

HANS HUSTRU OG HENDES BRYSTER

Af Jann Scheuer (Københavns Universitet)

1. Den Danske Ordbog: projektet og dataindsamlingen

I denne artikel præsenteres og diskuteres resultater som er opstået i forbindelse med Den Danske Ordbogs dataindsamling. Den Danske Ordbog, herefter undertiden kaldet DDO, er naturligvis et projekt der har til formål at producere en ordbog. Den Danske Ordbog er finansieret af Carlsberg-fondet og Kulturministeriet, der har ofret ca. 30 millioner på foretagendet, 15 millioner hver. Ordbogen har til huse på Københavns Universitet Amager, og den beskæftiger 10 redaktører på fuld tid, plus hvad der er af studentermedhjælp, edb-konsulenter, administrativt personale, osv. Tidsrammen for ordbogens tilblivelse er meget snæver: projektet løber fra 1991 til 1999, og der skal produceres en ordbog på 6 bind af ODS-størrelse, med i alt omkring 100.000 opslag. Første bind skal komme primo marts 1998, sjette og sidste bind primo september 1999.

For den naive betragter er der to slags ordbøger her i verden: de normative og de deskriptive. Den sondring er basal. Den Danske Ordbog bekender sig til den deskriptive tro, og med det valg følger en masse problemer, både praktiske og teoretiske. Jeg skal kort beskrive nogle af de designmæssige beslutninger der er truffet på Den Danske Ordbog.

Man må jo naturligvis forholde sig, når man vil være deskriptiv, til hvilket sprog man vil beskrive. Hvad det angår blev det bl.a. besluttet at sproget skulle være nutidigt; det skulle være produceret i tiåret 1983-92. Det skulle desuden være både fagsprog og alment sprog, både talesprog og skriftsprog, sprog produceret af og til forskellige aldersgrupper, osv. Der blev altså sat visse minimumskrav til spredningen - i praksis blev spredningen enorm. Det blev også besluttet at ordbogen skulle indsamle et korpus af størrelsesorden 40 millioner ord. Det svarer til mere end 100.000 normalsider. På baggrund bl.a. af det tekst-korpus og med andre eksisterende ordbøger som model, skal Den Danske Ordbogs ordbog produceres.

Enhver kan regne ud at 8 år er meget kort tid til at fremstille en ordbog af

den art og størrelsesorden - 6 bind af ODS-størrelse - med de ressourcer der er til rådighed. Det er oplagt at DDO har brug for at skyde genvej hvor det kan lade sig gøre, og en af de genveje der byder sig i vore dage er edb. Indsamlingen af de 40 millioner ord foregik per edb. Der fandtes heldigvis i forvejen datamatiserede tekst-korpora som ordbogen kunne trække på. Således bliver alle folketingsforhandlinger skrevet ind på edb. På Aarhus Universitet er der lavet en tekst-database som indeholder skønlitteratur. Flere aviser har alle deres artikler på edb. Danmarks Radio har undertekster på edb. På Københavns Universitet fandtes sociolingvistiske interview på edb.

Det var lidt vanskeligere at få datamatisk adgang til fx private breve, lejlighedssange og den slags, og der blev i forbindelse med de genrer udført en hel del reelt indtastningsarbejde. Men også de folkelige og dagligdags genrer frembragt af den uprofessionelle tekstproducent blev samlet ind.

De resultater der præsenteres i denne artikel foredrag stammer fra det omtalte tekstkorpus.

2. Databasen

Resultatet af dataindsamlingen blev et korpus af den ønskede størrelsesorden, og det er suverænt det største af sin art i landet. Der er blevet fremstillet programmer til effektiv søgning i korpusset og til statistisk beregning af forekomster, korrelationer, osv. Programmerne er faktisk meget brugervenlige, og alt i alt er basen blevet en guldgrube som kan bruges til meget andet end at producere ordbøger med. Jeg kender ikke DDO's planer med hensyn til basens anvendelse efter ordbogsredaktionens afslutning, men databasen: det store tekstkorpus og de programmer der arbejder på den, er efter min mening et værktøj som sprogforskningen i al almindelighed kan have stor gavn af.

