

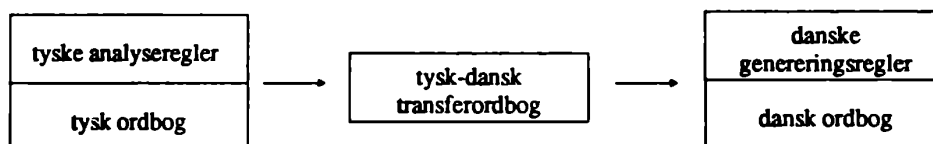
## ***Maskinoversættelse af tyske NPer***

***Ellen Christoffersen og Margrethe H. Møller***  
***Handelshøjskole Syd, Kolding***

Ved Institut for Erhvervsforskning i Kolding har en projektgruppe fra Handelshøjskole Syd siden sommeren 1989 arbejdet med udviklingen af det tysk-danske modul til maskinoversættelsessystemet METAL.

METAL er et af de "gamle" maskinoversættelsessystemer i den forstand, at man har arbejdet på udviklingen af det i 20-30 år. Først kun i USA, ved University of Texas, og siden midten af firserne også i Europa, idet den tyske Siemens-koncern købte systemet. I dag er der METAL-arbejdsgrupper i USA, Tyskland, Belgien, Spanien og Danmark. Siemens sælger systemet med sprogparrene tysk-engelsk, engelsk-tysk og tysk-spansk, og andre sprogpar er under udvikling. I den danske arbejdsgruppe udvikler vi moduler med dansk, i første omgang det tysk-danske genereringsmodul.

METAL arbejder efter transfermodellen:



Den danske generering bygger på et analyseresultat, der dels er fremkommet vha. morfologisk og syntaktisk analyse af den tyske tekst, dels omfatter de semantiske oplysninger, de enkelte leksemer er forsynet med i den tyske ordbog.

METAL skal først og fremmest oversætte fagsproglige tekster. I det lingvistiske arbejde har den danske arbejdsgruppe derfor taget udgangspunkt i fagsproglige tekster. Vi ønskede at begynde vort arbejde med enkle tekniske tekster som f.eks. styklister og indholdsfortegnelser til brugervejledninger, og derfor var det også naturligt for os i første projektfase at begynde med at oversætte NPer.

Da vi kunne oversætte indholdsfortegnelser, satte vi en foreløbig slutstreg under NP-genereringen og gik over til at oversætte sætninger.

I dette indlæg vil vi beskæftige os med nogle principielle problemer i maskinoversættelse, eksemplificeret ved to problemer, vi ofte er stødt på i vore fagsproglige NPer. Det drejer sig dels om et *leksikaliseringsproblem*. Hvilke oplysninger skal være tilgængelige i ordbøgerne og hvilke oplysninger kan man beregne sig til? – og dels om et *strukturproblem*: En struktur på udgangssproget kræver somme tider en strukturændring på målsproget, og i andre tilfælde kan den gengives ved samme struktur på målsproget. Her må man søge at finde lingvistiske variable og værdier, der både kan beregnes og også er tilstrækkelige til at træffe valget mellem de to muligheder.

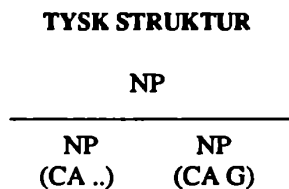
## ***Et strukturproblem: Tyske genitivkonstruktioner***

(Ellen Christoffersen)

Først vil jeg gøre rede for det problem, der opstår, når en bestemt struktur i udgangssproget modsvares af flere forskellige strukturer i målsproget. Et sådant problem møder vi, når der i den tyske tekst optræder et nominalsyntaxme som attribut til et andet nominalsyntaxme. Det attributive nominalsyntaxme står i genitiv, mens det overordnede nominalsyntaxmes kasus er bestemt af dets syntaktiske status i sætningen.

### ***Tysk struktur***

Det tyske analyseresultat af denne struktur ser ud som vist på figur 1.



