

JURAPROJEKTET

ERFARINGER OG RESULTATER

Inge Gorm Hansen, Institut for Engelsk, HHK.

Juraprojektet blev iværksat ved HHK i efteråret 1984 med pilotprojektet: Database til terminologisk information og generering af ordbøger. På symposiet i 1985 i Helsinki gav Steffen Leo Hansen en statusrapport over projektet, som er offentliggjort i symposiets proceedings ¹⁾. Projektgruppen ^{*}) har i begyndelsen af 1987 afsluttet projektet og udarbejdet en større rapport (ca. 500 sider incl. bilag) til Egmontfonden, som har støttet projektet økonomisk. En oversigtsartikel om pilotprojektet er udkommet i HHKs publikation ARK no. 36 ²⁾. I det følgende vil projektets enkelte faser således ikke blive gennemgået i detaljer, idet der henvises til ovennævnte publikationer.

Målet med dette indlæg har i højere grad været at fremlægge nogle konkrete arbejdsresultater og perspektivere dem i forhold til fremtidig terminologisk og leksikografisk forskning.

Oversigt over pilotprojektet

Formål: at udvikle og afprøve en korpusbaseret arbejdsmetode, der forener terminologiske og leksikografiske principper, og som skal resultere i

- en flersproget terminologi-database
- sprogretnings- og brugerbestemte tosprogs-ordbøger
- den fagsproglige komponent af en større etsprogsordbog for danske brugere

^{*}) Steffen Leo Hansen, Ole Helmersen, Bodil Nistrup Madsen, Hanne Puggård, Joan Haff Tournay, Charlotte Werther, Anne Zoëga og Inge Gorm Hansen

- Sprog: dansk, engelsk, spansk
- Emne: civilproces - borgerlige domssager i 1. instans
- Baggrund: - manglen på egnede, fagspecifikke ordbøger for danske brugere
- ønsket om at kombinere anvendelsen af EDB med nyere forskningsresultater inden for terminologi og leksikografi ved fremstillingen af ordbøger
- Målgruppe: danske brugere
- oversættere og tolke
- jurister
- advokatsekretærer
- undervisere i juridisk fagsprog
- jurastuderende
- erhvervsproglige studerende
- Arbejdsfaser: - oprettelse og anvendelse af tekstkorpus
- inventarisering af fagsproglige termer
- oprettelse af en flersproget terminologi-database
- generering af ordbøger

Emnet for pilotprojektet er som omtalt fagsprogligt. Jura er valgt som fagområde, da der inden for dette oversættes og tolkes meget, og da der generelt savnes egnede juridiske ordbøger. Samtidig udgør emnet et område, hvor man må forvente at blive præsenteret for en lang række metodiske og indholdsmæssige problemer, når det skal bearbejdes terminologisk og leksikografisk. Især ækvivalensproblematikken er et meget vanskeligt område indenfor juridisk sprog, hvor man arbejder med vidt forskellige referencerammer i form af de forskellige juridiske systemer.

Arbejdsfasen, oprettelse af tekstkorpus, har været projektets mest omfattende og har bestået af

1. etablering af tekstkorpus
2. fastlæggelse af korpusstruktur
3. oprettelse af et maskinlæsbart korpus
4. udvikling af programmel til automatisk korpusanalyse

Pilotprojektets to sidste arbejdsfaser, oprettelse af en flersproget terminologidatabase og generering af ordbøger, har især omhandlet anvendelse af DANTERM-recorden i forbindelse med oprettelse af arbejdsdatabasen JURTERM.

Anvendelse af det maskinlæsbare korpus

I det følgende vil anvendelsen af det maskinlæsbare korpus, blive uddybet med eksempler på søgning efter data i korpus til indlæsning i terminologidatabasen. Der vises eksempler på, hvorledes data genereres fra terminologidatabasen til ordbogsprofiler beregnet for forskellige brugergrupper.