Databasen nåede at blive nogenlunde færdig til brug mens jeg endnu arbejdede på DDO og jeg brugte (det kan jeg sige nu, hvor jeg ikke længere er ansat) en del tid på at sidde og læse i den, simpelthen fordi de tekster den indeholdt var så interessante. For nogles vedkommende slog autenticiteten mig. For illustrationens skyld kan jeg kort nævne et par genrer: strikkeopskrifter fra Femina, farvandsudsigter, (som nævnt) folketingsdebatter, erotiske noveller fra

Cupido, lægebrevkasser, sexbrevkasser og fra de mere private sfærer: lejligheds-sange, trusselsbreve - der var fx breve fra Holger Danske til kvindelige flygtninge-venlige politikere der var præcis så modbydelige som den slags breve er. Men de mest rystende dokumenter stammede fra Nationalmuseets dagbogsprojekt. Der var breve fra gamle mennesker der helt tydeligt forsøgte at tilfredsstille deres behov for at kommunikere i alenlange beretninger om dødsfald og sygdom i familie og øvrige omgivelser. Der var også en nøjagtig beretning fra en jysk pige som var arbejdsløs, og hvis eneste menneskelige kontakt i løbet af den dag der var tale om, var med den læge der skulle behandle hendes dårlige ryg.

Det er svært at formidle det indtryk sådan en tekst giver, men oplevelsen af virkelighed var slående. Det er naturligvis et subjektivt indtryk, men indtrykket er, når man kigger i Den Danske Ordbogs database, at den beskriver sproget i det virkelig liv i en udstrækning man ikke er vant til, heller ikke (eller slet ikke) som sprogforsker.

Det skal for en god ordens skyld nævnes at alle teksterne i korpus naturligvis er anonymiserede.

3. Resultater: undersøgelsen af *hans*, *hendes*

Det kvalitative indtryk, læseoplevelsen i forbindelse den enkelte tekst i databasen, er én ting. Man kan også gå kvantitativt til korpus. De programmer der er skrevet til at få adgang til DDO's korpus med, kan udføre et antal kvantitative statistiske beregninger, som kan give et fantastisk overblik over sproget i korpus som et hele.

Der kan bl.a. meget hurtigt vises samtlige forekomster af et givet ord i korpus, der kan regnes på frekvensen af ord, og der kan også regnes på hvilke ord der typisk optræder efter hinanden i korpus. Og meget mere.

Skema 1 herunder viser et resultat af sidstnævnte undersøgelsestype. Det er blevet undersøgt hvilke ord der optræder efter de possessive pronomener *hans* og *hendes*. I venstre side står de ord der optræder efter *hans*, i højre de ord der optræder efter *hendes*. I mange tilfælde danner pronomenet og det følgende ord et hypotagme. I *hans*-kolonnen har vi således *hans sønner*, *hans brødre*, osv.

	HANS	Klasse	T-score	HENDES	Klasse	T-score
1	sønner	s	8.25	hår	s	-9.27
2	brødre	s	7.61	øjne	s	-8.86
3	værker	s	7.29	mor	s	-8.05
4	mænd	s	7.29	kinder	s	-7.55
5	opfattelse	s	6.81	krop	s	-7.36
6	politiske	adj	6.77	stemme	s	-7.21
7	død	s	6.55	forældre	s	-6.92
8	kolleger	s	6.53	far	s	-6.91
9	venner	s	6.13	ben	s	-6.61
10	billeder	s	6.06	ansigt	s	-5.96
11	eller	kon	6.05	hud	s	-5.47
12	parti	s	6.04	skød	s	-5.38
13	regering	s	5.99	veninder	s	-5.16
14	musik	s	5.9	lille	adj	-5.07
15	tilhængere	s	5.62	lår	s	-5.06
16	politik	s	5.46	trusser	s	-4.55
17	egen	adj	5.36	smil	s	-4.42
18	film	s	5.31	mund	s	-4.39
19	kammerat	s	5.28	veninde	s	-4.30
20	indsats	s	5.28	taske	s	-4.26
21	kontor	s	5.20	læge	s	-4.22
22	ideer	s	5.18	bluse	s	-4.14
23	trofasthed	s	5.17	børn	s	-4.12
24	firma	s	5.11	varme	adj	-4.01
25	eget	adj	5.07	søstre	s	-4.00
26	tid	s	4.97	dør	s	-3.91
27	værk	s	4.93	hånd	s	-3.90
28	produktion	s	4.91	ældste	adj	-3.78
29	afgang	s	4.78	barn	s	-3.76
30	land	s	4.75	hvide	adj	-3.43
31	medarbejderes		4.66	hals	s	-3.43
32	styre	s	4.57	latter	s	-3.42

