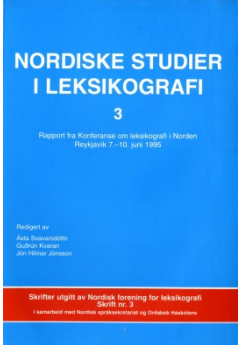


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Anvendelse af korpora i opbygningen af ordbøger til maskinoversættelsessystemet METAL	
Forfatter:	Lotte Weilgaard Christensen	
Kilde:	Nordiske Studier i Leksikografi 3, 1995, s. 421-431 Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi

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 Nordiske studier i leksikografi (1-5) 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.

Anvendelse af korpora i opbygningen af ordbøger til maskinoversættelsessystemet METAL

This paper presents some of the results that the Danish METAL group has achieved as to the production of lexicons for the MT system METAL, and especially our recommendations regarding the use of various documentation categories.

Existing MT lexicons from the German-English language pair, lists of terms for inhouse usage, as well as domain specific corpora form the basis of the lexicons. From experience we find that lexicons should mainly consist of words and terms extracted from technical corpora.

However, technical documentation often consists of words that should not be entered into lexicons; this might be due to terminological inconsistency, the formation of ad hoc compounds, etc. I close this paper by dealing with some of the lexical problems we had to cope with when translating a specific technical text.

1 Indledning

I denne artikel vil jeg gøre rede for, hvordan vi ved Handelshøjskole Syd, DK har opbygget ordbøger til maskinoversættelsessystemet METAL, og hvordan vi har bearbejdet de kilder, der har dannet grundlaget for vore maskinoversættelsesordbøger (herefter anvendes forkortelserne MO-system og MO-ordbog). Vi har forsøgt at tillempe terminologiske arbejdsmetoder og principper i ordbogsopbygningen til et morfo-syntaktisk MO-system. Nøgleordene i artiklen er således fagsprog, terminologi og maskinoversættelse.

2 Ordbøger i METAL-systemet

I den danske METAL-gruppe har vi udarbejdet en prototype for et tysk-dansk modul til MO-systemet METAL; det blev færdigt i 1994. Mine erfaringer bygger hovedsageligt på ordbogsarbejdet, hvor jeg bl.a. har arbejdet med en bilteknisk ordbog.

Ordbogsopbygningen i METAL, der er opbygget efter transfermodellen, er skitseret i følgende figur:

<i>Tysk monoordbog</i>	<i>Tysk — Dansk transferordbog</i>	<i>Dansk monoordbog</i>
Auto NST	Auto NST ⇒ bil NST	bil NST

Som det fremgår, har METAL to monoordbøger for vores sprogpår, en tysk og en dansk, og desuden én transferordbog, her tysk-dansk. Substantivet *Auto* vil således være indlæst i den tyske monoordbog, ligningen *Auto* ⇒ *bil* i transferordbogen og *bil* i den danske monoordbog.

Monoordbøgerne indeholder, som det ses i eksemplet *kobling*, primært oplysninger om et ords bøjninger:

Ensproget ordbog

TRÆK	VÆRDI	FORKLARING
(„kobling“	NST	indgangsord, ordklasse
ALO	„kobling“	stamme
CL	(P-ER S-EN)	bøjninger
GD	(C)	grammatisk køn
LINK	S	fugebogstav for komposita
SX	(N)	naturligt køn
TYN	(CNC)	semantisk træk
...)		