*Figur 1*

Denne struktur finder vi i følgende eksempler hentet fra de fagsproglige tekster, vi har arbejdet med i forbindelse med oversættelse af tyske nominalsyntaxmer til dansk:

1. Eine Anzahl mehrsprachiger Lexikoneinträge  
Et antal flersprogede leksikonartikler
2. Die Vorschriften des Herstellers  
Producentens forskrifter
3. Der Austausch nichttragender Teile  
Udvekslingen af ikke-bærende dele
4. Die Verwindungsstabilität des Fahrzeuges und die Sicherheit der Insassen  
Køretøjets vridningsstabilitet og passagerernes sikkerhed
5. Folgende Verformungen des Rahmenbodenverzuges  
Følgende deformationer af rammebunden
6. Bremsdruckkreise der Antriebsachse  
Trækakselens bremsekredse
7. Ausbeulen leichter Blechschäden  
Opretning af lette pladeskader
8. Vor Beginn der Reparaturarbeiten  
Inden reparationsarbejdernes begyndelse
9. Ein Dutzend aktueller Probleme  
Et dusin aktuelle problemer
10. Rahmenkontrollmaße und Blechstärken der BMW-Modellreihen  
BMW-modelrækkernes rammekontrolmål og pladetykkelser

11. Der Anfang des Benutzerhandbuches  
Brugervejledningens begyndelse
12. Eine Reihe der Übersetzungsvorschläge  
En række af oversættelsesforslagene

Forsøger man at gruppere eksemplerne efter de kriterier, der anvendes i de gængse grammatikker, opstår følgende fire grupper:

**Objektiv genitiv:**

Som objektiv genitiv betegnes det "logiske objekt" for den handling, der er udtrykt i det verbalsubstantiv, som er kerne i den overordnede NP.

1. Ausbeulen leichter Blechschäden  
Opretning af lette pladeskader
2. Der Austausch nichttragender Teile  
Udvekslingen af ikke-bærende dele
3. Folgende Verformungen des Rahmenbodenverzuges  
Følgende deformationer af rammebunden

**Possessiv genitiv:**

Den possessive genitiv udtrykker et tilhørsforhold eller et direkte ejendomsforhold.

4. Die Verwindungsstabilität des Fahrzeuges und die Sicherheit der Insassen  
Køretøjets vridningsstabilitet og passagerernes sikkerhed
5. Rahmenkontrollmaße und Blechstärken der BMW-Modellreihen  
BMW-modelrækkernes rammekontrolmål og pladetykkelser
6. Bremsdruckkreise der Antriebsachse  
Trækakselens bremsekredse

**Subjektiv genitiv:**

Som subjektiv genitiv betegnes det "logiske subjekt" for den handling, der er udtrykt i det verbalsubstantiv, som er kerne i den overordnede NP.

7. Vor Beginn der Reparaturarbeiten  
Inden reparationsarbejdernes begyndelse
8. Der Anfang des Benutzerhandbuches  
Brugervejledningens begyndelse
9. Die Vorschriften des Herstellers  
Producentens forskrifter

**Paritiv genitiv:**

Den partitive genitiv er et genitivled, som betegner en mængde, der består af ensartede dele, eller en ensartet masse, medens den overordnede NP betegner en del af denne mængde eller masse.

