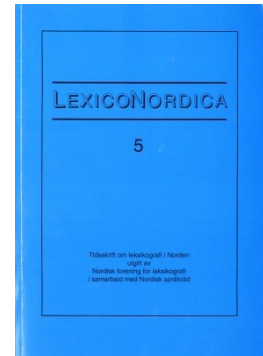


LexicoNordica

Titel: Norsk terminologisk database - NOT
Forfatter: Johan Myking
Kilde: LexicoNordica 5, 1998, s. 39-53
URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre LexicoNordica (1-16) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Johan Myking

Norsk terminologisk database – NOT

In this paper the project "Norwegian terminology database. Structure and applications – NOT" and the NOT termbase are presented and discussed. I attempt at giving an assessment of the general ideas underlying the development of the termbase as well as the challenges it encounters within the framework of current Norwegian terminography. The aim of being a "living termbank" rooted in specific projects is in partial conflict with the "generic use" of the base, a problem that could possibly be reduced by introducing corpus-related methods. The possibility of developing methods for validation by introducing concepts from sociolinguistics is discussed. The NOT project and termbase should be conceived of as examples of non-centralistic terminography within a network of language resources.

1. Innleiing

Feltet språkteknologi vil vera eit viktig satsingsområde for Universitetet i Bergen dei næraste åra. Denne satsinga er no organisert i Forskningsprogram for humanistisk informasjonsteknologi, HIT. Norsk termbank er organisatorisk og fysisk lagd inn i dette programmet og det tilhøyrande HIT-senteret. Terminologi og terminografi skal såleis gjerast til ein sentral del av satsinga. Det vil seia at den modellen, "konseptet", for terminologiarbeid som har rådd grunnen, må utviklast vidare. Men dette krev òg ein viss diskusjon og kritisk vurdering av modellen, og det er målet med denne artikkelen. Eg vil presentera og diskutera to relaterte fenomen som begge er refererte til med akronymet NOT:

- Norsk termbanks termdatabase NOT
- utviklingsprosjektet *Norsk terminologisk database. Struktur og applikasjonar*, der databasen NOT er det primære objektet

Den elektroniske termdatabasen NOT inneheld terminologi frå konkrete terminologiske prosjekt. Denne terminologien er utvida og samredigert med terminologi frå andre kjelder for å kunna gje eit mest mogeleg representativt uttrykk for dei respektive fagområda som basen dekkjer. Filosofien bak dette grunnlaget i konkrete prosjekt gjev opphav til ein del problem og utfordringar, knytte til:

- termdatabasar som normeringsverktøy
- sentralisert vs. desentralisert terminografi
- brukarrespons og validering
- balansen mellom "kontekstspesifikk" og "generisk" nytte

Hovudvekta i artikkelen ligg på dei språkpolitiske og sosiolingvistiske aspekta som ligg innebygde i dei faglege måla for NOT, og som førebels langt på veg ligg på eit programmatisk nivå. Når det gjeld dei teknologiske og dei organisatoriske aspekta, viser eg til Meurer (1997) og til Andersen (1997).

2. "NOT": konsept, prosjekt og termbase

2.1 "Levande termbank"

Sidan starten i 1981 har målet for arbeidet ved Norsk termbank vore omtala som "ein levande termbank". Dette er ingen presis fagterm, snarare uttrykk for eit "konsept" som omfattar m.a. at det terminografiske produktet – termbasen – skal veksa fram frå aktivt terminologiarbeid i aktuelle prosjekt og såleis vera ein refleks av den dynamiske termvoksteren i samfunnet. Det vil altså seia at spesifikke prosjekt i ein spesifikk kontekst utgjer ein sentral del av datagrunnlaget for databasen, og at databasen fungerer som ei overbygning for slike prosjekt. Medan eit prosjektavhengig termmateriale har spesifikk og avgrensa bruksverdi, må termbasen ha "generisk" verdi, dvs. gje eit representativt uttrykk for termininventaret i det fagområdet (dei fagområda) som er representert(e) i basen.