Pilotprojektets korpus består af et "manuelt" og et maskinlæsbart korpus, hvor det sidste er en delmængde af det første. Det manuelle korpus er emnedækkende for civilproces og udarbejdet i form af en bibliografi i samarbejde med juridisk laboratorium ved Københavns Universitet. Det har ikke været muligt indenfor projektets rammer at etablere det totale korpus i maskinlæsbar form. Teksterne i det maskinlæsbare korpus er således primært valgt ud fra de specielle krav, man må stille til terminologisk dokumentationsmateriale og på baggrund af teksttypologiske overvejelser og spørgsmålet om repræsentativitet.

Det samlede korpus - dansk, engelsk, spansk - består af ca. 1 million løbende ord og indeholder følgende teksttyper: love, lovkommentarer, lærebogstekster, retsafgørelser og partsskrifter (formularer og autentiske dokumenter).

Hvis man sammenligner jurkorporus med andre korpora, kan det forekomme meget omfattende, når man tager i betragtning, at det dækker et ret begrænset emneområde indenfor et enkelt af juraens fagområder. Dette skal ses på baggrund af førmtalte specielle krav til korpus, som projektgruppen har formuleret. Det har været et væsentligt krav til korpus, at det såvel kvalitativt som kvantitativt skulle være dækkende for alle inden for det udvalgte emneområde forekommende teksttyper. Korpus skal ikke blot kunne anvendes til at registrere sprogbrugen i de forskellige teksttyper, men også fungere som en videns/dokumentationsbase, hvor man kan hente faktuelle oplysninger og søge efter definitioner af termer. Målet har således været, at korpus ved sin sammensætning kunne tilgodese ønsket om at registrere både lingvistiske og ekstralingvistiske faktorer i forbindelse med de enkelte termer.

Dette skal ses i sammenhæng med den forskel, der traditionelt er mellem leksikografiske og terminologiske arbejdsmetoder. Leksikografen vil normalt tage det enkelte ord som sit udgangspunkt og beskrive hele ordets betydningsfelt, hvor terminologen isolerer én betydning i det samlede betydningsfelt som led i en systematisk beskrivelse af et begreb, dets relationer til over- og underbegreber etc. Det foreliggende projekt er en blanding af de to nævnte arbejdsmetoder, hvorfor der stilles specielle krav til sammensætningen af projektets tekstkorpus.

I den konkrete anvendelse af det maskinlæsbare korpus har der været gennemført en række forsøg med søgning efter data med ovennævnte to mål for øje: 1) registrering af sprogbrug og 2) søgning med henblik på at finde definatorisk materiale til indholdsbeskrivelse af de enkelte termer.

I forbindelse med 2) har der været gennemført forskellige forsøg med strategier i on-line søgning. Teksterne er klassificeret med nøgleord og ved hjælp af nøgleordskoderne har man forsøgt at fin-

de frem til tekststeder med faktuelle oplysninger, som kunne bidrage til definition af termerne. Andre forsøg er gået ud på at søge på forskellige "streng". Man har f.eks. kunnet søge på "stævning + forstås" for at identificere et tekstafsnit indeholdende ordene "ved stævning forstås", eller f.eks. "stævning ...er". Det har givet resultater i visse tilfælde, men er et område, som bør undersøges nøjere, da de forsøg der har været gennemført endnu hviler på et spinkelt empirisk materiale. Der er således i forbindelse med eksemplet "stævning" valgt en definition, som ligger udenfor det maskinlæsbare korpus som den overordnede definition samt en udvidet definition, som er trukket fra det maskinlæsable korpus.

Med hensyn til 1) sprogbrugsundersøgelserne, registrering af fraseologi, kollokationer etc. er der udarbejdet en række programmer, som er beskrevet i projektgruppens rapport. Steffen Leo Hansen har i forbindelse med projektrapporten udarbejdet Jurkorpus 1 og 2, hvori er vist eksempler på konkordansprogrammer m.v. Det første nummer af Inst. f. Datalogivistik's publikation LAMBDA ³⁾ indeholder instituttets programbibliotek, hvor jurkorpusprogrammerne er dokumenteret.