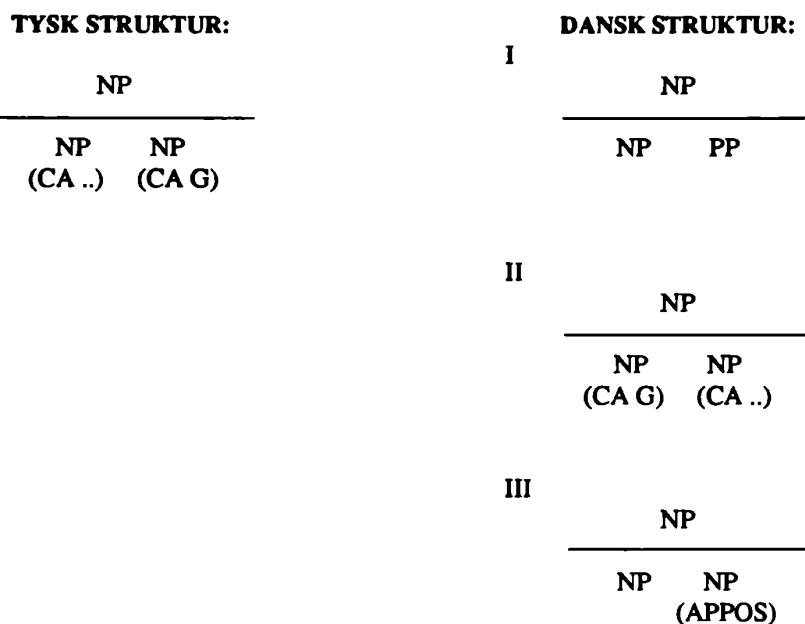
10. Eine Anzahl mehrsprachiger Lexikoneinträge  
Et antal flersprogede leksikonartikler
11. Ein Dutzend aktueller Probleme  
Et dusin aktuelle problemer

12. Eine Reihe der besten Übersetzungsvorschläge  
 En række af de bedste oversættelsesforslag

I de tyske eksempler er der strukturel overensstemmelse, men semantisk er der tale om en stor variationsbredde.

**Danske strukturer**

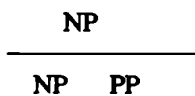
Til ovennævnte tyske struktur svarer tre strukturer på dansk, der ser ud som vist i figur 2. Den af strukturerne, der forekommer hyppigst i de fagsproglige tekster, vi har arbejdet med, er nævnt først, og den mindst hyppige sidst.



Figur 2

**Dansk struktur I:**

Det tyske attribut i genitiv skal på dansk erstattes af en præpositionsforbindelse. Denne struktur anvendes på dansk ved transfer af objektive genitiver og visse partiive genitiver.

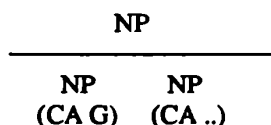


- |   |                    |
|---|--------------------|
| 1. En række af oversættelsesforslagene<br>Eine Reihe der Übersetzungsvorschläge | (partiiv genitiv)  |
| 2. Opretning af lette pladeskader<br>Ausbeulen leichter Blechschäden            | (objektiv genitiv) |
| 3. Udvekslingen af ikke-bærende dele<br>Der Austausch nichttragender Teile      | (objektiv genitiv) |

4. Følgende deformationer af rammebunden (objektiv genitiv)  
 Folgende Deformationen des Rahmenboden-verzuges

**Dansk struktur II:**

Det tyske attribut i genitiv skal på dansk være en NP i genitiv, der står foran den overordnede NP. Denne struktur anvendes på dansk ved transfer af subjektive og possessive genitiver.

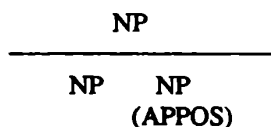


5. Inden reparationsarbejdernes begyndelse (subjektiv genitiv)  
 Vor Beginn der Reparaturarbeiten
6. Brugervejledningens begyndelse (subjektiv genitiv)  
 Der Anfang des Benutzerhandbuchs
7. Producentens forskrifter (subjektiv genitiv)  
 Die Vorschriften des Herstellers
8. Køretøjets vridningsstabilitet og passagerernes sikkerhed (possessiv genitiv)  
 Die Verwindungsstabilität des Fahrzeuges und die Sicherheit der Insassen
9. BMW-modelrækkernes rammekontrolmål og pladetykkelser (possessiv genitiv)  
 Die Rahmenkontrollmaße und Blechstärken der BMW-Modellreihen
10. Trækakselens bremsekredse (possessiv genitiv)  
 Bremsdruckkreise der Antriebsachse