Transferordbøgerne indeholder ordets egentlige betydning i form af ordligninger, der som et minimum omfatter oplysninger om grundform og ordklasse for både udgangs- og målsprog samt fagområde. Det ses i ordbogsartiklen for *Kupplung*, der ligger i den biltekniske ordbog (ME-CAR):

Tosproget ordbog

„Kupplung“ NST ⇒ „kobling“ NST
 Pref S.0.0.00 Tag (ME-CAR)
 <<HHS Lw 29-Apr-91>>

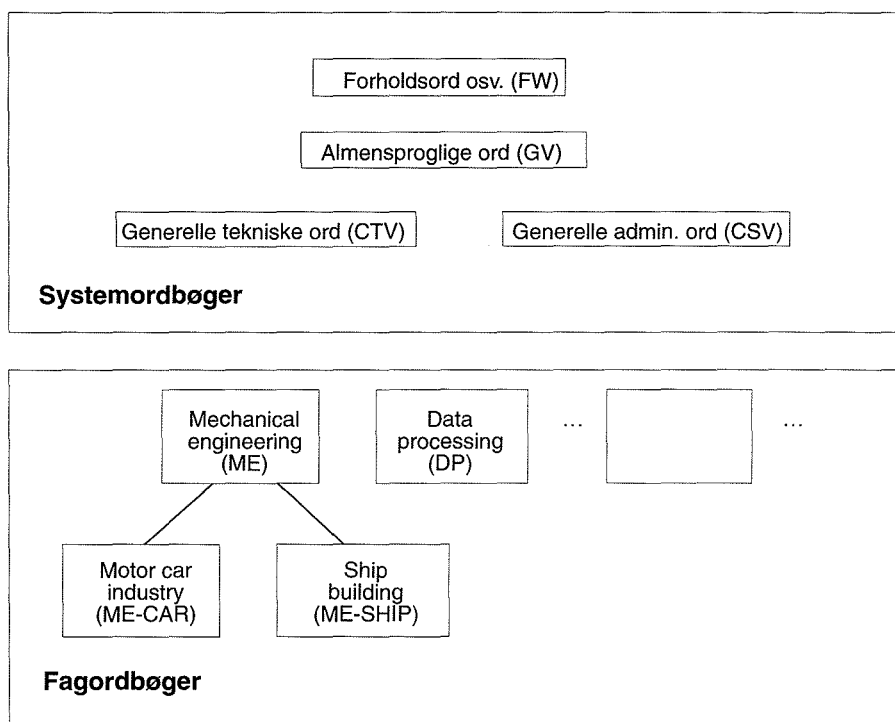
Transferordbøgerne i METAL er hierarkisk opbygget og består overordnet af en gruppe af systemordbøger, bl.a. en almensproglig ordbog og en generel teknisk ordbog, samt en gruppe af fagspecifikke ordbøger. Ved at inddele transferordbøgerne i fagspecifikke ordbøger opstår der en systematisk makrostruktur, der kan være mere eller mindre detaljeret.

3 Makrostruktur som hjælp til disambiguering

Hvis en term har flere betydninger, kan en human oversætter vha. definitioner og forklaringer finde frem til den rigtige oversættelse.

Den mulighed har maskinen ikke! Derfor må man i ordbøgerne anføre oplysninger, der er så entydige, at maskinen kan foretage det rigtige valg.

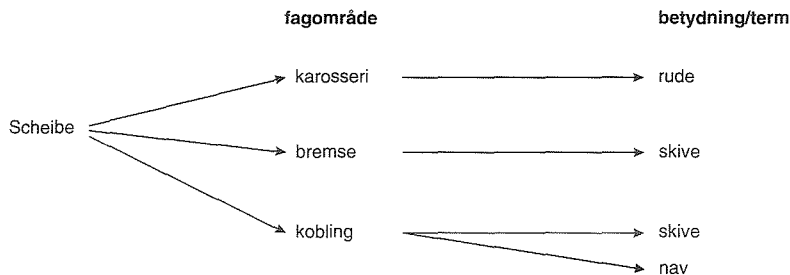
Angivelse af fagområde i transferartiklerne er én måde at hjælpe maskinen til at disambiguere korrekt. Det vil med andre ord sige, at mulighederne for disambiguering også afhænger af den valgte makrostruktur.



