	39%
<i>gik/gick</i> i konstruktioner med ”skolord”	87	4,1%	7	1,1%
<i>gik/gick</i> i konstruktioner med <i>altså/alltså, bare/bara, også/också och så</i>	294	13,8%	27	4,7%

Tabellen visar att en personlig anknytning är mycket vanligare i BySoc än i den svenska korpuser. Den personliga bakgrundens roll visas också av den roll prat om skolan, *gik i skole, gik i samme klasse*, spelar i den danska korpuser jämfört med den svenska.

Slutligen visar tabellen något om de informella kontexter *gik/gick* ingår i tillsammans med adverb som *altså/alltså, bare/bara, også/också och så*.

8. Været

Danska *været*: (övervikt 325 % över svenska *varit* i GSLC)

Været är aktiv perfekt particip av *være*. I danska byggs perfektum och pluskvam-perfektum på hjälpverbet *have* eller *være* + (e)t-form av perf. part. T.ex. *vi har skrevet; han er løbet. Være* används med rörelseverb och verb som betecknar övergång från ett tillstånd till ett annat. I motsats till svenska utelämnas normalt inte hjälpverbet i bisatser (t.ex. *Jag påstår att det förekommit pest i Mars.*). Det är bara 99 gånger (2,5%) som *været* förekommer utan hjälpverb *have* i BySoc. I den svenska korpuser finns det 574 förekomster av *varit* och av *vart*. Av dessa förekommer 68 utan hjälpverb (11,8%), vilket visar skillnaden mellan svenska och danska på denna punkt.

Verbanvändningen är liknande på svenska och danska. När det i BySoc står *det har været ret stor afstand imellem jer* kan detta lätt översättas på svenska: *det har varit (ett) rätt (så) stort avstånd mellan er*. Om man pratar om förflutna händelser utan bestämda tidsangivelser är det naturligt att använda perfektum som i BySoc: *nu havde jeg været sergent i livgarden; han havde været skibspræst*.

En anledning till den danska övervikten för *været* är grammatisk. Danska använder ibland *været* för att bilda passiv. Se följande exempel ur korpusen.

Danska: mange år siden jeg har været dumpet
Svensk översättning: många år sedan kuggades jag

Den danska korpusen skiljer sig något från det som är brukligt i svenska. Nedan några exempel ur BySoc.

Danska: en dag hvor Nyboder har været (smykket med flag)
Danska: så ha+ har vi aldrig været generet af ham
Danska: jeg synes aldrig jeg har været kuet = undertryckt)
Danska: eller at det bare har været nævnt
(+ = self intereruptio)

En annan typ av användning av denna konstruktion är aktiv, som framgår av följande exempel ur BySoc:

Danska: nu har vi været kommet ind
Danska: hun (havde) altså hellere været blevet her
Danska: der har han jo faktisk været flyttet

En preliminär genomgång av materialet visar att i den danska korpusen har vi 134 fall av den aktiva typen av konstruktioner och i GSLC bara 11. Det enda fallet i GSLC informell som kanske representerar konstruktionen i fråga är följande

Svenska: han e(h) knuffade omkull mej ja(g) vart överrumplad

Om vi relaterar antalet förekomster av den aktiva konstruktionen till alla förekomster av *været* respektive *varit*, kommer vi fram till att i BySoc står de för 3,4% och i GSLC för 1,9. Det handlar alltså om en relativt marginell förekomst, som dock är intressant och bidrar till att det uppstår ett gap i användning av *været* och *varit* mellan korpusarna.

9. Kom

Danska *kom* (övertikt 297 % över svenska *kom* i GSLC)

Liksom för *gik* kan vi konstatera att preteritumformen *kom* är vanligare på danska än på svenska. Däremot är presensformen *kommer* vanligare på svenska än på danska.