2.2 Prosjektet NOT

Prosjektet *NOT – Norsk terminologisk database. Struktur og applikasjonar* skal vara i tre år frå 1997, og har to siktemål:

a) Internt: å utnytta og tilretteleggja det akkumulerte terminologiske materialet ved Norsk termbank for anvende og teoretiske forskingsformål. Grunnstammen i dette materialet er termdatabasen NOT. Særleg vekt skal det leggjast på korpusrelatert metodikk og utvikling av infrastruktur. Prosjektet skal på denne måten knytast til paraply-samarbeidet Norsk *infrastruktur for språkteknologi* (NIFST). Jf. Andersen (1998).

b) Eksternt: å utvikla ein prototyp på ein fleksibel, fleirspråkleg og distribuert termdatabase, som skal fungera både som arbeidsverktøy i terminologiarbeid og som terminologisk oppslagsverk. Det skal utviklast formidlingsformer og oppslagsrutiner gjennom bruk av nye medium (CD-ROM, Internett). Databasen skal òg kunna inngå som modul i integrerte terminologihandterings- og omsetjingssystem. Jf. Meurer (1997).

Som ein del av dette eksterne arbeidet er det òg eit mål å utvikla databasen NOT vidare når det gjeld terminologisk innhald. Det vil seia utviding med nye fagområde, raffinering av det eksisterande innhaldet og formatet, kort sagt: vanlege terminografiske arbeidsoppgåver.

2.3 Termdatabasen NOT

Termbasen NOT er hovudproduktet ved Norsk termbank. Basen fungerer som ei "elektronisk fagordbok", hovudsakleg mellom engelsk og norsk (bokmål). Den eksisterande brukarversjonen av basen vert køyrd under Windows og er marknadsført kommersielt. Oppdatering og brukarkontakt vert ivareteke gjennom informasjonsbladet *NOT-nytt*.

Namnet NOT (NOrsk termbanks Termdatabase) vart først brukt i 1987. Det akkumulerte terminologiske materialet frå einskilde større prosjekt vart på det tidspunktet formatert og tilrettelagt for stormaskiner og distribuert gjennom ei abonnementsordning. Abonnentane var i denne fasen først og fremst oljeselskap og ingeniørselskap som hadde levert "råmaterialet" til den største termsamlinga. Denne løysinga var dyr, men kunne forsvarast med den høge og direkte brukarnytten og den høge oppdateringsfrekvensen som var mogeleg, éin gong i kvartalet. Med overgang til salsversjon i Windows fekk ein til ei billigare løysing som fleire kunne ta i bruk, men med ein lågare oppdateringsfrekvens.

NOT dekkjer i dag 38 fagområde. Med omsyn til kvantitet er dei viktigaste områda desse: ADMINISTRASJON, BORING, ELEKTRO, FYSIKK, GEOLOGI, KJEMI, KVALITETSSIKRING, MEKANIKK og PROSESS. Kvart enkelt fagområde ligg inne på éi separat kjeldefil. Det finst altså totalt 38 kjeldefiler. Desse filene er bygde opp av postar. I samsvar med den onomasiologiske tradisjonen i terminologiarbeidet dekkjer kvar post alle opplysningar om eitt omgrep. Framleis i samsvar med denne tradisjonen er basen strukturert ut frå prinsippet om homonymi. I dag dekkjer basen ca. 30.000 omgrep, realiserte ved ca. 90.000 termar (hovudtermar og synonym på begge språka medrekna). Basen er ikkje fullt utbygd med nynorske ekvivalentar på alle

fagområda. Grammatisk informasjon ligg ikkje i basen, men det er planlagt å leggja inn slik informasjon dei næraste par åra.

Dei ulike informasjonskategoriane ligg i separate felt i posten. Dei viktigaste felte er: norsk hovudterm, norsk synonym, engelsk hovudterm, engelsk synonym, norsk definisjon, engelsk definisjon. I tillegg finst det felt for referansar, fagområde, kommentarar, kontekst, godkjeningsdato osv. Nokre felt er obligatoriske, andre ikkje. Talet på felt kan utvidast etter behov.

=		
	N hvd	produksjonsbrønn
	ref	ODBOR
	gkj	NT-OD 87-09-01*
	N nyn	produksjonsbrønn
	ref	NT:OR
	N def	Brønn som produserer eller er i stand til å produsere hydrokarboner
	ref	OD:NPD - Contribution No 33
	E hvd	production well
	ref	GFA 000 Intro E:13
	ref	OD:NPD - Contribution No 33
	reg	GFAhoved 85-06-13**
	reg	Statoilstav-Note 87-07-08*
	reg	NT 87-12-08***
	E syn	producing well
	kon	28 "producing wells" (14 in each shaft of which 2 may be subsea completions)
	E syn	output well
	ref	DGI:6020
	eg	Statoilstav-Note 87-07-08*
	. brk	BORING
	. inn	89-01:GS,94-06:EE
=		

Figur 1. Døme på post (i "råformat", ikkje tilgangsgrensesnitt)

Denne posten har, som vi ser, opphav i ein spesifikk kontekst (GFA = dokumentasjonen til oljeplattformen Gullfaks A), men supplert med andre kjelder (t.d. OD = Oljedirektoratet).