Transferordbøger i METAL

I den hierarkiske struktur i transferordbøgerne er **systemordbøgerne** med almene og almentekniske ord placeret øverst. Nedenunder findes **fagordbøger**; de kan være underinddelt, f.eks. er den maskintekniske ordbog inddelt i en ordbog for biler og én for skibe. Inden en tekst sendes til oversættelse, angives den mest specifikke af de ordbøger, systemet skal slå op i; for de biltekniske tekster er det naturligvis den biltekniske ordbog. Dér starter systemet sin søgning. Hvis den søgte term ikke findes i den biltekniske ordbog, bevæger systemet sig op på næste niveau, dvs. i den overordnede maskintekniske ordbog (ME), og så fremdeles.

Som den biltekniske ordbog ser ud i dag, vil *Scheibe* blive oversat med alternativerne *rude/skive/nav*:



Men som det ses af ovenstående figur, vil problemet til dels kunne løses ved at underinddele den biltekniske ordbog i flere mindre ordbøger med hver sit delområde, her karosseri,

bremse og kobling. Inden for delområdet kobling vil der dog stadigvæk være alternativerne *skive* og *nav*.

Kilderne og de fagområder, der behandles i dem, vil være bestemmende for, hvordan den endelige systematik i transferordbøgerne kommer til at se ud. Men grunden til, at vi opbygger ordbøgerne med en systematisk makrostruktur er den søgestrategi, METAL bygger på. Ud fra en terminologisk betragtning, kan man sige, at det er første skridt på vejen mod begrebsystemer.

4 Kilder anvendt i opbygningen af METAL-ordbøgerne

I dette kapitel skal jeg kort skitsere, hvordan vi har brugt vores kildemateriale. Vi havde nok en ide om, hvordan den ideelle vej så ud, men bl.a. fordi vi er en del af et internationalt projekt, har vi undertiden været nødt til at gå anderledes til værks, end det, der efter vores mening, var mest hensigtsmæssigt.

4.1 Systemordbøgerne

Til opbygningen af systemordbøgerne for det tysk-danske modul tog vi først udgangspunkt i de eksisterende METAL-systemordbøger fra det tysk-engelske sprogpar. — Det viste sig dog, at ordbøgerne efter vores opfattelse var for store og emnemæssigt for brede.

Vore erfaringer i dag peger i retning af, at man kan tilvejebringe systemordbøger mere målrettet ved at opbygge dem på baggrund af ordenes eller fagordenes frekvens i fagsproglige tekstkorpora end ved at overtage eksisterende systemordbøger fra andre sprogpar. **Korpusbaserede systemordbøger** vil komme til at indeholde ord, der optræder på samme måde i mange forskellige typer tekniske fagtekster, og som altid oversættes ens. Da vores mål for vores ordbøger er kun at optage de ord, der er relevante for maskinoversættelse af tekniske tekster, vil vi for det dansk-tyske modul afprøve en metode, hvor systemordbøgerne så godt som udelukkende er baseret på tekstkorpora. Vi håber på den måde at opnå en bedre oversættelses kvalitet og bedre hastigheder.

Vi har således fokuseret på indholdet i system- eller basisordbøgerne, og ikke som de fleste på deres størrelse.

4.2 Fagspecifikke ordbøger

Udviklingen af de fagspecifikke ordbøger har overvejende været korpusbaseret. I modsætning til systemordbøgerne har det lige fra starten været vores strategi, at det vigtigste grundlag for de fagspecifikke ordbøger var de **fagsproglige korpora**, vi har arbejdet med i forbindelse med systemudviklingen. Vi har dermed skullet tage højde for alle de termer og termstrukturer, der fandtes i vore korpora.

Det lyder ideelt og dermed burde lemmaselektionen eller termselektionen, som jeg vil kalde det, være klart afgrænset, men fagtekster indeholder ofte ord, som man ikke vil og heller ikke bør medtage i ordbøgerne. Det vil jeg komme tilbage til.

4.2.1 Kildemateriale til den biltekniske ordbog

For det tysk-danske modul har jeg primært arbejdet med terminologien inden for det biltekniske område.