Her skal blot omtales to konkordansprogrammer, som forfatteren har arbejdet med. Det drejer sig om et liniekonkordansprogram (KWIC-index) og et sætningskonkordansprogram. I liniekonkordansprogrammet er der mulighed for ved søgning på en bestemt term at opstille en liste med alle belægssteder, hvor den pågældende term f.eks. er midtstillet eller står først eller sidst på linien. Der vises således blot de ord, som omgiver den pågældende term begrænset til én linie. Med sætningskonkordansprogrammet fremkommer ved søgning på f.eks. ordet stævning alle de sætninger i korpus, som indeholder ordet stævning. Søgeordet kan "highlightes". I forbindelse med udarbejdelsen af recorden for "stævning", som vises senere, har forfatteren gennemført søgninger efter kollokationer m.v. både ved hjælp af KWIC-indexet og sætningskonkordansen. Søgningerne er gennemført og resultaterne registreret for

hver enkelt af teksttyperne i korpus med henblik på at sammenligne sprogbrugen i de forskellige teksttyper og resultaterne fra de to konkordansprogrammer. En overraskende høj procentdel af de registrerede kollokationer kunne "fanges" ved gennemgang af KWIC-indexet, som er meget overskueligt og hurtigt at arbejde med. Sætningskonkordanserne var naturligvis betydeligt mere omfattende og mere tidkrævende at arbejde med. En sætningskonkordans for ordet stævning havde således et omfang af 60 sider og KWIC-indexet blot 10, hver dækkende 418 belægssteder i den danske del af korpus.

Det er forfatterens opfattelse, at de muligheder for registrering og sammenligning af sprogbrug i forskellige teksttyper på dansk og det/de relevante fremmedsprog, som et maskinlæsbart korpus giver, vil kunne forbedre kvaliteten i fagsprogsordbøger betydeligt. Ved anvendelse af et tekstkorpus vil man kunne få svar på en lang række af de problemer, som gængse ordbøger ikke giver løsningen på i dag. Der tænkes her på de mange forbindelser, i hvilke et ord optræder i forskellige tekster, hvilke præpositioner, verber etc. de enkelte termer kan forbindes med, aspekter, som sjældent belyses i fagordbøger. Den metode, som oversættere af højt specialiserede fagsproglige tekster anvender til løsning af denne type sproglige problemer, er normalt at arbejde med originale kilder omhandlende det pågældende emne og ved læsning af disse indsamle oplysninger om sprogbrugen. Denne metode er meget tidsrøvende og ville kunne lattes betydeligt ved anvendelse af maskinlæsbare korpora.

Anvendelse af arbejdsdatabasen JURTERM

Efter udvælgelse af teksteksempler, definitioner m.m. fra det maskinlæsbare korpus lægges disse oplysninger over i en fil, og de pågældende data kan overføres maskinelt til databasens records, (jf. bilag 1). Oplysninger, som hentes fra kilder udenfor det maskinlæsbare korpus, indtastes. I første omgang udfyldes records

med danske oplysninger ved hjælp af en "maske" med på forhånd indsatte feltnavne, dernæst udfyldes for hver record engelske, spanske etc. "delmasker", som flettes sammen med de danske oplysninger.

Projektgruppens arbejde med DANTERM-recorden har bevirket, at den er blevet udvidet og ændret med henblik på at kunne tilgodese de krav, der er opstået i forbindelse med juraprojektet.

DANTERM-recorden er meget omfattende, da den er opbygget med henblik på at kunne rumme alle de oplysninger, som en række forskellige brugere måtte have behov for at registrere. I den enkelte record kan man således begrænse sig til kun at udfylde de kategorier, som er relevante for et givet projekts målgruppe.