Der er ikke tale om en blot og bar undersøgelse af frekvens. Et ord figurerer ikke i listen til venstre fordi det **hyppigt** optræder efter *hans*, men fordi det **signifikant** optræder **hyppigere** efter *hans* end efter *hendes*. Og omvendt.

Ordene er i hver kolonne ordnet efter faldende signifikans. At forbindelsen *hans sønner* optræder hyppigere end *hendes sønner* er mest signifikant i *hans*-kolonnen, mindst signifikant er at *hans styre* optræder hyppigere end *hendes styre*. Alle ord på listen er dog signifikant knyttet til enten *hans* eller *hendes*; listen angiver signifikante statistiske forskelligheder.

Listen er redigeret i forhold til de oprindelige uddata. De uredigerede uddata gengives bagerst i artiklen. Der var en del "støj", og den er blevet fjernet. Støj er i denne sammenhæng ord som var kommet på listen fordi de indgår i uproduktive forbindelser: fx i titler på blade, som i tilfældet *Hendes Verden* og titler af en anden art som i *Hendes Majestæt*. Ordene *verden* og *majestæt*, og andre ord, er fjernet fra listens hun-side. Støj er også ord som udelukkende kan optræde efter *hans* og slet ikke efter *hendes*, og omvendt. Fjernet af den årsag er blandt andet ordene i denne artikels titel: *hustru* og *bryster*, som ellers stod næstøverst i hver sin kolonne. Man kan ikke vellykket sige *hendes hustru* og *hans bryster* og derfor er det ikke interessant at de to ord er knyttet til de respektive pronomener. Det fortæller os intet om sprogbrugernes frie valg.

Hvis sprogbrugen, som den fremstår i databasen, var kønsneutral skulle der ikke være nogen ord tilbage efter sådan en udrensning. Hvis alle ord, der kan optræde efter både *hans* og *hendes* gjorde det uden signifikant forskel i hyppighed, ville der ikke være nogen signifikant forskel på brugen af *hans* og *hendes*. Men det er der så rigeligt, som man kan se. Og som man også kan se: ordene skriger på fortolkning!

Når man ser på ordene virker det oplagt at inddele dem i nogle betydningsmæssige kategorier. Og ydermere: ordene i sig selv synes at opfordre til nogle bestemte kategorier. Jeg tager de tungest vejende og mest iøjnefaldende kategorier først, dem der er flest ord i højst oppe på listen.

Kvinden som kød- og handelsvare: I *hendes*-kolonnen er der en masse ord - halvdelen eller flere - der karakteriserer kvinden fysisk. Jeg tænker her på forbindelser som

hendes: hår, øjne, kinder, krop, stemme, ben, ansigt, hud, skød, lår, smil, mund, hånd, hals, latter

Lader man blikket glide hen over listen, er det iøjnefaldende at det ikke er en hvilken som helst karakteristik. Det drejer sig i høj grad om kvindens fysiske handelsværdi - på kødmarkedet, som man siger. Kvinden som mandens begærsobjekt opgjort i kødkapital. Det er heller ikke en hvilken som helst beklædningsgenstand der figurerer højst blandt *hendes*-ordene: *trusser* hører par excellence til

i den seksuelle diskurs. Her er altså det billede sprogbrugen i perioden 1983-92 giver af kvinden. Dette er de sproglige forbindelser, de hypotagmer, der særligt kendetegner hende.