Den biltekniske ordbog er den største af de fagspecifikke ordbøger, vi har opbygget. Den er tillige den af vores fagspecifikke ordbøger, der er baseret på det mest alsidige kildemateriale både mht. kildekategorier og kvalitet.

Grundlaget for den biltekniske ordbog har primært været tysk undervisningsmateriale for mekanikere og en reparationsvejledning samt sekundært en virksomhedsintern ordliste.

Undervisningsmaterialet er et parallelt korpus, der foreligger på tysk som udgangssprog og i dansk oversættelse. Da fagfolk har efterredigeret den danske oversættelse meget grundigt, har vi været i en ideel situation, hvor kildematerialet på begge sprog er formuleret præcist på såvel sætnings- som termniveau. Sproget i disse tekster opfylder kravene til fagsproglige tekster om entydighed, terminologisk konsistens og sprogøkonomi.

Kvaliteten i reparationsvejledningen, der kun forelå på tysk, er ikke på højde med undervisningsmaterialet. En sidste og meget vigtig kilde i oversættelsesarbejdet har været fagfolkene selv.

Fra reparationsvejledningen har vi oversat kapitler inden for forskellige delområder, som også er behandlet i undervisningsmaterialet, f.eks. motor, køler, udstødning, bremses. Ved at behandle flere mindre områder fra begge korpora, kunne vi opnå en god terminologisk afdækning af et antal klart afgrænsede delområder og samtidig en vis bredde i ordbogen.

5 Oversættelsesopgave fra BMW

Vi har for nylig lavet en oversættelsesopgave for BMW. På baggrund af den skal jeg nu prøve at eksemplificere det praktiske ordbogsarbejde i forbindelse med maskinoversættelse. Mit første indtryk var, at teksten fra BMW med sine korte, overskuelige sætninger ville være velegnet til maskinel oversættelse.

Et af emnerne på denne konference er, hvordan de enkelte kilder påvirker lemmaselektionen. Det skulle vise sig, at denne tekst er af en kvalitet, der ikke bør overføres til ordbøgerne, og langt fra den kvalitet, der specielt var kendetegnende for undervisningsmaterialet.

Nogle af de forhold, der har vanskeliggjort termselektionen er:

- at teksten er skrevet i **versaler**
- at den indeholder mange **inkonsekvente forkortelser**
- at teksten er præget af andre **inkonsekvente stavemåder**
- at teksten er præget af **synonymi**
- at der er anvendt mange **engelske låneord**
- at teksten er skrevet i **telegramstil**

Jeg skal i det følgende gennemgå de nævnte forhold nærmere, og se på, hvordan vi så vidt muligt har undgået at overføre disse „unoder“ til ordbøgerne. En stor del af dette arbejde består i at forbedre kildeteksten ved manuel eller automatisk preeditering.

Først er det dog vigtigt at slå fast, at nogle termer er bedre egnede til at kommunikere en teksts indhold videre end andre. I valget af termer følger vi de regler, man i terminologilæren har opstillet for termdannelse.

Til de vigtigste **termkrav** hører, at termer skal være motiverede, at de skal være afledningsproduktive og følge det almindelige sprogs syntaktiske regler.

Det har vist sig, at der kan opnås en bedre kvalitet i oversættelserne, hvis kravene følges. Nogle krav vil med maskinoversættelse endda kunne overholdes og opfyldes bedre og mere konsekvent, end humane oversættere har gjort det.

Det er især ved vores valg af de danske ækvivalenter disse termkrav har betydning, eftersom den tyske terminologi på forhånd er givet i teksten. Men i det omfang det er muligt, inddrages termkravene også for den tyske dels vedkommende.