3. Termdatabasar som normeringsverktøy

Det første store råmaterialet for NOT var den terminologien som vart utvikla i samband med fornorskinga av oljespråket på 1980-talet, først

og fremst gjennom "Terminol-prosjektet" i 1984–85 (Roald 1985). Dette arbeidet kravde terminologisk gjennomarbeiding av eit spesifikt tekstmateriale, først og fremst ca. 10.000 sider med opplæringsdokumentasjon. Terminologien vart analysert og deretter samordna med eksisterande terminologi frå andre kjelder, så langt det var mogeleg. Det samla terminologiske materialet var dermed ei blanding av eksisterande og ny terminologi, men sjølv på felt der det fanst terminologi frå før, representerte materialet ei kraftig utviding. Derfor framstod materialet som eit vesentleg kvalitativt og kvantitativt tilskot til norske språkressursar.

All terminografi er i prinsippet normativ, og alle termbasar er i prinsippet "normeringsverktøy". Ambisjonsnivået bak databasen NOT var heilt frå starten høgt når det gjaldt styring av språkbruken: for det første når det gjaldt variantavgrensing og synonymkontroll, for det andre når det gjaldt termdanningsmønster og i synet på "god" og "dårleg" språkbruk i terminologi. Det er derfor dekning for å hevda at NOT har hatt ein meir eksplisitt språkpolitisk profil enn mange andre termbankar, og det heng sjølv sagt saman med den høge symbolverdien dette terminologiarbeidet hadde. Desse erfaringane er problematiserte i ymse andre arbeid (t.d. Rangnes 1997, Myking 1996).

I mange samanhengar kom det kanskje til uttrykk ein litt urealistisk optimisme med tanke på kor mykje samordning, harmonisering og standardisering av terminologi som kunne oppnåast gjennom dette arbeidet. Fordi terminologiarbeidet var nyskapande og konteksten nasjonaløkonomisk og politisk svært viktig, tenkte ein seg at NOT med tida kunne få ein viktig funksjon som overordna normeringsverktøy for norsk fagspråk generelt – både reint fagleg og i nokon grad politisk. Men funksjonen som overordna språkpolitisk instrument kan vera vanskeleg å kombinera fullt ut med funksjonen som oppslagsverktøy i arbeidet med fagspråklege tekstar i ein spesifikk samanheng, t.d. i omsetjing. I denne siste funksjonen vil éin spesifikk database oftast vera berre eitt av fleire arbeidsverktøy. NOT har i praksis vore marknadsført og utnytta som éin av fleire terminologiske ressursar for norsk.

4. Sentrale vs. "sentralistiske" termbasar

I fleire av dei nordiske landa er terminologiarbeidet identifisert med ein sentralinstitusjon, der det til kvar institusjon høyrer ein termbank/termbase: TNC med Termdok i Sverige, TSK med TEPA i Finland. Namnet på den danske termbanken Danterm signaliserer den same

nasjonale statusen, sjølv om det er eit nettverk av institusjonar som står bak den, og ikkje éin sentralinstitusjon.

Det er ikkje i det heile mogeleg eller ønskjeleg å sjå vekk frå behovet for sentrale terminologiske ressursar, dvs. nasjonale normeringsverktøy med eit overordna språkpolitisk mandat. Tvert om er vel stoda oftast den at slike ressursar er for svakt utbygde. I norsk samanheng var Norsk termbank rundt 1980-talet tiltenkt ei rolle som nasjonalt terminologisk "arkiv" eller termbase i ein nasjonal terminologisk infrastruktur. Men seinare har det utvikla seg slik at det finst ein sentralinstitusjon i Oslo, Rådet for teknisk terminologi (RTT), som produserer terminologiske ordbøker, og så Norsk termbank i Bergen, der vi har databasen NOT og driv terminologiske prosjekt på eksternfinansiert grunnlag, men i ein forskingskontekst. Det terminografiske arbeidet i Noreg har såleis i lang tid hatt to hovudtyngdepunkt.