Som nævnt har pilotprojektets målgruppe været oversættere og tolke, jurister, advokatsekretærer, undervisere i juridisk fagsprog, jurastuderende og sprogstuderende. Med henblik på at gennemføre forsøg med forskellige typer af ordbogsartikler/profiler til de nævnte målgrupper har det været nødvendigt at udfylde de fleste af recordens felter. Ved symposiet udleveredes to eksempler på udfyldte records for "stævning" i terminologidatabasen, som skulle illustrere anvendelsen af de enkelte kategorier til registrering af oplysninger af både sproglig og ekspertfaglig art. Det drejer sig om en record indeholdende dansk, engelsk og spansk og en record indeholdende engelsk. De udfyldte records er meget omfattende især på grund af de faktuelle oplysninger, hvor mange sider tekst er registreret og f.eks. stævningsformularer er indlæst. Af pladshensyn er de udfyldte records ikke medtaget her.

Eksempler på brugerprofiler

Som nævnt skulle de udfyldte records dække alle de definerede brugergruppers behov, og i ovennævnte records for termen stævning er de fleste felter udfyldt. På basis af oplysningerne i recorden kan der således sammensættes forskellige sæt af oplysninger -

brugerprofiler - til de enkelte målgrupper. Hver profil udgør en delmængde af oplysningerne i den samlede record. Der er endnu ikke foretaget forsøg med automatisk generering af sprogretnings- og brugerbestemte ordbøger, da der på nuværende tidspunkt ikke er indlæst et tilstrækkeligt stort antal records i databasen. Automatiseringen af genereringsprocessen vil bl.a. bestå i anvendelse af på forhånd definerede profiler. For at illustrere, hvordan sådanne profiler kan sammensættes, er der udarbejdet eksempler herpå.

Ved symposiet udleveredes følgende eksempler på brugerprofiler:

- 1) Profil sammensat med henblik på en jurastuderende
- 2) Profil sammensat med henblik på en oversætter - sprogretning dansk-engelsk.
- 3) Profil sammensat med henblik på en jurist - sprogretning dansk-engelsk.

Alle eksempler er udarbejdet for termen stævning.

Her skal af pladshensyn kun vises 1) og 2), da 3) er meget omfattende. Feltet FACTS indeholder således 8 sider tekst.

Nedenfor vises 1), som er en dansk record til brug for f.eks. en jurastuderende.

Som det ses, drejer det sig her om en profil sammensat udelukkende på basis af de danske oplysninger i recorden. Den danske del af terminologidatabasen vil således kunne anvendes i forbindelse med uarbejdelse af en et-sprogs systematisk definitions- og sprogbrugsordbog. Det skal understreges, at den profil, der er sammensat for en jurastuderende kun er et eksempel og ikke en fastlagt profil. Den studerende kan have behov for at få defineret et begreb, og få oplysninger om over- og underbegreber. Der kan ligeledes være behov for informationer om sprogbrug, når den studerende skal skrive opgaver, udfærdige eksempler op proceskrifter m.v.