5.1 Versaler

Teksten er skrevet i versaler. Vi ønsker naturligvis ikke at optage termerne i ordbøgerne i denne form. METAL kan ikke uden videre håndtere hele tekster, der er skrevet i versaler. Men systemet behandler enkeltord i versaler som konstanter, der ikke skal oversættes. Det vil oftest være firma- eller produktnavne, der netop skal overtages direkte i oversættelsen.

Vi konverterede først dokumentet til tekstbehandlingsprogrammet Word. I Word er der mulighed for automatisk at ændre versaler til små bogstaver. Da de fleste ord var substantiver, udnyttede vi en facilitet, der ændrer versalerne til små bogstaver på nær det første bogstav, som i overensstemmelse med den tyske retskrivning for substantiver skrives med stort.

5.2 Inkonsekvente forkortelser

Forfatteren har af edb-tekniske pladshensyn anvendt flere inkonsekvente forkortelser. Eksempler på disse forkortelser er:

- (1) *Bremsflüssigkeitsbeh.* for 'Bremsflüssigkeitsbehälter'
- (2) *Zylinder Sch.Bügel* for 'Zylinderschutzbügel'
- (3) *Z.-Vert.* for 'Zündverteiler'
- (4) *Geri.* for 'gerissen' i konteksten *Naht Aufgeg., Geri.*

Som det ses, er der stor forskel i gennemskueligheden af disse forkortelser. Ofte hjalp det mig, at der var tale om en terminologi, jeg før havde beskæftiget mig med. Eksempel (1) *Bremsflüssigkeitsbehälter* hører til de mere gennemskuelige. Eksemplerne (2)–(3) kunne jeg finde i deres fulde form andre steder i teksten ved at søge på dele af forkortelserne og sammenligne deres kontekst, andre kunne jeg gætte mig til ud fra konteksten. Eksempel (4) *Geri* med stort G voldte en del besvær, men også her hjalp konteksten *Naht Aufgeg., Geri* mig til sidst. Det store G var en rest fra versalerne i det oprindelige dokument. Da *gerissen* forekom over 300 gange skrevet ud, var det ikke umiddelbart indlysende, at der var tale om en forkortelse for *gerissen*.

5.3 Inkonsekvente stavemåder

I teksten er der flere inkonsekvente stavemåder i forbindelse med binde- eller fugestreg i komposita, og der er eksempler på anvendelse af flere mulige ortografiske varianter.

Mange tyske **komposita** ses skrevet både med og uden bindestreg. For oversættelses-systemet opfattes disse som to forskellige termer, ligesom det også vil være tilfældet ved forskellige fugebogstaver. Da vi ikke ønsker at have redundans i vore ordbøger, har vi besluttet os for en strategi, hvor vi principielt kun indlæser formerne uden bindestreg.

- (5) a *Tankentlüftungs-Ventil* (4)
- b *Tankentlüftungsventil* (1)
- (6) a *Kupplungs-Seite* (1)
- b *Kupplungsseite* (1)
- c *Kuppl.-Seite* (1)

Tankentlüftung i eksempel (5)a og (5)b optræder fire gange med bindestreg og en gang uden. Men det er mit generelle indtryk, at anvendelsen af bindestreger er ret tilfældig.

Termen *Kupplungsseite* i eksempel (6) forekommer både med og uden bindestreg samt i forkortet form; de er alle hver især realiseret en gang. Denne term er blot et af mange eksempler på, at flere af de nævnte forhold kan optræde side om side.

- (7) a *Blasbildung* (11)
- b *Blasenbildung* (40)

I eksempel (7) *Blasbildung/Blasenbildung* er termerne dannet med og uden fugebogstaver. *Blasbildung* optræder i alt 11 gange i teksten, mens *Blasenbildung* optræder 40 gange; noget taler altså for at vælge denne form. Frekvensoplysningerne fremgår af nogle for-analyselister, som METAL genererer, og som er udgangspunkt for leksikonarbejdet.