**Dansk struktur III:**

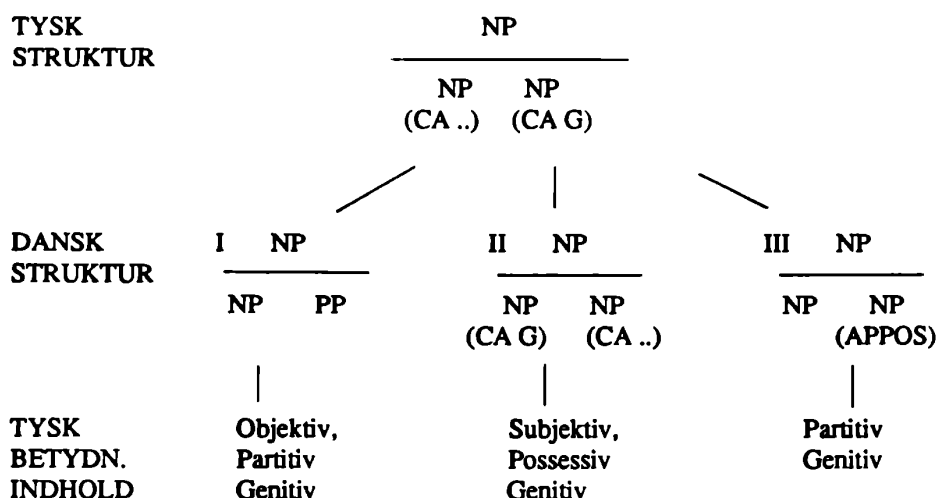
Det tyske attribut i genitiv skal på dansk være en apposition med samme syntaktiske status som den overordnede NP.

Denne struktur anvendes på dansk ved transfer af visse partitive genitiver, nærmere betegnet artens genitiv.



11. Et antal flersprogede leksikonartikler (partitiv genitiv)  
 Eine Anzahl mehrsprachiger Lexikoneinträge
12. Et dusin aktuelle problemer (partitiv genitiv)  
 Ein Dutzend aktueller Probleme

Et betydningsindhold på tysk behøver ikke at være knyttet til en bestemt struktur på dansk. Således omfatter både dansk struktur I og III visse partitive genitiver (se figur 3).



Figur 3

Inden for den ene tyske struktur kan vi altså konstatere en semantisk differentiering i 4 betydningsvarianter, nemlig objektiv, subjektiv, possessiv og partitiv genitiv. Dette forhold er imidlertid kun en begrænset hjælp til den strukturelle disambiguering, der skal foretages fra tysk til dansk.

### *Hvordan kommer vi frem til den rigtige danske struktur?*

Det ville være ideelt, om de differentieringskriter, som vi har brug for, blev leveret af den tyske analyse. Dette er imidlertid ikke tilfældet. For tiden analyseres en tysk genitivkonstruktion som en NP, der består af to NPer, hvoraf den sidste står i genitiv (se figur 1). Det er der en historisk forklaring på, idet det første tyske analysemodul blev udviklet til sprogparrat tysk-engelsk, hvor det ikke har været nødvendigt at levere disse informationer. Der er derfor behov for, at den tyske gruppe udvikler analysen på dette punkt.

Da vi ikke har kunnet vente på at få de manglende oplysninger leveret af den tyske analyse, har vi selv måttet finde løsninger for at kunne danne den rigtige struktur på dansk.

Vi har derfor udviklet en procedure, der beregner, under hvilke betingelser en tysk genitiv skal gengives med den danske struktur I, II og III:

- |     |                                 |                                  |
|-----|---------------------------------|----------------------------------|
| I   | tysk genitiv - dansk PP         | (objektiv og partitiv genitiv)   |
| II  | tysk genitiv - dansk s-genitiv  | (possessiv og subjektiv genitiv) |
| III | tysk genitiv - dansk apposition | (partitiv genitiv)               |

Selve proceduren bearbejder genitiverne i den omvendte rækkefølge, således at den mest specielle og lettest identificerbare struktur behandles først.