Sjølv om det finst ein sentralinstitusjon, kan norsk terminografi derfor ikkje karakteriserast som "sentralistisk". Behovet for utbygde elektroniske terminologiske leksikon – termbasar – er stort og aukande, både i nasjonal, regionalt nordisk og europeisk samanheng. I Noreg er det fleire viktige fagmiljø som har sett i gang arbeid med å byggja opp ein elektronisk infrastruktur for terminologi og terminologirelaterte språkressursar. Dette skjer først og fremst for å løysa eigne interne behov, og det skjer utan ei sentralt koordinert satsing på terminologiutvikling. Nokre viktige døme på dette:

a) Den norske *sentralinstitusjonen* for terminologiarbeid er Rådet for teknisk terminologi, som publiserer fagordbøker innanfor mange ulike fagfelt. Det metodiske grunnlaget er systematiske (områdedekkjande) terminologiske undersøkingar utførte av terminologiske arbeidsgrupper. Det er såleis tale om terminologiarbeid i ei "klassisk" utgåve.

b) Informasjonsforvaltinga innanfor det *medisinske fagområdet* er grunnlagd på viktige klassifikasjonssystem for sjukdomstilstandar, diagnosar, utstyr m.m. (ICD, ICPC, DRG osv.), og tilpassinga ("lokaliseringa") til norsk kviler på ei omfattande terminologisk gjennomarbeiding. Denne terminologiske infrastrukturen skal ha forgreiningar til m.a. medisinske forskingsmiljø og alle sjukehus. Norsk termbank har gjennomført fleire slike prosjekt, særleg den norske utgåva av den internasjonale sjukdomsklassifikasjonen (ICD-10, 1996). Gjennom dette prosjektet har Norsk termbank etablert kontakt og samarbeid med t.d. KITH i Trondheim, som er ein av dei sentrale aktørane på dette området, og som arbeider for å utvikla ein nasjonal medisinsk terminologidatabase på grunnlag av arbeidet med klassifikasjonar og

kodeverk (Buan 1997). Ein slik medisinsk termbase vil krevja ein omfattande innsats med normering av det medisinske fagspråket, både på omgreps- og uttrykkssida. Men ein slik termbase ville òg vera eit effektivt verktøy for ei slik normering.

c) Eit omfattande terminologisk standardiseringsarbeid på miljøsektoren i europeisk samanheng er koordinert i Noreg av Norsk allmennstandardisering (NAS). Norsk termbank tek her del i utviklinga av ein *Norsk miljøtesaurus* på grunnlag av *General European Multilingual Environment Thesaurus*.

I norsk samanheng finst det dessutan mange døme på at visse bransjar, utdanningsetatar og -institusjonar og forlag går saman om konkrete terminologiprojekt der dei større terminologiske institusjonane ikkje er kopla inn, og der arbeidet ikkje har noko nasjonalt, "overgripande" mål som peikar ut over seg sjølv.

Klassifikasjonssystem og dokumentbaserte tesaurusar av dei typene som er nemnde her, er ikkje systematiske framstillingar av "kunnskapen" i eit fagfelt slik eit terminologisk oppslagsverk tradisjonelt er oppfatta. Dei er spesifikke (og omfangsrrike) tekstlege framstillingar av eit utsnitt av "verda", altså av ei samling "objekt", slik dei er organiserte for visse avgrensa formål. Men klassifikasjonane er oftast oppbygde etter hierarkiske prinsipp som tilsvarer dei strukturprinsippa ein nyttar i terminologiarbeid, og ein stor del av tekstmassen i ein klassifikasjon er termar. Det same gjeld jo i særleg grad tesaurusar. Arbeidet med desse to typene oppslagsverk – klassifikasjonar og tesaurusar på den eine sida, strukturerte terminologiar på den andre – har derfor svært mykje til felles. Medverknad frå terminologar har vorte viktigare i t.d. det medisinske klassifikasjonsarbeidet (Hirs 1993), og det har alltid vore eit nært samarbeid mellom terminologi, standardisering og tesaurusutvikling.

Ingen norske fagmiljø har i dag eit samla fagleg ansvar og forvaltingsansvar for forskning på dette området, og dette representerer eit problem for den samla slagkrafta i arbeidet. Prosjektnamnet *Norsk terminologisk database* signaliserer ein ambisjon om å fylla ein slik funksjon. Eitt av måla med prosjektet er som nemnt å utvikla ein prototyp på ein database og eit "konsept" for korleis eit slikt overgripande terminologiarbeid kan organiserast, både fagleg, administrativt og teknologisk (stikkord: direktetilgang, Internett m.m.).