5.3.1 Metal Pattern Matcher (MPM)

Til METAL-systemet hører forskellige hjælpeværktøjer. Et af disse hedder Metal Pattern Matcher; det er et **mønstergenkendelses- og batch-hjælpeprogram**, der automatisk kan redigere en tekst enten før eller efter oversættelse.

Dette hjælpeværktøj kan i kildefilen erstatte de stavemåder, som vi ikke ønsker at optage i ordbøgerne, med den rigtige stavemåde, dvs. den stavemåde vi ønsker at optage i ordbøgerne.

Det anføres i en fil, at *Blasbildung* — fra eksempel (7)a — hver gang det forekommer, skal læses som (7)b *Blasenbildung*, der er den term, der er indlæst i ordbøgerne.

5.4 Synonymer

Teksten er som nævnt præget af synonymer.

- (8) a *Anhängermodul*
- b *Anhängemodul*
- (9) a *Antriebseinheit*
- b *Antriebs-Einheit*
- c *Antriebsteil*

Nogle af disse synonymer har vi valgt at behandle ved hjælp af Metal Pattern Matcher. Det er synonymer af typen som vist i eksempel (8)a *Anhängermodul* og (8)b *Anhängemodul*. Der er tale om en slags ortografiske varianter.

Synonymer som i eksempel (9) *Antriebseinheit/Antriebsteil* optages som selvstændige artikler i vores ordbøger, da de som i eksemplet kan have forskelligt køn og dermed forskellig bøjning af artikler og adjektiver mv. Det kan der ikke tages højde for i Metal Pattern Matcher.

5.5 Engelske låneord

I BMW-teksten er der anvendt en del engelske låneord som *Airbag*, *Kickdown*, *Memory*. Min umiddelbare holdning var, at jeg måtte undgå at optage disse anglicismer i ordbøgerne. Det skyldtes bl.a., at de engelske låneord også skulle indlægges i den tyske monoordbog, som den tyske METAL-gruppe har ansvaret for. Det var efter min opfattelse den tyske gruppe, der skulle afgøre, om den kritikløst ville overtage de engelske låneord.

På det tidspunkt stødte jeg på en artikel om engelske låneord i tyske bildokumenter af Sverre Vesterhus (1991). I sin undersøgelse kommer Vesterhus frem til, at de tyske bilmærker alle ligger under den gennemsnitlige frekvens for engelske låneord på 0,85 pr. side, dog med en undtagelse, nemlig BMW.

Gennemsnitlig frekvens	0,85
BMW	1,07
Ford	0,77
Opel	0,74
Audi	0,67
Porsche	0,41
Volkswagen	0,41
Mercedes-Benz	0,39

(Vesterhus 1991)

For de tyske bilmærker tegnede BMW sig med 1,07 for den højeste frekvens for engelske låneord, mens Mercedes-Benz lå lavest med 0,39. — På baggrund af disse tal er jeg efter min mening nødt til at acceptere de engelske låneord og optage i hvert fald nogle af dem i ordbøgerne. Det er interessant, at de engelske låneord ikke kun optræder i én teksttype, nemlig reklamematerialet, men også i BMWs mere interne fagdokumentation, hvorfra låneordene sandsynligvis endda stammer.

Mange engelske låneord findes endnu ikke i tyske ordbøger, og for flere mangler der endnu et bøjningsmønster, fordi lånene er relativt nye. I maskinoversættelse skaber det problemer mht. de grammatiske oplysninger, som skal anføres i den tyske monoordbog aht. sætningsanalysen.

Og i BMW-teksten er der anvendt en telegramstil, som bevirker, at det oftest ikke er muligt at udlede køn eller bøjningsformer.

Nogle af de engelske låneord har jeg optaget i ordbøgerne, mens jeg har valgt at behandle andre med Metal Pattern Matcher. Jeg har i det følgende grupperet nogle af låneordene efter, hvordan jeg har behandlet dem.

Da *Recall Campaign* i eksempel (10) også forekommer som oversættelseslånet *Rückrufaktion* har jeg i Metal Pattern Matcher erstattet *Recall Campaign* med *Rückrufaktion*.