Manden som hegemonisk operator: Lader man blikket løbe ned over *hans*-kolonnen, er det helt oplagt at der er stor forskel. Mændene har intet kød. Men hvad har de så? De kategorier som springer i øjnene på mandesiden, har at gøre med det åndelige og offentlige virke, dets produkter og magtrelationer. I disse kategorier passer hovedparten af de ord der optræder i *hans*-kolonnen. Det åndelige virke og dets produkter udtrykkes af forbindelser som

hans: værker, opfattelse, billeder, politik, musik, film, indsats, kontor, ideer, firma, værk, produktion

og magtrelationer i dette virke udtrykkes af

hans: mænd, parti, regering, tilhængere, medarbejdere, styre

Det der portrætteres i kolonnen med *hans*-ord er kort sagt den hegemoniske magt. Produkterne af *hans* værk er kunstneriske såvel som samfundsregulerende, hvad enten det nu drejer sig om statsadministrationen, de private firmaer eller om kunst- og kulturlivet. Alle de meningsdannende instanser er dækket. Og det er så det billede sprogbrugen har af manden.

Familien: Det er oplagt at familien også er en kategori der har relevans. Hvad familien angår, kan man notere sig at mandens position defineres ved at han har sønner og brødre, dvs. jævnbyrdige og underordnede af hankøn. Den definition giver ham en placering i toppen af et mandsfikseret familiehierarki. Kvindens position i familien defineres derimod af relationer der går både op og ned. Vi finder forbindelserne:

hendes: mor, forældre, far, børn, barn

Hendes placering er ikke i toppen, der er et forældreniveau over hende. Til forskel fra manden har hun både far, mor og forældre. Og til forskel fra manden har hun børn, ikke bare sønner.

Andre relationer: De øvrige relationer, relationerne ud til de mennesker man møder uden for familien beskrives for *hans* vedkommende med forbindelserne:

hans: mænd, kolleger, venner, tilhængere, kammerat, medarbejdere
og for *hendes* vedkommende:

hendes: veninder, veninde, søstre, læge

Mandens relationer er tilsyneladende mere professionelle. Det lader til at hun kun omgås sit eget køn, når man ser bort fra lægen.

Ikke-substantiver: Det er altovervejende substantiver der optræder på listen. De forbindelser som ikke indeholder substantiver er i *hans*-kolonnen

hans: politiske, eller, egen, eget

Blandt *hans*-ordene optræder konjunktionen *eller*. Det skyldes at man oftest nævner ham først og skriver *hans* eller *hendes*, og ikke omvendt. Derudover finder vi ordene *eget* og *egen*, der har med det at eje at gøre. Semantisk har *egen* og *eget* den egenskab at fokusere opmærksomheden på den ejende person og det suveræne ejendomsforhold. *Hans eget hus* i forhold til *hans hus* betyder i højere grad at det er *ham* der ejer, og at han gør det alene, suverænt. Manden som besidder er fremme her.

Forbindelser uden substantiv i *hendes*-kolonnen er:

hendes: lille, varme, ældste, hvide

Blandt hun-ordene har vi altså adjektiverne *lille*, *varme* og *hvide*. Formentlig optræder **substantiverne** i listen med samme fordeling efter *hendes* alene og efter *hendes lille*, *hendes varme* og *hendes hvide*. Det er derfor sandsynligt at det vi ser er beskrivere i hypotagmer der beskriver kvindens fysik, hypotagmer som *hendes lille hånd*, *hendes varme smil*, osv. Og sexdyret får så i øvrigt tænderne trukket ud: det er en rigtig lille hvid prinsesse der er til salg.

Død og afgang: Manden er for øvrigt den der holder op - enten med at leve eller også med at arbejde. På *hans*-listen figurerer *død* og *afgang*. Det har sat sine spor i DDO's database at mænd er døde eller gået af. Det sætter spor i historien når mænd gør den slags; ikke når kvinder gør det. På listen figurerer også *hans tid*, som må høre hjemme i samme semantiske univers, i beskrivelsen af den historisk betydende mand.