Ingen ting tyder likevel på at planar om eitt og berre eitt nasjonalt normeringsverktøy for terminologi skal verta lanserte eller kunna realiserast på eit politisk plan, sjølv om dei tekniske løysingane hadde

komme på plass. Ambisjonsnivået må derfor konsentrerast om nettverkssamarbeid og eit fagleg arbeid for korleis kontekstspesifikke termsamlingar kan utviklast til "generiske" verktøy. Modellen som terminografi av typen NOT meiningsfullt kan inngå i, må derfor vera "distribuert": Mest mogeleg tilgang og kontakt med sluttbrukarar, oppdragsgjevarar og andre språkressursar, men med eit sentralt knutepunkt i det faglege og administrative apparatet rundt basen.

Når det gjeld nettverkssamarbeid, er behovet understreka både gjennom nordiske og europeiske tendensar. På nordisk plan er det t.d. lansert planar (Westerberg 1997) for ein nordisk termbank, *Nordterm-Net*. Gjennom Internettet er dette mogeleg på ein heilt annan måte enn for berre for få år sidan. Omdøypinga av dette Nordterm-prosjektet, frå "Samnordisk termbank" (Hjulstad/Westerberg 1997) til *Nordterm-Net*, er i seg sjølv karakteristisk for denne nyorienteringa.

Konsekvensen av dette er at brukarane må sjå på dei ulike terminologiske ressursane som komplementære, ikkje ekvivalente, sjølv om dei t.d. dekkjer det same nasjonalspråket. Eitt og same fagområde kan m.a. ha ulik dekning i to termbasar. Dette kan i sin tur henga saman med datagrunnlaget og den innsamlingsmetodikken som ligg til grunn for dei ulike basane. Dette er til sjuande og sist fastlagt av formålet og brukargruppa som det terminografiske produktet har.

5. "Brukarsirkelen" – validering og kontekstforskning

I moderne terminografi er behovet for effektiv oppdatering på grunnlag av tilbakemelding frå brukarane vurdert som svært viktig. For NOT har denne kontakten til no vore ivareteken via meldingsbladet *NOT-nytt*, men siktemålet er å etablera ein meir formell og systematisk kontakt med brukarmiljøa. Det bør vera ein "brukarsirkel" mellom terminolog, termbase og brukar, som sikrar maksimal relevans og aktualitet på dei terminologiske data i basen. Ein viktig del av denne prosessen er 'validering', som er definert m.a. slik:

A validated system is one which has been proved to do what it purports to do. The proof of validation is obtained through collection and evaluation of data, preferably, beginning from the process development phase and continuing through into the production phase. (U.S. Food and Drug Administration, "FDA".)

Terminografiske og leksikografiske produkt er reiskapar i kommunikasjon, men òg i normativt språkarbeid. Terminologi og leksikografi kan derfor relaterast til Einar Haugens kjende modell for språkplan-

legging med fasen 'implementering', iverksetjing (Haugen 1987). Forsking i implementeringsproblem er mindre utbygd enn forskning i andre delprosessar av språkplanlegging, som t.d. kodifisering, men er derfor etterlyst i norsk normeringsdebatt (m.a. Omdal 1996). Studiet av brukarrespons på terminologi og terminografiske produkt kan definerast som ein del av eit slikt implementeringsstudium. Det kan gjelda problem som brukaraksept og brukarbehov, funksjonalitet, dekningsgrad m.m., både på ordnivå og for leksikografiske produkt som heilskap.

Innanfor terminologi ser det ut til at desse spørsmåla enno ikkje er drøfta særleg sterkt, endå om merkelappen "sosioterminologi" ofte er nytta med referanse til dette problemområdet (t.d. Maurais 1993). Av konkrete prosjekt finst det rett nok eit EU-finansiert prosjekt om validering av terminologiresursar, Interlinguistic terminology validation (Interval, 1997). Dette prosjektet dekkjer 7 EU-språk – finsk er med av dei nordiske språka – og har både akademiske og kommersielle partnerar.

Eit prosjekt av typen Interval inneheld likevel ikkje slik teori- og metodediskusjon som trengst for å utvikla terminologisk og terminografisk arbeid som vitskaplege disiplinar. På same måten som ved språkplanlegging generelt er det rimeleg å venta at ein vesentleg del av teori- og metodegrunnlaget kan hentast frå sosiolingvistikk og sosiologisk orientert språkforsking.