EKSEMPEL 1

1.
DA_TERM støvning
DA_DEF 1 støvning: sagsøgerens første processkrift, der skal indeholde påstand, en kort fremstilling af de kendsgerninger, hvorpå påstanden støttes, og angivelse af de beviser som påberåbes samt parternes navne og adresser og retten, hvor sagen anlægges, jfr. Rpl. § 348.
(kilde: vEyJ - 5 s. 212)
- 2 Stævningen skal indeholde:
- 1) parternes navn og adresse, herunder angivelse af en postadresse i Danmark, hvortil meddelelser til sagsøgeren vedrørende sagen kan sendes, og hvor forkyndelse kan ske,
 - 2) angivelse af den ret, ved hvilken sagen anlægges,
 - 3) sagsøgerens påstand,
 - 4) en kort fremstilling af de kendsgerninger, hvorpå påstanden støttes, og
 - 5) angivelse af de dokumenter og andre beviser, som sagsøgeren agter at påberåbe sig.
- (kilde: Rpl § 348 stk. 2)
- DA_EXT Stævningen skal være udfærdiget i A 4-format og indleveres til den kompetente ret med 3 genparter, samt en genpart af de dokumenter, som sagsøgeren agter at påberåbe sig, for så vidt de er i hans besiddelse, jfr. § 348, stk. 2 og 3.
(kilde: GomC, s. 63)
- DA_SYNTAG affatte støvning, (kilde: GomC s.65-70)
støtte afgørelse på støvning , (kilde: GomC s.72)
anføre i støvning , (kilde: DRSH s.1, KarR s.3117, DRL s.2)
angive i støvning , (kilde: GomC s.72-78, KarR s.3139-3140, Rpl § 354)
bekendt; gøre sagsøgte bekendt med støvning, (kilde: GomC s.108-111)
beramme en støvning til forberedelse, (kilde: GomC s.205-210)
berigtige støvning, (kilde: KarR s.3143-45)
dom efter støvning , (kilde: DSGDFU)
forhåndskontrol af støvning , (kilde: GomC s.72-78, note 23)
forhåndsprøvelse af støvning , (kilde: KarR s.3141)
forkynde støvning for én , (kilde: GomC s.63-65)
forkynde støvning for én på bopælen ved en anden, (kilde: KarR s.3100)
forkyndelse af støvning , (kilde: KarR s.3100, GomC s.63-65)
forsyne støvning med påtegning , (kilde: KarR s. 3139-40)
fremlægge støvning, (kilde: GomC s.111-113, DSSK1)
genparter af støvning , (kilde: KarR s.3139-40, DVL s.2)
i; støvning i en sag, (kilde: DRSH, s.1)
imødegå støvning , (kilde: GomC s.247-249)
indgive støvning, (kilde: KarR s.3151-52)
indlevere støvning til berammelse, (kilde: KarR s.3114)
indlevere støvning til retten, (kilde: GomC s.63-65, Rpl § 251, KarR s.3119-20)
indlevering af støvning , (kilde: GomC s.63-65, KarR s.3121, Rpl § 354)
krav til støvning , (kilde: KarR s.3152)
lyder; støvning der lyder på, (kilde: DRSH s.1)
mangelfuld støvning, (kilde: KarR s.3141)
mangler ved støvning , (kilde: KarR 3136, GomC s.111-113)
meddele en støvning berammelsespåtegning, (kilde: KarR s.3157)
mod; støvning mod sagsøgte, (kilde: KarR s.3139-40)

modtage stævning i sag, (kilde: GomC s.148-150)
modtagelse af stævning, (kilde: GomC s.550-555)
opregne i stævning, (kilde: GomC s.72-78)
prøvelse af stævning, (kilde: KarR s.3141)
påberåbe i stævning, (kilde: DRSH s.1, KarR s.3100, GomC s.557-560)
påstand; stævning med påstand om, (kilde: DRSH s.1)
påstå; stævning, hvorved en sagsøgt påstås tilpligtet, (kilde: KarR s.3139-40)
påtegning på stævning, (kilde: Rpl § 360, DSUR2 s.1)
reparere mangelfuld stævning, (kilde: KarR s.3136)
rette stævning imod én, (kilde: GomC s.540)
tage stævning til følge, (kilde: GomC s.72-78)
tilbagekalde stævning, (kilde: KarR s.3141)
udarbejdelse af en ny stævning, (kilde: GomC s.159-161)
udformning af stævning, (kilde: KarR s.3069-3395, GomC s.114-119)
udfærdige stævning, (kilde: KarR s.3139-40, Rpl § 354, GomC s.63-65)
udfærdigelse af stævning, (kilde: KarR s.3139-40)
udtage stævning i sag, (kilde: KarR s.3124-25)
udtage stævning mod én eller for én ved værger, (kilde: KarR s.3125-26, DANSUR2 s.2)
udtagelse af stævning, (kilde: GomC s.412-413, KarR s.3133-34, DSGDFU s.22)
underskrive stævning, (kilde: KarR s.3125, GomC s.270-274)
sagens forberedelse