Sære ord: Ordet *ældste* optræder som nævnt også på *hendes*-siden, og det er lidt underligt. Men i sig selv betyder *hendes ældste* jo noget, så måske er der tale om børn. Mere sært er *hendes dør*. Forklaringsforslag modtages gerne...

3.1. Dekonstruktionen af kvinden

En af mine studerende sagde, konfronteret med listen, noget i stil med: *det er den sædvanlige dekonstruktion af kvinden*. Jeg er ikke helt sikker på hvad hun mente, men det forekommer mig at der er noget om snakken. Kvinden beskrives i et ben, et smil, nogle lår. Det er fragmenter. Dels fragmenter af hendes krop. Men menneskets fysik er i sig selv et fragment af den menneskelige totalitet. Og det merkantile mandlige begærs øje, som er i funktion i denne sammenhæng, er kun én synsvinkel på fysikken. Så også hvad perspektivet angår er der tale om et fragment og en dekonstruktion.

Dekonstrueres manden i samme grad? For mig at se beskriver *hans*-ordene i højere grad manden i sin totalitet, manden som helstøbt operator i den samfundsmæssige diskurs, den diskurs hvor åndsformerne dannes og alle beslutninger træffes. Mandens fysik beskrives som nævnt ikke. Men er han derfor dekonstrueret? Ikke i samme grad vil jeg sige, og ikke med samme konsekvens, hvad magt angår. Kvinden nægtes sin ånd, manden sin krop. Kvindens tab er det største, al den stund det er ånden, der regerer denne verden. Mandens tab efterlader en stadig helstøbt og stadig fungerende magtinstans.

3.2. Udsagnets tyngde

Hvor tungt vejer så det resultat der præsenteres? Som nævnt: korpus er 40 millioner ord, mere end 100.000 normalsiders tekst, og der er investeret stor opmærksomhed i at indsamle et korpus der indeholder tekster med meget stor spredning. Den Danske Ordbog kunne sagtens have indsamlet en langt større database hvis der ikke havde været stillet de krav til repræsentativitet som faktisk blev stillet. Man kunne have skaffet 100 millioner ord fra avisartikler så nemt som ingenting, men det ville ikke have været lige så interessant.

Jeg har ikke tal på hvor mange forekomster af *hans* og *hendes* der egentlig er i korpus, men de to ord hører næppe ligefrem til de lavfrekvente. Det er en sag der kan diskuteres, men for mig vejer dette resultat meget tungt; jeg tror på DDO-databasens udsagn.

med undersøgelser som den der er blevet præsenteret. Man kan fx give sig til at selekttere i genrer, og undersøge om tendensen er den samme i offentlige og private genrer. Og om tendensen er den samme i tale- og skriftsprog. Man kan selekttere i aldersgrupper og man kan også selekttere i køn. Man kan undersøge om kvinder beskriver deres eget køn på samme måde som på listen. Undersøgelser på udvalgte genrer eller udvalgte afsendere m.m. kan uden videre udføres med de programmer der findes på DDO i dag.

Man kan naturligvis også undersøge andre grammatiske sammenhænge: fx hvilke verber der optræder med *han* og *hun* som subjekt. På den måde kunne man undersøge om de aktionsarter der knytter sig til de to køn fordeler sig efter samme skema som genstandene gør, om ideologien kommer til udtryk på samme måde. I listen med *hans*- og *hendes*-ordene er det interessant at mange af ordene i *hans*-kolonnen netop er produkter af aktioner eller arbejde: fx *hans værk*, *hans indsats* og *hans produktion*. Ingen af ordene i *hendes*-kolonnen er produktet af aktioner - hvis vi tillader os at se bort fra børnene. *Hendes lån*, *hendes stemme* og *hendes læge* er ikke produkter hun selv har skabt gennem sine handlinger, det er ikke genstande hun kan forme eller har magt og kontrol over. Det er attributter der bare klæber til hendes person. Så i betragtning af hvad listen med *hans* og *hendes*-ordene viser os mht. til handlinger og deres produkter, må det forventes at en undersøgelse af verber der har *han* og *hun* som subjekt, også vil give kønsspecifikke resultater.