#### *Tysk genitiv → dansk apposition*

Denne danske struktur skal anvendes, når den overordnede tyske NP angiver et mål, en mængde o.l. (substantiver som *Anzahl, Meter, Kiste, Glas*), og når den overordnede NP er uden determiner.

Den praktiske løsning har bestået i, at vi i den tysk-danske ordbog har forsynet de tilsvarende danske substantiver (*antal, meter, kasse, glas*) med et træk, der viser, at det tyske genitivattribut modsvares af en dansk apposition (trækket *Marker of Genitive* med værdien *APPOSition*).

Tysk genitiv → dansk s-genitiv

De tyske genitiver, der på dansk skal gengives som s-genitiv, kan identificeres ved hjælp af tre betingelseskomplekser.

Det første betingelseskompleks anvender som kriterium, om der på tysk som determiner for overleddet anvendes enten den bestemte artikel eller 0-artikel.

Her er der anvendt et slags "kollokationkriterium".

I andet betingelseskompleks anvendes som hovedkriterium den regel, at alle danske subjektive genitiver optræder som s-genitiv. Opgaven består så i at identificere den overordnede tyske NP som subjektiv genitiv. Det er den, hvis kernen er et "intransitivt" verbalsubstantiv, eller kernen er et "transitivt" verbalsubstantiv og underleddet kan fungere som agens, dvs. i den tyske ensprogede ordbog er forsynet med det semantiske træk HUMAN eller POTent.

Den praktiske løsning består i, at vi slår op i den danske ordbog for at kunne hente oplysningen om, hvorvidt den danske ækvivalent til kernen i den overordnede tyske NP er et verbalsubstantiv, og for at få oplysning om dets transitivitet.

Er det danske substantiv et verbalsubstantiv, har vi for at kunne hente denne oplysning forsynet det med trækket DEVERBAL. Kan verbalsubstantivet forbindes med en objektiv genitiv er det i ordbogen forsynet med trækket TRANSitivity med værdien True.

I tredje betingelseskompleks identificerer vi en possessiv genitiv ved udelukkelsesmetoden dvs. ved at konstatere, at det overordnede substantiv ikke er forsynet med trækket DEVERBAL i den danske ordbog og ikke har fået tildelt et træk i den tysk-danske ordbog, der viser, at det tyske genitivattribut skal oversættes ved en dansk præpositionsforbindelse.

Tysk genitiv → dansk PP

De resterende tyske genitiver (objektiv, partitiv genitiv) oversættes ved danske præpositionsforbindelser. Den danske præposition er enten den, der måtte være angivet på den overordnede NPs kemesubstantiv ved trækket Marker og Genitive, ellers er det præpositionen "af".

### ***Et leksikalsk problem: Oversættelse af tyske komposita***

(Margrethe H. Møller)

I et maskinoversættelsessystem er der principielt to måder at oversætte komposita på: man kan enten leksikalisere dem eller oversætte dem kompositionelt.

Ved leksikalisering optager man kompositumet i ordbøgerne som ét ord. Ved kompositionel oversættelse optager man de led, kompositumet er sammensat af, i ordbogen – medmindre de findes der allerede – og lader systemet oversætte dem ét for ét og sammenføje dem til et kompositum på målsproget.

Sammensætning er en meget anvendt orddannelsesmekanisme i tysk fagsprog, og man kan ikke leksikalisere alle komposita. I praksis vil man oversætte kompositionelt, når det overhovedet kan lade sig gøre, og kun ty til leksikalisering, når der ikke er andre muligheder. Det sparer kodningsarbejde, idet leksikalisering indebærer, at der oprettes artikler i alle tre ordbøger, og det giver mulighed for at behandle sammensætninger, der ikke har været registreret tidligere.

Det, jeg gerne vil vise, er, hvornår der ikke er andre muligheder end leksikalisering, og hvilke ulemper leksikaliseringen på den anden side har.

Først og fremmest er kompositionel oversættelse kun mulig, når det tyske og det danske kompositum har samme struktur. Hvis det tyske kompositum skal oversættes ved et dansk simpleksord eller et flerleddet udtryk, som vi har mange eksempler på i de fagtekster, vi beskæftiger os med, er vi nødt til at leksikalisere kompositumet.