Ein termbase som er utvikla i eit forskingsbasert miljø bør såleis svara på behovet for validering med å gje eit tilskot – med overføringsverdi – til det metodiske og teoretiske grunnlaget for terminologisk validering generelt. Rundt databasen NOT må det såleis t.d. utviklast:

- teoretisk kompetanse i validering spesifikt retta mot språklege (terminologiske) data
- rutinar og metodar for validering av informasjonsinnhald og struktur i NOT, dessutan konkrete resultat
- rutinar, metodar og prinsipp for overføring av data frå NOT til andre typar språkteknologiske produkt innanfor forskingsprogrammet HIT og eksternt

I eitt av prosjekta som er planlagt rundt NOT, men enno ikkje finansiert og igangsett, er validering og kvalitetssikring på grunnlag av brukaranalyse eit hovudmål (*NOT-konval*, jf. Andersen/Myking 1997). Ein viktig del av det metodegrunnlaget som kan utnyttast, er utvikla ved NT gjennom dei sosiolingvistiske delane av prosjektet *Nordsjøen som*

språkleg-kulturelt laboratorium. Fleire arbeid frå dette prosjektet handlar om sosiolingvistisk analyse av termdanning og termdanningsmønster (jf. t.d. Rangnes 1997).

6. "Generisk nytte"

6.1 Datagrunnlag og "kontekstuell skeivheit"

Det overordna målet for terminologiarbeid er å gje ein uttømmende og strukturert representasjon av kunnskapen innanfor eit fagområde, slik denne kunnskapen er nedfelt i inventaret av omgrep og termar. Den representasjonen av eit fagområde som termbasen skal gje, må følgjeleg vera "generisk", dvs. ikkje bunden til spesielle kontekstuelle variantar av den faglege aktiviteten (t.d. kjemisk terminologi i variantar nytta innanfor éi einskild kjemisk bedrift). Eller så må slik kontekstuell variasjon i alle fall vera identifisert og beskriven på ein måte som gjer det mogeleg å kontrollera han.

Grunnlaget for databasen NOT vart lagt på eit tidspunkt då korpusbasert metodikk ikkje var vanleg i terminologiarbeidet. Materialet vart i hovudsak bygd opp gjennom ein kombinasjon av dei to tradisjonelle metodane i terminologiarbeidet, den *punktuelle* og den *systematiske* (jf. Picht 1983): Prosjektet tok utgangspunkt i tekstar til bruk i ein heilt konkret samanheng, og målet var omsetjing til norsk. Men den terminologiske informasjonen vart strukturert i relevante omgrepfelt og samordna med meir "generiske" oppslagsverk og kjelder så langt det var mogeleg. Denne kombinasjonen var nødvendig for å løysa dei store arbeidsoppgåvene på 1980-talet, først og fremst oppbygginga av den norske oljeterminologien, og gjorde det mogeleg på svært kort tid å byggja opp eit terminologisk materiale med høg bruksverdi både for dei opphavlege brukarane og i ein vidare kontekst.

Dette forklarar kvifor datainnsamling med basis i konkrete prosjekt har vore ein viktig del av "filosofien" bak NOT. Tilknyttinga til omfattande og aktuelle prosjekt kan – som nemnt ovanfor – gje mange fordelar, særleg med omsyn til kvantitet, brukarkontakt og aktualitet. Men det finst òg problematiske sider, først og fremst knytte til dei metodiske problema rundt representativitet og generisk nytte: Det som er adekvat i éin kontekst, er ikkje per definisjon representativt for det sentrale i fagområdet. Dette problemet er t.d. utførleg drøfta av Wright (1997:147ff.).

Den "generiske" verdien av NOT har vorte utvikla etter kvart, gjennom utviding med nye fagområde og gjennom kvalitativ bear-

beiding av det opphavlege materialet. Særleg to arbeidsoppgåver er viktige:

- siling av "hapax", dvs. tilfeldige ord- og morfemkombinasjonar
- systematisk omgrepsanalyse og komplettering av omgrepssystem

Denne typen oppretting skuldast stort sett tolkingsproblem ved den engelske strukturen i nominalfrasar. Dette problemet er velkjent og mykje diskutert i samband med "intuitiv" ekserpering, dvs. ekserpering utan støtte i korpusbasert eller (halv)automatisk termekserpering.