DA_SYST
DA_POS
DA_REL

2

BC-GEN	processkrift
BC-PART	skriftveksling
NC-GEN	byretsstævning landsretsstævning ad citationsstævning interventionsstævning
NC-PART	påstanden i stævningen krav i stævningen sagsfremstillingen i stævningen sagsbehandlingen i stævningen stævningens fremstilling af kendsgerninger bevisangivelsen i stævningen bevisfortegnelsen i stævningen stævningens dokumentfortegnelse stævningens bevisfortegnelse stævningens beløb stævningsbeløb bestævnet beløb påstævnet beløb
PREC	udtagelse af stævning; skriftveksling
SUCC	indlevering af stævning til retten; svarskrift
CAUS	søgsmålsgrund
RC	ankestævning at stævne at bestævne at indstævne at påstævne stævnevarsel stævningsgebyr stævningsmand stævningsmandsforkyndelse

Nedenfor vises den anden profil, som er sammensat med henblik på en oversætter:

1. EKSEMPEL 2

DA_TERM stævning
DA_GRAM sub -en -er
DA_STYL lov, lovkommentar, lærebog, processkrift, retsafgørelser, retsbogstilførsler
DA_SYNTAGEN
forkynde en stævning for én (kilde: GomC s.63-65) [1]
forkyndelse af stævning (kilde: KarR s.3100, s.3142-43) [2]
forsyne stævning med påtegning (kilde: KarR 3139-40) [3]
fremlægge stævning (kilde: GomC s.111-113, DSSK1) [4]
genpartere af stævning (kilde: KarR s.3139-40, DVL s.2) [5]
i; stævning i en sag (kilde: DRSH s.1) [6]
lyder; stævning der lyder på (kilde: DRSH s.1) [7]
påstand; stævning med påstand om (kilde: DRSH s.1) [8]
påtegning på stævning (kilde: Rpl § 360, DSUR2 s.1) [9]
reparere mangelfuld stævning (kilde: KarR s.3136) [10]
udtage stævning (kilde: KarR s.3124-25, DANSUR2 s.2) [11]
udtagelse af stævning (kilde: KarR s.3133-34, GomC s.65-70 DSGDFU s.22) [12]
EN_TERM writ of summons
EN_ALT ABBR writ
EN_GRAM sub -s
EN_STYL lovstof, lærebog, processkrift
EN_SYNTAGDA
amend; to amend the writ (kilde: RSC O.13,r.7) [10]
copy of the writ (kilde: RSC O.13,r.7) [5]
indorse; to indorse a writ (kilde: OHAC s. 194) [3]
indorsement of the writ (kilde: RSC O.18,r.15) [9]
issue; to issue a writ (kilde: RSC O.13,r.6) [11]
the issue of the writ (kilde: RSC O.29,r.1) [12]
produce; produce a writ (kilde: RSC O.42,r.5) [4]
serve; to serve a writ on sby (kilde: OHAC s. 206) [1]
service of the writ (kilde: OHAC s. 139) [2]
a writ claiming ... (kilde: OHAC p194) [7 + 8]
writ; the writ in the action (kilde: RSC O.18,r.6) [6]
EN_NOTE 'writ' og 'writ of summons' er alternative former og den forkortede form 'writ' er i ovenstående tekst- og syntagmeeksempler anvendt i henhold til almindelig praksis, jfr. nedenfor.
Rules of the Supreme Court Practice 1985 indeholder i O.1,r.4(1) en række definitioner, bl.a. følgende:
"In these rules, unless the context otherwise requires, the following expressions have the meanings hereby respectively assigned to them, namely - ... "writ" means a writ of summons"
EN_EQDA Den danske term stævning og den engelske term writ (of summons) er ikke ækvivalente, da de to retssystemer er fundamentalt forskellige.
I engelsk ret findes en række forskellige typer stævninger alt efter sagens genstand og ved hvilken domstol sagen anlægges, j.fr. beslægtede begreber anført under RC i recorden. På basis af en sammenligning af de anførte definitioner og det under EXPL. anførte skønnes det hensigtsmæssigt at anvende writ (of summons) til oversættelse af stævning. Endelig skal nævnes at dette iflg. gængse ordbøger er alm. oversættelsespraksis.