Vi skelner mellem kontekstuaafhængig og kontekstafhængig kompositionel oversættelse.

- (1) Serien | produktion – serie | produktion
- (2) Dienst | programm – hjælpe | program

I eksempel (1) kan begge kompositumets led oversættes direkte, uafhængigt af konteksten. Det tyske *Serie* oversættes til dansk *serie*, og det tyske *Produktion* oversættes til dansk *produktion*. I eksempel (2) er oversættelsen af førsteleddet kontekstafhængig: normalt oversættes det tyske substantiv *Dienst* ved det danske substantiv *tjeneste*, men når det står sammen med *Programm*, oversættes det ved verbet *hjælpe*. Dette kan angives i transferordbogen.

Danske komposita kan have fugebogstaver, og som hovedregel gælder det, at et givet ord altid anvender samme fugebogstav, når det indgår som første sammensætningsled i et toleddet kompositum. Det danske substantiv *ændring* anvender f.eks. fugebogstavet *s* som i *ændringsforslag*, *ændringsønske*. Vi angiver, hvilket fugebogstav der skal anvendes, i den danske ordbog.

Nogle ord anvender desuden en speciel allomorf, når de indgår i et kompositum. Det danske substantiv *maskine* anvender f.eks. allomorfen *maskin* som i *maskinfejl*. Også dette angiver vi i den danske ordbog.

Der kan være disambigueringsproblemer, f.eks. når analysen vælger substantivet *die Steuer* i stedet for verbet *steuern* i det tyske kompositum *Steuerzeile*, så vi får oversættelsen *\*skattelinie* i stedet for den rigtige styrelinie. Her er den eneste løsning leksikalisering af det tyske kompositum og dets danske oversættelse.

Men bortset fra det fungerer kompositionel oversættelse fint, når der er tale om toleddede komposita. Det bliver straks vanskeligere, når kompositumet har mere end to led. For det første er det da ikke muligt i den danske generering at beregne, hvor der skal indsættes fugebogstaver, for det andet kan vi ikke stole på at få den rigtige oversættelse af et led, hvis dette leds oversættelse er kontekstafhængig. Begge dele skyldes, at systemet ikke er i stand til at analysere sig frem til kompositumets interne struktur.

I eksemplerne:

- (3) [[Wörter | buch]eintrag]  
ordbogsartikel
- (4) [Fach[buch | handlung]]  
specialbog\_handel

ser vi, at substantivet *bog* skal have fuge-*s*, når det optræder som andet sammensætningsled som i *ordbog* i eksempel (3), men ikke, når det er første sammensætningsled som i *boghandel* i eksempel (4).

Da systemet ikke har kriterier til at afgøre hvordan komposita med flere end to led er opbygget, har man i den tyske analyse truffet et valg: man analyserer altid højrekursivt. Alternativet ville være at levere flere analyser. Det må imidlertid være legitimt at træffe et sådant valg i et operationelt maskinoversættelsessystem.

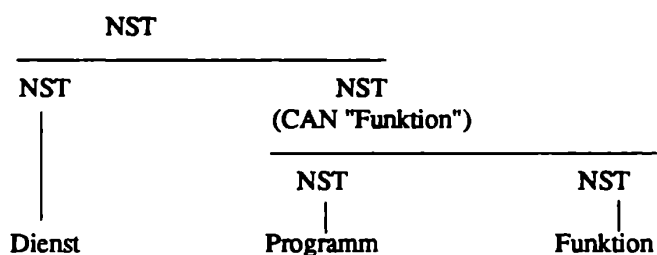
I den danske generering gør vi ikke noget forsøg på at vælge specielle allomorfer eller indsætte fugebogstaver i disse komposita, vi sammenføjer simpelt hen de kanoniske former, så vi får oversættelse (5) *ordbogartikel* og (6) *specialboghandel*. Vi overlader det til posteditoren at indsætte fuge-*s*. Hvis et sådant kompositum optræder gang på gang og posteditoren ikke ønsker hver gang at skulle indføre fuge-*s*, må han leksikalisere kompositumet.