Precis som för *gick* är med stor sannolikhet huvudanledningen till detta att BySoc till skillnad från GSLC innehåller samtal om personliga händelser i förfluten tid. Detta intryck förstärks av att 81% av alla *kom*:s förekomster i BySoc har personliga pronomen som subjekt eller objekt. Motsvarande siffror för svenska *kom* är 41% i GSLC och 50% i GSLC informell. Några typiska konstruktioner med *kom* som finns i BySoc men inte är lika vanliga i GSLC:

1. *kom til at V* (svensk motsvarighet: *kom att* m.fl. uttryck)
2. *kom an på* (svensk motsvarighet: *bero på*; *kom an på* finns men är inte så vanligt)
3. *kom af sted* (svensk motsvarighet: *kom iväg*)

Exempel:

1. og så kom alle børnene også til at tale sådan
2. nogen gange tre det kom an på om vi
3. hvor vi kom af sted i fjorten dage

10. Noget

Danska *noget* (övervikt 267% över svenska *något* i GSLC)

I siffrorna ovan har den svenska talspråksformen *nåt* sammanförts med *något*.

En av anledningarna till att danska *noget* är vanligare än svenska *något* (*nåt*) är att *noget* tycks ingå i fler fasta kollokations- och konstruktionstyper. Se tabellerna 6 och 7 nedan.

Tabell 6. Konstruktionstyper

Konstruktionstyp	Frekvens
det er ({altså/ikke}) noget	365
er der (ikke) noget	57
det var ({ikke/sådan/ikke sådan}) noget	291
var det (ikke (sådan)) noget	31
X {ikke} noget	1110
og (sådan) noget	1647
{eller } sådan noget	662
{af/med/ikke/till} sådan noget	97
X sådan noget	3347

Tabell 7. Något /nåt i vissa fraskontexter i GSLC

Kontextfras	Något	Nåt	Utbyggd konstruktion
de(t) e0/ä(r) nåt		77	
de(t) e0/ä(r) något	15		
om de(t) e0 nåt			14
eller nåt		192	
eller något	11		
eller nåt sånt			91
eller nåt sånt där			59
nåt sånt		226	
något sådant		7	
för nåt		51	
för något	6		
gör/göra något	70		
på nåt		619	
på något	165		
på nåt sätt			414
på något sätt			128
hitta på nåt			10
hitta på något			1
komma på nåt			17
komma på något			1
va(d) ä(r) de(t) för nåt	7		
SUMMA	274	1170	

Endast 22 olika fraskontexter är uppställda i tabell 6, men dessa räcker till för att täcka hälften av de fall där *noget* förekommer i BySoc-korpusen. De 6 vanligaste fraskontexterna står för 4553 förekomster (46%), dvs. nästan hälften av alla förekomster. Många av dem är möjliga ur svensk synpunkt. Detta gäller dock inte kombinationer med *sådan noget*, vilka står för 34,6% av alla förekomster av *noget*. Speciellt vanliga är kombinationerna *og sådant noget* och *eller sådan noget*. En annan vanlig kombination är *ikke noget*, som står för 11,5% av alla förekomster av *noget*.

Om vi behandlar de svenska uttrycken i tabell 7, ser vi att det inte finns någon direkt motsvarighet till *sådan noget* i den svenska korpusen. Den fras som kommer närmast är *nåt sånt (något sådant)* med annan ordföljd än på danska, men *sådan noget* är 148% (3347/226) vanligare på danska än svenska. *Sådan noget* får därför betraktas som mycket typisk för danskt talspråk. De fall som presenteras i tabellen utgör 48,8% av alla förekomster av *något* (N = 562) och 36,5% av alla *nåt* (N = 2572). Det är genomgående så att *nåt* är vanligare än *något* i alla kontexter. En förklaring till den danska korpusens övervikt verkar således vara att språkanvändarna i BySoc brukar ett litet antal färdiga frasformer (fraskontexter, moulds), medan det i svenskan verkar vara mer variation.

11. Var

Danska *var* (övervikt 265 % över svenska *var* i GSLC)

Den högre frekvensen för *var* (preteritum av *vara/være* i både danska och svenska, och på svenska även frågeord) kan förklaras med aktivitetskillnader i de danska och svenska inspelade materialen. Det danska materialet baseras till skillnad från det svenska på informella intervjuer där personliga redogörelser för tidigare händelser är viktigt. Detta tenderar att leda till att öka användningen av såväl användningen av preteritumformer av *være* som användningen av pronomenet *jeg*.