6.2 Koreferanse?

Potensiell koreferanse komprimert i ein nominalfrase er eit vanleg døme på dette problemet. I strengen "*ball block valve*" viser analysen at det er tale om to termar (resp. omgrep) med ulike motiverande drag:

ball valve --> *kuleventil* [form, konstruksjon]

block valve --> *stengeventil* [funksjon, formål]

Dersom denne typen strengar skulle få status av terminologisk oppslag, måtte det på norsk skje gjennom ei omskriving:

**stengeventil av kuletypen*

Jf. òg *displacer type level transmitter* --> *nivågiver av fortrengnings-type*.

6.3 Term eller skilt?

I ein spesifikk tekst kan det vera vanskeleg å skilja uttrykk som refererer til ein klasse av gjenstandar, frå uttrykk som refererer til ein unik gjenstand. I mange tilfelle vil det seia at oppslaget nærmar seg eit namn, og settet av termar beskriv snarare ein konkret gjenstand (t.d. ei plattform) enn eit fagleg område:

mesanindekk vs. *øvre mesanindekk*

Jf. "*boredekk*" vs. "*boredekket*".

Adjektiviske beskrivarledd som "øvre" kan likevel vera enklare å "avsløra" enn andre typar, som t.d.:

first stage separator -->

**første trinns separator* [lokalisering?]
førstetrinnsseparator [type? funksjon?]

Det første analysealternativet signaliserer ein svakare terminologisk status enn det andre.

6.4 "*Hapax*"

Den engelske nominalfrasestrukturen kan gjera det vanskeleg å dra grensene mellom det sentrale faglege omgrepet i frasen og påbyggingar som er meir eller mindre tilfeldige. Det som først er ekserpert, kan vera vanskeleg å sila vekk. Til overtermen *froth* = *skum* kan det t.d. leggjast adjektivledd som t.d.:

oil-rich froth *oljerikt skum*

– og problemet er å avgjera om dette er ei fagleg kategorisering eller berre ei tilfeldig beskriving. Liknande kontekstuelle påbyggingar kan skje gjennom substantiviske etterledd:

froth layer *oljeskumlag*

Og det same kan skje når overtermen sjølv er samansett:

froth separator *froth separator logic*
oljeskumseparator *oljeskumseparatorstyring*

Ikkje alle desse døma er henta frå basen, men problemtypane dei representerer, har spela ei rolle i utbygginga av basen. Mangelen på "generisk verdi" viser seg altså dels som redundans, dvs. innslag av materiale utan terminologisk status, og dels som "skeivheit", materiale som meir beskriv ein konkret/unik gjenstand enn eit fagområde.

Tilsynelatande er døma banale. Dei er likevel ikkje tilfeldige, men ei følge av den modellen for terminologiarbeid, "konseptet", som NOT er skapt av. Dette er problem av prinsipielt same type som andre fagmiljø

vil møta når dei skal samordna terminologirelevant materiale med terminologi i eigentleg meining i éin og same termdatabase. I medisinske klassifikasjonar kan t.d strengar av typen

Q37 Ganespalte med leppespalte
[Q37 Cleft palate with cleft lip]

og

Q37.2 Spaltet bløt gane med leppespalte, bilateral
[Q37.2 Cleft soft palate with cleft lip, bilateral]

brytast ned i termar som

ganespalte, *leppespalte* og *bilateral*, som har eigne oppslag i medisinske ordbøker (jf. ICD-10, 1996). Ved oppbygging av ein termbank er det derfor ikkje uproblematisk å handtera begge typane postar som 'terminologi' (jf Buan 1997:77).

Det finst sjølvsagt ulike måtar å handtera dette problemet på. Er målet å koplå slikt materiale opp mot ei eigentleg termsamling, kan det gjerast gjennom SGML-merking eller tilsvarande. Men er målet å utnytta materialet som kjelde for terminologi, må problemet med representativitet løysast.

7. Avslutning

Det "konseptet" som eg i denne artikkelen har referert til som "levande termbank", er framleis aktuelt og gyldig. Ikkje noko terminografisk produkt fyller funksjonen sin utan at det reflekterer ein oppdatert kunnskapsstatus innanfor det fagområdet eller dei fagområda som produktet dekkjer. Ved å knyta terminografisk arbeid til spesifikke prosjekt kan ein fanga inn noko av dynamikken i termvoksteren i samfunnet. Ein slik modell har likevel sine svake sider, først og fremst knytte til datainnsamling og representativitet.