I ovenstående eksempel er en del oplysninger sorteret fra i forhold til oplysningerne i den første profil. F. eks. er definitionerne ikke medtaget, og der er kun anført "sprogparrelaterede kollokationer. Selvom oversætteren først og fremmest har brug for at få at vide, hvad en term på udgangssproget kan oversættes med på målsproget, vil det normalt være nødvendigt inden for fagsprog at anføre en definition eller forklaring, da oversætteren ikke kan have fagindsigt i de mange områder han/hun præsenteres for i oversættelsesøjemed. Det skal igen understreges, at de viste ordbogsprofiler kun er eksempler. Det er ligeledes nødvendigt med oplysninger om ækvivalens, især indenfor et område som juridisk sprog, hvor retssystemer og retsvirkninger adskiller sig væsentligt fra land til land, og ækvivalensen derfor ofte kun er tilsyneladende.

Som nævnt ovenfor har projektgruppen ønsket at kunne registrere den type oplysninger om sprogbrug, som oversætteren for det meste savner i traditionelle fagsprogsordbøger. Det drejer sig om kollokationer m.v., som anført i feltet SYNTAG. Det er her muligt at kvalificere visse oplysninger, som udelukkende er relevante for ét sprogpar, f.eks. dansk-engelsk og ikke dansk-spansk. Feltnavnet DA-SYNTAGEN indeholder således "syntagmer", der er parallelle med engelske "syntagmer". I eksemplet er kollokationerne i det danske og det engelske felt anført hver for sig i alfabetisk rækkefølge. I en egentlig ordbogsartikel bør informationerne naturligvis flettes, således at parallelle danske og engelske teksteksempler står ved siden af hinanden. Det har imidlertid været vanskeligt at finde en metode til klassifikation af oplysningerne i feltet SYNTAG. Yderligere undersøgelser er nødvendige for at dette problem kan løses tilfredsstillende. Endvidere skal der arbejdes med metoder til automatisk sammenføring af parallelle kollokationer. I eksemplerne er relationerne mellem sprogparrene markeret med tal i kantet parentes.

Konklusioner og generelle bemærkninger

Ovenfor er de krav til korpussammensætning, som projektgruppen har stillet, kort beskrevet. Efter forfatterens opfattelse har

disse krav været for ambitiøse, da det kræver uforholdsmæssigt store tekstmængder, når man skal tilgodese et ønske om, at korpus skal kunne anvendes til sprogbrugsundersøgelser og samtidig fungere som en aktuel vidensbase. Såfremt en jurist skulle kunne "slå op" i korpus med henblik på at se, hvad der er gældende ret på et bestemt område, vil dette kræve et korpus, som kan fungere som et opdateret juridisk dokumentationssystem. Dette kunne måske gennemføres for et meget begrænset juridisk emne, men ville være urealistisk, hvis større fagområder skulle inddrages. Det ville være mere hensigtsmæssigt til vort formål at dele korpus i to, således at det ene tilgodeså sprogbrugssiden og det andet dokumentationssiden, hvor man måtte overlade vidensbaseopbygningen til andre og gennem samarbejde trække på disse vidensbaser. Via netværk er det allerede muligt at benytte forskellige juridiske dokumentationsbaser, og der arbejdes i flere lande med udvikling af juridiske eksperter-systemer. Det vil næppe være økonomisk muligt at opbygge og vedligeholde sådanne systemer alene til brug for ordbogsproduktion, selv med den nyeste OCR-teknologi.

Et andet problem i forbindelse med udnyttelse af tekstkorpus har været de meget store datamængder, som skulle gennemgås ved registrering af sprogbrug. Som nævnt fandtes 418 belægssteder for termen "stævning" alene i det danske korpus. Selv med systemets muligheder for at søge på delmængder i korpus virker on-line søgning tung, da man skal "blade" mange eksempler igennem og læse dem på skærmen. Det vil i sådanne tilfælde ofte være at foretrække at arbejde med udskrifter f.eks. i form af konkordanser etc.