12. Nok

Danska *nok*: (övervikt 249% över svenska *nog* i GSLC)

Danskans *nok* motsvaras i första hand av svenskans *nog* och särskilt i kombinationen *vist nok* av *væl*. *Nok* är liksom svenskans *nog* i olika kontexter synonymt med *tillräckligt*

och *kanske*. Troligen beror skillnaden på att *nok* för det första används mer i alla dessa funktioner på danska än på svenska och för det andra på att svenska har högre frekvens för de obesläktade orden *tillräckligt* och *kanske*.

5 Ord bland de 100 mest frekventa som är vanligare på svenska än på danska

Vi fortsätter nu med att studera ord som är mycket vanligare på svenska än på danska. Se tabell 8.

Tabell 8. Ord som är vanligare på svenska än på danska

Rang i GSLC	Ord	Antal	Dansk övers.	Överv. mot BySoc %
94	utan	1775	uden	457
23	här	14128	her	415
51	sen	3780	siden	404
78	ta	2306	tage	350
85	menar	2086	mener	331
65	kanske	3054	måske	296
97	sätt	1694	måde	289
99	tar	1661	tager	288
47	väl	4089	vel	281
61	sej	3350	sig	177
5	att	30998	at	172
43	ha	4668	have	169
22	för	14389	for	166
49	tycker	3904	synes	164
57	kommer	3425	kommer	158
31	ska	8258	skal	154
11	vi	18670	vi	147
50	säga	3816	sige	146
70	går	2665	går	145
21	den	14505	den	144
68	gör	2721	gør	143
13	man	17779	man	143
12	på	18038	på	142
69	blir	2699	bliver	141
98	andra	1687	andre	133
79	alla	2269	alle	129
58	lite	3412	lidt	126
34	vad	7730	hvad	123
27	kan	11157	kan	117
96	se	1741	se	114
1	det	79782	det	113
16	du	16539	du	112
60	tror	3369	tror	110

17	en	15789	en	108
52	vara	3748	være	106
39	nu	5247	nu	105
10	har	18800	har	104
18	ju	15618	jo	104
2	är	37942	er	104
80	in	2193	ind	101
88	säger	1894	siger	101
37	ett	6079	et	99
28	med	10466	med	97
19	men	14934	men	93
90	mer	1883	mere	92
3	och	35856	og	91
95	två	1771	to	91
74	ni	2401	I	89
82	upp	2133	op	87
71	vill	2613	vil	85
56	hur	3459	hvor	83
9	inte	19152	ikke	81
32	till	8254	til	81
6	jag	29174	jeg	78
7	så	26773	så	78
15	i	17244	i	77
73	ut	2467	ud	77
72	just	2595	lige	75
4	ja	33394	ja	74
25	m	12267	mm	71
41	av	4813	af	67
93	mej	1819	mig	65
33	nä	8060	nej	65
55	mycket	3480	meget	64
62	vet	3301	ved	62
77	bra	2321	godt	52
100	från	1653	fra	52
64	hade	3093	havde	52
42	också	4706	også	47

40	han	5173	han	41
89	väldigt	1890	meget	34
84	hon	2123	hun	33
86	någon- ting	2082	noget	22
81	jo	2137	jo	14

1. Utan

Svenska *utan*: (457% övervikt över danska *uden*)

Det svenska ordet *utan* är mer än fyra gånger så vanligt som det danska *uden*, vilket kan bero på att det inte bara har betydelsen 'i avsaknad av', utan också fungerar som konjunktion, och då på samma sätt som tyska *sondern*. Denna sistnämnda funktion saknas hos det danska *uden*. På danska har *men* istället detta som en av sina funktioner.

Vi kan se detta tydligt om vi söker på kollokationen *utan det* som förekommer 347 gånger i GSLC, i nästan samtliga fall följt av ett verb, ofta men inte nödvändigtvis i presens.