Recall Campaign bliver således slet ikke optaget i ordbøgerne, men registreret i en MPM-fil.

(10) *Recall Campaign* → *Rückrufaktion*

Metal Pattern Matcher indeholder også en anden del, hvor man i en fil kan anføre ord, som systemet skal markere som konstanter. På den måde ved systemet, at det ikke skal forsøge at oversætte ordene. Det vil som nævnt meget ofte være produktnavne. I eksemplerne (11)–(12) har den fagmand, jeg har arbejdet sammen med, opgivet, at *Park Distance Control* og *Interlock* skal overtages til dansk.

(11) *Park Distance Control* → *Park Distance Control*

(12) *Interlock* → *interlock*

Da jeg ikke umiddelbart har kunnet finde de nødvendige grammatiske oplysninger, har jeg valgt at behandle disse låneord i Metal Pattern Matcher, indtil jeg har fundet de manglende grammatiske oplysninger.

(13) *Fluidbloc* → *hydroleje*

(14) *Silentblock* → *gummibøsning*

Eksemplerne (13) *Fluidbloc* og (14) *Silentblock* har voldt mig de største problemer. Da disse låneord ikke overtages direkte til dansk, men har danske ækvivalenter, er jeg tvunget til at optage dem i ordbøgerne. *Fluidbloc* oversættes til *hydroleje* og *Silentblock* til *gummibøsning*. Jeg har valgt at kode dem med det køn og bøjningsmønster, jeg anså for mest sandsynligt, og har så i et kommentarfelt i ordbogsartiklerne anført, at kodningen senere skal efterredigeres. På grund af telemgramstilen er artikler ofte udeladt i teksten, derfor vil en forkert ordbogsoplysning om artikelbrug ikke komme til at genere sætningsanalysen i denne oversættelse.

5.6 Telegramstil

I BMW-teksten forekommer der en form for reduceret artikelbrug, der skaber en slags **pseudo-komposita** eller pseudo-termer, som vi ikke ønsker at optage i ordbøgerne. Nogle eksempler er *Frontverkleidung-Ziergitter* og *Korrektur-Buchung*:

(15) a *Frontverkleidung-Ziergitter*

b *Frontverkleidung des Ziergitters*

c *Frontverkleidung am Ziergitter*

(16) a *Korrektur-Buchung*

b *Buchung der Korrektur*

De fleste eksempler er reducerede genitivkonstruktioner som *Frontverkleidung-Ziergitter*, der i sin komplette form er *Frontverkleidung des Ziergitters* eller *Frontverkleidung am Ziergitter*. Genitiv-artiklen eller præpositionen 'an + artikel' erstattes af en bindestreg. Derved bliver der dannet nogle konstruktioner, der minder om komposita og derved får status som selvstændige termer uden at være det.

Eftersom der ikke er tale om regulære termer, bør man så vidt muligt undgå at optage dem i ordbøgerne. I det omfang det er muligt, vil jeg altså lade dem oversætte kompositionelt, hvor systemet oversætter *Frontverkleidung* for sig og *Ziergitter* for sig og derefter sammenføjer dem til et kompositum på målsproget.

Men er der tale om udtryk, hvor strukturen er forskellig på udgangs- og målsprog, er man tvunget til at optage dem i ordbøgerne for at kunne foretage de nødvendige strukturelle ændringer.

6 Konklusion

Maskinen har ingen viden om verden og derfor ikke samme muligheder for at vurdere en ordbogsartikel som en person. Derfor er det uhyre vigtigt, at de data, der indlæses i ordbogsartiklerne i et MO-system, er korrekte og præcise.