På det tidspunktet då grunnmaterialet i NOT vart bygd opp, hadde ein dårleg tilgang til (halv)automatisk ekserpering og materiale i maskinleseleg form. Desse føresetnadene er i dag endra. Det er derfor rimeleg å venta at korpusbaserte metodar, eventuelt saman med statistisk analyse (jf. t.d. Brekke m.fl. 1996), vil gje ei viktig hjelp til å nøytralisera denne typen skeivheit – sjølv når tekstgrunlaget i utgangspunktet er kontekstuelt avgrensa.

Litteratur

- Andersen, Øivin/Johan Myking 1997: *Norsk terminologisk database – kontekst og validering*. Prosjektbeskriving, Norsk termbank.
- Andersen, Øivin 1997: NOT-prosjektet: Norsk terminologisk database – Struktur og applikasjoner. I Hjulstad, H. (red.): *Nordterm '97*. Oslo: Nordterm/RTT (= Nordterm 9), 61–65.
- Andersen, Øivin 1998: Argumentstruktur og nominalisering. *Lexico-Nordica*, dette nr.
- Brekke, Magnar/Johan Myking/Kurshid Ahmad 1996: Terminology and lesser-used living languages. A critique of the corpus-based approach. *TKE '96*, Frankfurt: Indeks Verlag.
- Buan, Bjørn 1997: Utvikling av en nasjonal medisinsk termbase. I Hjulstad, H. (red.): *Nordterm '97*. Oslo: Nordterm/RTT (= Nordterm 9), 77–82
- Haugen, Einar 1987: *Blessings of Babel: bilingualism and language planning: problems and pleasures*. Berlin : de Gruyter.
- Hirs, Willem 1993: ICD-10, A missed chance and a new opportunity for medical terminology standardization. I: *Proceedings from 3rd Infoterm Symposium – Terminology work in subject fields*. Wien: IITF, 134–156.
- Hjulstad, Håvard/Kjell Westerberg 1997: Nordisk termbank – Drøfting av en mulig nordisk termbank. I *Terminologi og kvalitet*. Reykjavík: Nordterm (= Nordterm 7), 103–108.
- ICD-10, 1996: *Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer. 10. revisjon. Norsk utgave. Systematisk del og alfabetisk indeks*. Oslo: Statens helsetilsyn/Elanders forlag.
- Interval, 1997: *Interlinguistic terminology validation*. (<http://www.mcs.surrey.ac.uk/interval/>).
- Maurais, Jacques 1993: Terminology and Language Planning. I Sonneveld, H. & Loening, K.L. *Terminology. Applications in interdisciplinary communication*. Amsterdam/Philadelphia: Benjamins, 111–126.
- Meurer, Paul 1997: Et verktøy for terminologiske databaser. I Hjulstad, H. (red.): *Nordterm '97*. Oslo: Nordterm/RTT (= Nordterm 9), 66–76.
- Myking, Johan/Meurer, Paul 1996: Norsk terminologisk database (NOT). Prosjektbeskriving. *Småskrifter*, 7/97. Norsk termbank.
- Myking, Johan 1997: Oljespråket – kva kom det ut av laboratoriet? *Småskrifter* 25/97. Norsk termbank.

- Omdal, Helge (red.) 1996: *Språknormering og språkbrukar*. Kristiansand: HiA.
- Picht, Heribert 1983: Terminologiens prinsipper og metoder. I *Språk i Norden*. Nordisk språksekretariat, 51–69
- POINTER, 1995: *Proposals for an operational Infrastructure for terminology in Europe*.
- Rangnes, Odd Kjetil 1997: Oljeterminologi og språkholdninger: bruk av spørreskjema i lys av dobbel hermeneutikk. *Småskrifter* 1/97. Norsk termbank.
- Roald, Jan/Johan Myking/Per-Bjørn Pedersen 1985: *Terminol. Slutt-rapport fra et terminologisk prosjekt*. Norsk termbank. (= Norske språkdata 11.)
- Westerberg, Kjell 1997: Nordterm-Net. Nordisk terminologiinformasjon samlad på ett ställe. I Hjulstad, H. (red.): *Nordterm '97*. Oslo: Nordterm/RTT (= Nordterm 9), 121–128.
- Wright, Sue-Ellen 1997: Terminology Management for Technical Translation. I Wright, S.-E. & Budin, G. (eds.): *Handbook of terminology management*. Amsterdam/Philadelphia: Benjamins, 147–159.