Exempel:

- (3) alltså inte FEMHUNDRA som det står här utan det kostar halva det priset
- (4) inte så där väldigt strikt upplagt utan det va:r ganska känslomässigt

(VERSALER = kontrastiv betoning
:= ljudförlängning)

Ser vi så på de danska förekomsterna av *uden* och *uden det* så ser vi att det senare bara förekommer tre gånger i BySoc, och då inte alls som konjunktion, utan istället t e x så här:

- (5) man kan dårlig nok lukke døren inden / uden det runger hele vejen op
- (6) fordi man kan næsten ikke røre sig uden det kan høres derinde vel
- (7) ja det er måske oven i købet uden det er måske: oven i købet uden varme

2. Här

Svenska *här*: (415% övervikt över danska *her*)

Här är ca fyra gånger vanligare i svenska än i danska, och skillnaden förklaras med att de svenska uttrycken *det här*, *den här*, *dom här* och även *så här* inte alls är vanliga på danska (*så här* finns överhuvudtaget inte). *Sådan* som på danska kan användas som motsvarighet till det svenska *så här*, förekommer 13 104 gånger i den danska korpussen men bara 1 367 gånger i den svenska.

3. Sen

Svenska *sen*: (404% vanligare än danska *siden*)

Det svenska ordet *sen* (sedan) är betydligt mer frekvent än det motsvarande danska, *siden*, vilket förklaras med att danskan har flera andra ord som används på liknande sätt, t.ex. *senere* och *så*.

4. Ta

Svenska *ta*: (450% övervikt över danska *tage*)

Ta förekommer 2306 gånger i GSLC. Det borde inte vara så mycket vanligare än danska *tage*, eftersom BySoc kan uppvisa danska motsvarigheter till svenska uttryck som *ta bussen, vi kan ta en tisdag, ta emot* m fl. Emellertid är danska *tog* 56 % vanligare än svenska *tog* (558 mot 356, och då är 13 förekomster på danska med betydelsen 'tåg' borträknade).

Den troliga orsaken är att BySoc ev. innehåller proportionellt fler samtal där deltagarna berättar om förfluten tid, vilket kan ge upphov till denna effekt.

5. Menar

Svenska *menar*: (331% övervikt över danska *mener*)

Att *menar* är vanligare på svenska beror åtminstone delvis på att kollokationen *jag menar* oftare används för egenkommunikationsreglering än vad som är fallet på danska. Ser man enbart på detta, så finns *jag menar* 970 ggr i GSLC, mot 270 *jeg mener* i BySoc, och är därför 359 % vanligare på svenska, d v s en ännu större övervikt. Orsaken skulle kunna vara densamma som för *kanske* (se nedan).

6. Kanske

Svenska *kanske*: (296% övervikt över danska *måske*)

Svenskarna använder *kanske* nästan tre gånger så ofta som danskarna använder *måske*.

Troligen beror detta på att danskan vid sidan om *måske* använder andra ord, som *nok*, som vi ovan sett är mycket vanligare än svenska *nog*.

7. Sätt

Svenska sätt: (289% övervikt över danska *måde*)

En stor del av förklaringen ligger nog i att svenska *på något sätt* (och dess reducerade varianter *på nåt sätt* m fl) är oerhört mycket vanligare och förekommer 444 ggr, mot 15 för *på nogen måde*.

8. Väl

Svenska väl: (281% övervikt över danska *vel*)

På svenska kan väl användas som konsensus- eller konventionspåtryckande markör, vilket är ovanligare på danska.

(8) A: du har väl inte glömt att gå ut med soporna

Du har vel ikke glemt att gå ud med X är helt i sin ordning på danska, medan *vel i det kan han vel* markerar vaghet, inte konsensuspåtryckning.

9. Exempel

Svenska exempel: (269% övervikt över danska *eksempel*)

Övervikten blir ännu större om man också tillägger de reducerade svenska formerna *ti sempel* och *t sempel*, förutom det skriftspråksenliga *till exempel*.

Svenskan har också några fasta uttryck med ordet *exempel*, varav bl a följande kan återfinnas i GSLC:

(9) förhoppningsvis för att dom ska tjäna som avskräckande exempel

(10) jag skulle kunna nämna dussinvis med liknande exempel

(11) ja / det finns ju många exempel

(12) om man tar X då som exempel

(13) det är ett utmärkt exempel

Dessa finns troligen även på danska (*afskrækkende eksempel, lignende eksempel, mange eksempler, som eksempel* och *udmærket eksempel*). Dock tycks de inte förekomma i BySoc.