Desuden er vi i MO-ordbøgerne nødt til at tage højde for hver eneste af de ord og termer, der findes i en tekst, der skal maskinoversættes. Opbygningen af vores systemordbøger vil fremover ganske vist være korpusbaseret, men som i traditionelt leksikografarbejde kan vi så at sige springe et udtryk over. Men i forbindelse med konkrete maskinelle oversættelsesopgaver er der ingen vej udenom. Der er man nødt til leksikalsk at bearbejde korpus ned til mindste detalje. Vores erfaringer med fagtekster er, at de ofte lader noget tilbage at ønske. BMW-teksten er et af de værre eksempler, men ikke enkeltstående.

Endelig har vores arbejde med MO-oversættelse vist, at det i meget høj grad er ordbøgerne, der bestemmer, hvor god en maskinoversættelse bliver.

Mine eksempler fra BMW-teksten viser, at det er vigtigt at behandle teksten og korpora — i det hele taget — systematisk og lægge strategier for, hvordan forskellige problemer skal håndteres. — Det nævnte foranalyseprogram er én hjælp til at behandle tekster systematisk, men er langt fra tilstrækkeligt, f.eks. giver det ikke nogen oplysninger om semantiske relationer, hvilket er vigtigt for at kunne opbygge en konsistent terminologi i det mindste for målsprogets vedkommende og for at kunne placere termerne i de rigtige fagspecifikke ordbøger, hvilket har betydning for makrostrukturen.

Den systematiske gennemgang af BMW-teksten har vist, at det var nødvendigt med ganske meget pre-editering, hvis ordbøgerne bagefter skulle have en kvalitet, der var til at leve med. For at kunne oversætte teksten er den oprindelige tekst blevet en hel del forbedret, enten manuelt eller automatisk vha. Metal Pattern Matcher. Det vil med andre ord sige, at ordbogsarbejdet har været med til at forbedre kvaliteten i kildeteksten. Når der er tale om maskinoversættelse, kan man derfor også vende problematikken om og spørge: Hvordan påvirker ordbogsarbejdet kildematerialet? Men man kan og skal selvfølgelig ikke omskrive hele teksten; det er klart! Hvis det er nødvendigt, skal en tekst ikke maskinoversættes.

Bibliografi

- Vesterhus, Sverre A. 1991: Anglicisms in German car documents. I: *Language International* 3.3, 10–15.
- Weilgaard Christensen, Lotte 1992: Ordbøger i METAL. I: Henrik Holmboe (ed.): *Skriften På Skærmen*; Datalingvistisk Forening, Årsmøde nr. 2, Handelshøjskole Syd, 13.–14. maj 1992. Århus: Afdeling for Datalingvistik, Handelshøjskolen i Århus, 31–41.
- Weilgaard Christensen, Lotte 1993: Ordbøger i METAL — Anvendelse af terminologiske principper til opbygning af ordbøger i METAL. I: Bertha Toft (ed.): *Terminologi, edb & videnssteknik*, 2. nordisk symposium 1992, Varde. Kolding, 202–221.

- Weilgaard Christensen, Lotte 1994: Termvalg i fagtekster og maskinoversættelsesordbøger. I: *Jubilæumsskrift for HHS*. Kolding, 135–149.
- Weilgaard Christensen, Lotte/Ellen Christoffersen 1993: Oversættelse med METAL. I: *Sprog & Erhverv, Erhvervsprogligt Forbund 7–8/93*, 13–15.
- Weilgaard Christensen, Lotte/Ellen Christoffersen/Gert Engel 1991: *Aufbau und Wartung der METAL-Wörterbücher* (internt papir).
- Weilgaard Christensen, Lotte/Ellen Christoffersen/ Margrethe M. Møller 1991: Ordbøger i METAL. I: *Handelshøjskole Syd Forskning, Temanummer 3 af HHS-Information om erhvervsproglig forskning ved HHS*. Kolding, 6–7
- Weilgaard Christensen, Lotte/Ellen Christoffersen/ Margrethe M. Møller 1994: Metoder og strategier i et maskinoversættelsesprojekt. I: *Festskrift for Gert Engel*. Kolding, 233–247.