4. Konklusioner, konsekvenser

Hvordan skal Den Danske Ordbog give et eksempel på den typiske brug af ordet *hans*? Hvis man som eksempel vil vælge det mest repræsentative, kan man tage det hyppigst forekommende hypotagme med *hans* som bestemmerled. Det hypotagme kender jeg ikke, og det fremgår ikke af listen her. Det eksempel er imidlertid heller ikke særlig interessant, har Saussure lært os. Det der først og fremmest karakteriserer *hans* er modstillingen til andre ord i paradigmet, de øvrige possessive pronomener, og da især til det pronomen som kun adskiller sig ved genus. Og det er præcis hvad listen viser os. Det er simpelthen en god beskrivelse af sproget i 1983-92 at give eksemplerne *hans værker*, *hans parti*, og med samme

genus. Og det er præcis hvad listen viser os. Det er simpelthen en god beskrivelse af sproget i 1983-92 at give eksemplerne *hans værker*, *hans parti*, og med samme argumentation *hendes skød*, *hendes krop*. Og det er slet ikke en beskrivelse af sproget hvis man giver eksempler som *hans skød* og *hendes parti*. Ikke hvis repræsentativitet er målet, og jeg har svært ved at se hvordan man kan lave deskriptive ordbøger som ikke sigter efter netop det mål.

Ordlisterne beskriver en patriarkalsk ideologi. Der gives dette billede af kønnene: manden er et åndsvæsen, i familien en høvding, i samfundet i øvrigt besidderen af den hegemoniske magt. Kvinden er en kødvare. Og det billede skal reproducere i en deskriptiv ordbog.

Hvis man er utilfreds og uenig med den ideologi, kan man fristes til at slå til lyd for at ordbøger skal være normative, at man må undlade at reproducere og derigennem forstærke de chauvinistiske strømninger listen beskriver. Men den tankegang er forkert. Jeg synes egentlig ikke der er noget galt i at forsøge at opdrage folk med normative ordbøger, især ikke når det undertiden er det folk gerne vil have. Men der er noget galt i at undlade at lave deskriptive ordbøger. Den deskriptive ordbog er et vigtigt historisk dokument, og DDO portrætterer samfundet og samfundsideologien som den ser ud i tiåret 1983-92.

Det står værre til med den sproglige fremstilling af kønnene end den naive betragter kunne tro, og det er uhyre vigtigt at få det faktum nedfældet og synliggjort, som det bliver i en deskriptiv ordbog. Gør man ikke det, risikerer man at ideologierne bliver overset, eller fortrængt, fordi de ikke passer ind i samfundets foretrukne selvbillede.

Litteratur

- Lis Jacobsen 1912: *Kvinde og mand. En sprogstudie fra dansk middelalder*. Gyl dendalske boghandel, Nordisk forlag. København og Kristiania. (Se s. 227-30.)
- Karen Margrethe Pedersen 1994: Om Ømålsordbogen, i Jacobsen, Henrik Galberg: *Selskab for Nordisk Filologi København. Årsberetning 1992-1993*. København, s. 41-57. (Se s. 46ff.)

Jann Scheuer: Hans hustru og hendes bryster (uredigerede uddata)