En ytterligare anledning till att exempel är vanligare på svenska är att dess pluralform är identisk med singularformen (*ett exempel – flera exempel*), medan danska har en skiljaktig pluralform (*et eksempel – flere eksempler*).

10. Sej

Svenska *sej*: (177% övervikt över danska *sig*)

Att *sig, sej* är vanligare på svenska än danska indikerar att några speciella reflexiva konstruktioner är vanligare i svenskan, eller att reflexiva konstruktioner i allmänhet är det.

11. Att

Svenska *att*: (172% övervikt över danska *at*)

Svenskan uttrycker ofta framtid med *kommer att*, vilket danskan inte gör. En enkel sökning i den svenska korpusen gav ca 400 träffar på *kommer att*, och då tillkommer formuleringar som *kommer det att, kommer hon att*, o.s.v. Eftersom *at/att* förekommer drygt 10 000 fler gånger den svenska korpusen än i den danska, förklarar detta dock bara en del av skillnaden. En möjlig förklaring är att ordet oftare reduceras bort som bisatsinledare i danska än i svenska, t.ex i *jeg tror han kommer*.

12. Ha

Svenska *ha*: (169% övervikt över danska *have*)

Det svenska ordet *ha* är både verb (*ha - har - hade -haft*) och återkopplingsord (*jaha*). I det senare fallet finns ingen motsvarighet på danska.

6. Slutord

I denna uppsats har vi velat peka på några av de möjligheter en frekvensbaserad jämförelse av talspråskorpora från olika språk kan ge.

Vår utgångspunkt har varit en jämförelse av de 100 mest frekventa orden. Vi har exemplifierat vår metod genom att studera historiskt besläktade ord i de två språken. Givet att dessa ord mycket ofta visar stor överensstämmelse i sin användning mellan danska och svenska, är det av intresse att undersöka varför vissa ord är signifikant mycket vanligare i det ena språket än i det andra.

Vi exemplifierar den typ av resultat undersökningen ger genom att presentera preliminära resultat för tolv danska ord som är signifikant vanligare än besläktade svenska ord och för tolv svenska ord som är signifikant vanligare än besläktade danska ord.

Vår undersökning visar att förklaringar till skillnaderna är av följande typer.

- (i) Skillnader i aktivitet och samtalsämnen. Exempel på skillnader, som kan förklaras på detta sätt är t.ex. att danska personliga och possessiva pronomen samt preteritumformer som *blev*, *kunne*, *gik*, *var*, *været*, *kom* och *fik*, alla är vanligare än sina svenska motsvarigheter. Motsatsen gäller i stort sett för presensformerna, där de svenska formerna är vanligare. Förklaringen är som vi sett att BySoc innehåller informella samtal om personliga händelser i förfluten tid. Detta är också en bidragande faktor till att ord som *sammen* (vad man gör tillsammans) och *altså* (informell stil) blir vanligare i BySoc.

Samma skillnad ligger troligen bakom att det svenska ordet *eksempel* är vanligare än sin danska motsvarighet. Ordet passar bättre för de samtalsämnen som finns i den svenska korpusen, men är kanske mindre lämpligt i informella samtal.

- (ii) Skillnader som beror på att ett vanligt ord i ett språk motsvaras av flera relativt vanliga ord i det andra språket. Detta gäller t.ex. det danska ordet *sammen*, som motsvaras av de svenska orden *sammen*, *ihop* och *tillsammans*, *kunne* (*kunna*, *kunde*, *kunnat*), *nok* (*nog*, *kanske*, *tillräckligt*).

Det gäller även de svenska orden *utan* som motsvaras av danska *uden* och *men* eller *sen* (*siden*, *senare*), *kanske* (*nok*, *måske*), *eksempel* (*eksempel*, *eksempler*, *ha* (*ja*, *have*)).

- (iii) Skillnader som beror på att det ena språket innehåller populära och användbara konstruktioner, fraser eller idiom, innehållande det aktuella ordet som saknas i det andra språket. Ett exempel här är användningen av *ik* som yttrandefinal eliciteringspartikel på danska. Denna användning finns överhuvudtaget inte för svenska *inte*. Däremot finns en liknande funktion hos *va* och *eller*, som saknas på danska.

Ett annat exempel är det svenska uttrycket *kommer att* för att tala om framtiden, som saknas på danska.