Med henblik på at lette overførslen af data fra korpus til terminologidatabase har projektgruppen ønsket mulighed for at kunne se dele af korpus og dele af terminologidatabasen samtidig. Forbedrede metoder - f.eks. "vinduer" med interaktion mellem disse - ville gøre arbejdet med udfyldning af records nemmere.

Selvom det anvendte edb-system (DANSTATUS) ikke umiddelbart er tilstrækkeligt fleksibelt som arbejdsredskab i forbindelse med overførsel af data fra korpus til terminologidatabasen, viser erfaringerne fra pilotprojektet generelt, at den edb-støttede arbejdsmetode med on-line søgning og anvendelse af udskrifter i form af f.eks. konkordanser er et værdifuldt supplement, som letter og forbedrer arbejdet i sammenligning med traditionelle arbejdsmetoder. F.eks. kan udvælgelse af konteksteksempler, genfindning af belægssteder og teksttypologiske undersøgelser gennemføres langt hurtigere og med større pålidelighed.

Som tidligere nævnt har et af målene med projektet været at forsøge at forene terminologiske og leksikografiske arbejdsmetoder og principper. Indenfor leksikografien forskes bl.a. i ordbogtypologi og artikelstruktur, områder som er særdeles relevante for pilotprojektet. Brugeraspectet i forbindelse med ordbogsproduktion er således et fælles problem for leksikografer og terminologer. Der arbejdes med receptions- og produktionsordbøger eller passive og aktive ordbøger, hvor man skelner mellem brugere med og uden modersmålskompetence og indretter kravene til artikelstruktur og indhold derefter. Målet er bl.a. at "spare" en række oplysninger f.eks. af encyclopædisk art, hvor disse oplysninger forudsættes at være brugeren bekendt.

Spørgsmålet er, om disse teorier og metoder umiddelbart kan anvendes i forbindelse med fagsproglig leksikografi og terminologi. Det er forfatterens opfattelse, at der på dette punkt må skelnes mellem fagsprog og ikke-fagsprog, da f.eks. "modersmålsbrugeren" vil have en betydeligt mindre sproglig kompetence og begrænset viden om faktiske forhold, når det drejer sig om fagsprog. Det ville være ønskeligt med en fælles forskningsindsats indenfor terminologi og leksikografi med fokus på brugerbehov generelt.

Også inden for området oversættelsesteori synes der at være behov for en øget forskningsindsats med henblik på fagsprog og fagsproglig oversættelsesstrategi. Her er det især ækvivalensproblematikken, som inden for et område som oversættelse af juridiske tekster volder store vanskeligheder.

Afslutningsvis skal de videre perspektiver for juraprojektet omtales. Som det er fremgået af ovenstående, bør der arbejdes videre med flere af projektets faser, især de to sidste faser, oprettelse af en terminologidatabase og generering af ordbøger. Det videre arbejde vil være koncentreret om udbygningen af terminologidatabasen og udvikling af metoder til automatisk generering af sprogretnings- og brugerbestemte ordbøger. Det vil være ønskeligt, om sprogene fransk og tysk, som oprindeligt planlagt inddrages i projektet. Hvorvidt projektet kan videreføres afhænger af, om der kan tilvejebringes de tilstrækkelige menneskelige og økonomiske ressourcer. Det kan i den forbindelse oplyses, at medlemmer af projektgruppen forhandler med et forlag, som er interesseret i at udgive en dansk-engelsk juridisk ordbog i samarbejde med det erhvervsproglige fakultet.

Referencer:

- 1) Symposium om Datorstö Terminologi och Lexikografi, Helsingfors 1985, s. 29
- 2) Ark 36, HHK, 1987
- 3) Lambda 1, Institut for Datalogvistik, HHK 1986