Den Danske Ordbog

Undersøgelse af forbindelser med *hans* og *hendes*

	HANS	Klasse	T-score	HENDES	Klasse	T-score
1	kone	s	30.16	mand	s	-28.86
2	hustru	s	12.72	bryster	s	-10.08
3	folk	s	11.72	verden	s	-9.57
4	disciple	s	8.87	hår	s	-9.27
5	sønner	s	8.25	øjne	s	-8.86
6	brødre	s	7.61	mor	s	-8.05
7	værker	s	7.29	kinder	s	-7.55
8	mænd	s	7.29	krop	s	-7.36
9	opfattelse	s	6.81	stemme	s	-7.21
10	politiske	adj	6.77	kjole	s	-7.00
11	pik	s	6.77	forældre	s	-6.92
12	død	s	6.55	far	s	-6.91
13	kolleger	s	6.53	majestæt	s	-6.67
14	venner	s	6.13	ben	s	-6.61
15	billeder	s	6.06	kusse	s	-6.06
16	eller	kon	6.05	ansigt	s	-5.96
17	parti	s	6.04	mands	s	-5.89
18	regering	s	5.99	elsker	s	-5.72
19	musik	s	5.99	hud	s	-5.47
20	tilhængere	s	5.62	skød	s	-5.38
21	politik	s	5.46	veninder	s	-5.16
22	egen	adj	5.36	lille	adj	-5.07
23	film	s	5.31	lår	s	-5.06
24	kammerat	s	5.28	trusser	s	-4.55
25	indsats	s	5.28	ægtemand	s	-4.44
26	kontor	s	5.20	smil	s	-4.42
27	ideer	s	5.18	mund	s	-4.39
28	trofasthed	s	5.17	veninde	s	-4.30
29	firma	s	5.11	taske	s	-4.26
30	eget	adj	5.07	læge	s	-4.22
31	rige	s	4.99	bluse	s	-4.14
32	tid	s	4.97	børn	s	-4.12
33	værk	s	4.93	varme	adj	-4.01
34	produktion	s	4.91	søstre	s	-4.00
35	afgang	s	4.78	dør	s	-3.91
36	land	s	4.75	hånd	s	-3.90
37	medarbejderes		4.66	ældste	adj	-3.78
38	ord	s	4.58	barn	s	-3.76
39	styre	s	4.57	hvide	adj	-3.43
40	hold	s	4.50	hals	s	-3.43
41	efterfølger	s	4.48	latter	s	-3.42

Mette Kunøe og Erik Vive Larsen (udg.):
5. Møde om Udforskningen af Dansk Sprog,
Århus 1995

UDSKRIFTER AF DANSKE SAMTALER

Præsentation af nyt korpus med udskrifter af danske samtaler

Af Jakob Steensig (Aarhus Universitet)

1. Introduktion

På Institut for Lingvistik i Aarhus er vi nogle studerende og ansatte som de sidste to-tre år har arbejdet med analyser af danske samtaler. Vi er nu ved at oprette et egentligt data-korpus, som skal indeholde bånd og udskrifter af de samtaler vi efterhånden har fået optaget. Korpus kommer til at hedde *LAUS* (Lingvistik Aarhus Universitet, Samtalekorpus), og det er meningen at det skal stå til rådighed for forskere og studerende i hele landet.

Formålet er at udvide datagrundlaget for empiriske studier af dansk talesprog. Vi forestiller os at korpus kan bruges fx af forskere som har udviklet hypoteser på basis af egne data, og som gerne vil afprøve disse på yderligere data, af studerende som mangler samtaledata til opgaver eller projekter, af lærere som gerne vil bruge autentiske eksempler, m.fl.

2. Praktiske oplysninger om *LAUS*

Korpus kommer til at indeholde et indeks og en række bånd med tilhørende udskrifter. Til hver optagelse vil der foreligge et *datablad* med en kort beskrivelse af deltagere, optagelsesomstændigheder m.m. Generelt vil man kunne få adgang til både bånd og udskrifter mod at garantere anonym behandling af data og mod et løfte om at stille evt. nye udskrifter til rådighed for *LAUS*. I enkelte tilfælde kan der være knyttet specielle anonymiseringskrav til bestemte optagelser; det vil i så fald fremgå af databladene.

Vi har søgt penge til medhjælpstimer til at udarbejde datablade og evt. et søgeprogram, give databladene et fælles format, samle udskrifter og evt. tilpasse dem vores minimale udskriftsnormer (se nedenfor), og til at komplettere de udskrifter der allerede er lavet. Hvis vi får pengene vil korpus blive tilgængeligt i løbet af foråret 1995. Nærmere oplysninger kan fås ved henvendelse til undertegnede på Institut for Lingvistik, Aarhus Universitet, 8000 Århus C, tlf.