- (iv) Skillnader som beror på att några av ett ords funktioner är mycket ovanligare på det andra språket. Ett exempel är här danska *blev* som i passivfunktion oftast motsvaras av s-passiv på svenska. Det tycks också gälla ett ord som *altså*, som är populärare på danska än på svenska och som i jämförelse med svenska tycks ha en något utvidgad användning i danskan. Ett annat exempel är danska *noget*, där konstruktionen (*og/eller*) *sådan noget*, förutom att framstå som typiskt dansk, tycks vara vanligare än de svenska motsvarigheterna (*och/eller*) *nåt sånt*. Om vi ser på svenska ord som är vanligare, finns uttryck med *här* (*det här, dom här* och *så här*) som är ovanliga eller inte finns på danska. De motsvaras istället av ord som *sådan* och *denne, dette*. Ett annat exempel utgörs av verbet *menar* i frasen *jag menar* i egenkommunikationsreglerande funktion, t.ex. *jag tar bussen jag menar tåget*, som är en mycket ovanligare funktion för det danska verbet *mener*. Även ordet *sått* och frasen *på nåt sätt* är här exempel, vilkas danska motsvarigheter kanske inte i första hand är *på någon måde* utan snarare t.ex. kanske det ovan nämnda *eller sådan noget*. Det svenska ordet *väl*, som konsensusframkallare, är också vanligare i denna funktion än det motsvarande danska ordet *vel*. Möjligen hör också observationen att vissa reflexiva verb (eller kanske reflexiva verb i allmänhet) är vanligare på svenska hemma här.

Vår undersökning av 12 högfrekventa ord med övervikt på det ena språket har också lett till vissa observationer om grammatik. Vi har observerat att det är vanligare på danska än på svenska att utelämna *att* som bisatsinledande konjunktion, medan det på svenska är vanligare att utelämna *ha/vara* i perfektumkonstruktioner. Dansk passiv bildas mycket oftare än svensk med *bli, vara* och *få*, där svenskan istället använder s-passiv. Svenska tycks använda fler reflexiva verb. Svenska har dessutom ett perifrastiskt uttryck för framtid, *kommer att*, som inte tycks användas på danska.

En svaghet med den metod vi använt är att vi blir beroende av vad som händelsevis kommit med i vår korpus. Vi ser detta i hur ofta samtalsämne och aktivitet spelar roll för förklaringar av de skillnader vi observerar.

Den stora styrkan är att de resultat vi får fram kommer från autentiskt språkbruk, snarare än från mer eller mindre rimliga intuitioner (ofta normativa) om hur det borde vara.

2 Litteraturreferenser

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*. Forthcoming in van Kuppevelt, J. (ed.) *Current and New Directions in Discourse and Dialogue* (Proceedings from SIGDial workshop Aalborg Aug. 2002). Kluwer Academic Publishers.

Christensen, R.Z. 1999. *Dansk grammatik for svenskere*. Lund: Studentlitteratur.

Cramer, J. & Larsen, E. V. 2001. *Dansk som nabosprog - Dansk grammatik for svensktalende*. Aarhus: Aarhus Universitetsforlag.

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.

Molde, B. (2000) Nordstedts Dansk-Svenska Ordbok, Nordstedts Ordbok.

Munch-Petersen, V. P.; E. Harmann (2001) Svensk-Dansk Ordbog, Gyldendal.

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

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

GCLC

The corpus contains 25 activity types with totally 361 recordings and 1360919 tokens.

Activity type	Recordings	Tokens
Auction	1	14452
Bus Driver-Passenger	1	1363
Church	2	10274
Consultation	16	34865
Court	6	33869
Dinner	5	30738
Discussion	35	242294
Factory Conversation	5	29047
Formal Meeting	13	223262
Games & Play	1	6120
Hotel	9	18950
Informal Conversation	17	82051
Interview	57	392219
Lecture	2	14682
Market	4	12581
Party	1	4485
Phone	32	14700
Retelling Of Article	7	5331
Role Play	3	8136
Shop	54	52507
Task-Oriented Dialogue	26	15475
Therapy	2	13841
Trade Fair	16	14353
Travel Agency	39	39743
Unknown	7	45581