Det, at komposita altid analyseres højrekursivt, får også konsekvenser for den kontekstafhængige kompositionelle oversættelse i transferordbogen:

- (5) Datenbankdienstprogramm  
databasehjælpeprogram
- (6) Dienstprogrammfunktion  
\* tjenesteprogramfunktion

Eksempel (5) bliver oversat korrekt, fordi det tyske substantiv *Dienst* med den højrekursive analyse bliver genkendt som førsteled i kompositumet *Dienstprogramm*. Det gør det derimod ikke i eksempel (6), jf. nedenstående analysetræ, og den kompositionelle oversættelse bliver derfor forkert.



Hvis man vil være sikker på at få oversat *Dienstprogramm* rigtigt i alle sammenhænge, må man leksikalisere det.

I praksis vil en bruger af systemet vælge leksikalisering, når der er tale om fagsproglige termer. Det er vigtigt, at de bliver oversat korrekt og på samme måde hver gang.

Samtidig med, at man løser et problem ved at leksikalisere, skaber man imidlertid et andet. Man ødelægger nemlig muligheden for at oversætte sideordnede komposita med ellipse.

På tysk udelader man ofte andet sammensætningled, hvis det er fælles for to sideordnede komposita:

- (7) Änderungs- oder Umstellungswunsch  
ændrings- eller omstillingsønske
- (8) Maschinen- oder Bedienungsfehler  
maskin- eller betjeningsfejl
- (9) Kopf- und Fußzeile  
top- og bundlinie

I eksempel (7) – (9) er andetleddene *Wunsch*, *Fehler* og *Zeile* fælles for de to sideordnede komposita. I eksempel (7) skal det danske førsteled *ændring* have fuge-s. I eksempel (8) anvendes den specielle allomorf *maskin* af det danske førsteled *maskine*. I eksempel (9) skal det tyske substantiv *Kopf*, da det indgår i et kompositum med andetleddet *Zeile*, have den kontekstafhængige oversættelse *top* i stedet for den normale, direkte oversættelse *hoved*.

En forudsætning for, at vi kan behandle sideordnede komposita med ellipse er, at den tyske analyse får mulighed for at analysere konstruktionen kompositionelt. Kompositaenes bestanddele skal altså findes i ordbøgerne, og kompositaene må ikke være leksikaliserede. Dette skyldes, at analysen følger princippet om "the longest possible match".

Hvis man f.eks. har leksikaliseret det tyske kompositum *Fußzeile*, får man ikke en kompositionel analyse. Det betyder, at *Fußzeile* i eksempel (9) ikke bliver genkendt som et kompositum. *Kopf* bliver dermed heller ikke genkendt som førsteled i et kompositum med *Zeile* som andetled. Vi får

altså ikke den rigtige kontekstafhængige oversættelse af *Kopf*, nemlig *top*, men derimod den normale oversættelse *hoved*, altså *\*hoved- og bundlinie*.

Så kan man naturligvis vælge at leksikalisere sideordningen *Kopf- und Fußzeile*, men en sådan leksikalisering kan gribe om sig. De to komposita kan jo også optræde i andre kombinationer og med andre konjunktioner end *und*:

**Kopf- und Fußzeile**  
**Kopf- oder Fußzeile**  
**Kopf- bzw. Fußzeile**  
**Kopf- Fußzeile**  
**Fuß- und Kopfzeile**  
**etc.**

Komposita optræder hyppigt i tyske fagsproglige tekster, både alene og i sideordninger med ellipse. I vort arbejde med fagsproglige korpora har vi været nødt til at leksikalisere mere, end vi fra begyndelsen havde forestillet os.

**Ellen Christoffersen og Margrethe H. Møller**  
**Institut for erhvervsforskning**  
**Handelshøjskole Syd**  
**Engstien 3**  
**DK-6000 Kolding**