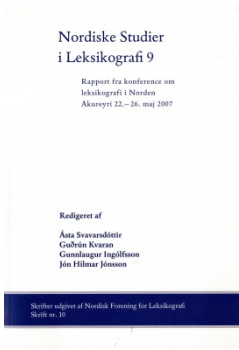


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Fra redaktørens værksted - Om færdigredigeringen af stor Tysk-Dansk Ordbog	
Forfatter:	Torben Christiansen	
Kilde:	Nordiska Studier i Lexikografi 9, 2008, s. 69-76 Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007	
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.

TORBEN CHRISTIANSEN

Fra redaktørens værksted Om færdigredigeringen af stor Tysk-Dansk Ordbog

From the editor's workshop – the final editing of Stor Tysk-Dansk Ordbog. After the authors had handed in the final data files for the dictionary, it fell to me as publishing editor to carry out a cross check in a number of areas in collaboration with the authors. We went through the dictionary data in a transverse fashion to eliminate errors and ensure the necessary consistency. We used as our basis a log file which served both as a to-do list and a memory aid for the editor and the project. In my paper I present some methods to survey and edit a dictionary of this large size, and I present some representative examples: 1. Statistics of fields and their content. 2. Spell check of all Danish words. 3. Check of nouns and their endings and gender within individual entries. 4. Comparison of entries: 50 semantic fields and super-/subordinate concepts such as chemical elements, the grading scale, and Greek letters. 5. Alphabetization of bold parts of idioms.

1. Sigtet med færdigredigeringen

I slutningen af 2005 udkom Gyldendals store *Tysk-Dansk Ordbog*. Den er på 2.125 sider og indeholder 153.000 opslagsord og 42.000 faste udtryk og vendinger. Henrik Bergstrøm-Nielsen og Preben Späth har udarbejdet artiklerne, der er gennemlæst og kommenteret af Peter Jürgensen og Birgit Lohse.

Efter at forfatterne havde afleveret alle artiklerne i den foreløbigt færdige version, kontrollerede jeg som forlagsredaktør i samarbejde med forfatterne ordbogen på en række områder. Formålet med dette arbejde, der strakte sig over ca. halvandet år, var at forbedre ordbogens kvalitet og forhindre senere dyre og besværlige rettelser. I denne artikel har jeg søgt at udvælge nogle af de redigeringsmetoder, der kan have interesse for andre. Jeg koncentrerer mig om ordbogens *indhold*, idet forskellige ordbøgers struktur vil afhænge af det anvendte ordbogsredigeringsprogram mv.

Ordbogen er redigeret i Compulexis, der er et program med en ret "flad" tagtekst-struktur, der består af et feltnavn plus et feltindhold. Artiklerne er lagret i en database. Hvis man skønner, at det er tidsbesparende, kan man eksportere data til WordPerfect, foretage gennemgående ændringer ved hjælp af søg og erstat, makroer el.lign. og genimportere data i ordbogen.

2. Kontrol på tværs af artiklerne

Ved afleveringen af artiklerne var der læst korrektur. Man kan imidlertid ikke gå ud fra, at alt er i orden, og forlagsredaktøren kan roligt være lige så mistænksom som en fængselsbetjent. Det skyldes, at der er mange ting i spil samtidig, når forfatterne gennemgår en artikel: Der er mange forskellige oplysningstyper som opslagsord, bøjningsformer og oversættelser, og der er både tysk og dansk tekst. Det er umuligt at overskue det hele på en gang, og det er svært at være systematisk og huske, at artikler som fx ugedagene bør være ensartede. Når man sammenligner artikler og felter “på tværs”, vil man uvægerlig få øje på fejl og inkonsekvenser. I den sidste fase gælder det følgelig om at isolere ét fænomen ad gangen og undersøge det.

Færdigredigeringen skulle derfor ikke bestå i endnu en gennemlæsning af ordbogen fra a til z. I stedet skulle ordbogens tekst angribes på tværs af artiklerne ved hjælp af metoder og ressourcer, som forlaget stillede til rådighed.

Mine vigtigste værktøjer i færdigredigeringen var: en feltopdelt artikelstruktur, statistikfunktioner, gode søgemuligheder i basen, fleksible printformer, gennemgående ændringer i tekstbehandling samt en logfil, hvor jeg minutiøst registrerede alt kontrolarbejde.

3. Logfil

“Erfaring er evnen til at genkende en fejl, når man begår fejlen for anden gang,” siger man. Men for at undgå fejl kan man notere sine erfaringer i en logfil, der både fungerer som en *to do*-liste og som hukommelse for redaktøren og projektet – nu og fremover. Også andre ordbogsprojekter kan have gavn af søgninger, der enten er generelle eller som kan tilpasses de aktuelle sprog. I figur 1 er der eksempler på generelle søgninger, mens figur 2 indeholder nogle af søgningerne vedr. opslagsord. Jeg har lavet ca. 600 tjek og foretaget godt 2.000 søgninger.

GEN01	Typografi
GEN02	Duplicates check
GEN03	Field code count
GEN04	Forkortelser generelt
GEN05	Forkortelser – enkelte
GEN06	Forkortelser: jm, jn, js
GEN07	ng./ngn./ngt./nt.
GEN08	Enkeltord
GEN09	Stavekontrol – dansk
GEN10	Retskrivningsordbogens nye udgaver 1996 og 2001
GEN11	Stavekontrol – tysk
GEN12	Danske bogstaver i tyske felter
GEN13	Tyske bogstaver i danske felter
GEN14	Bogstavet l i stedet for tallet 1

GEN15	Tal (i stedet for bogstaver)
GEN16	Romertal med punktum
GEN17	Spatium ved bindestreg mellem tal
GEN18	Manglende bindestreg mellem tal og "årig"
GEN19	Ubalancerede parenteser i felter
GEN20	Brøker
GEN21	Bindestreg ændres til tankestreg
GEN22	trykstreg ændres til accentueret bogstav (é)
GEN23	... el. ... ændres til ... (udeladelsestegn, ellipse)
GEN24	Konkurrerende ord
GEN25	??
GEN26	Trykstreger (hovedtryk og bitryk) i tyske ord
GEN27	Trykstregers placering ved -tion
GEN28	personligt subjekt: flere angivelsesmåder
GEN29	"" og »«, replikker og citater
GEN30	Komma
GEN31	overdreven mangler -t
GEN32	Sekvenser af 3 ens bogstaver: aaa/.../ååå
GEN33	Manglende mellemrum efter a. kolon, b. komma, c. lighedstegn, d. punktum
GEN34	Mellemrum ved parentes, spørgsmålstegn og udråbstegn
GEN35	Entities

Figur 1. Nogle generelle tjek, der ikke vedrører specifikke felter.

HWD01	begynder med -
HWD02	,
HWD03	() eller /
HWD04	Komma mellem HWD og REF
HWD05	Feltrækkefølge i artiklens hoved
HWD06	*
HWD07	Uregelmæssige verber
HWD08	Trykstreg i sammensat verbum
HWD09	Trykstreg i sammensat verbum
HWD10	Manglende tryk på -lei
HWD11	Manglende tryk på -rei

Figur 2. Nogle tjek vedr. opslagsord.

4. Overblik over ordbogens indhold

Ud fra en statistik over antal felter og deres indhold grupperede jeg felterne således:

- Felter med tysk indhold, der egner sig til en tysk stavekontrol.
- Felter med dansk indhold, der egner sig til en dansk stavekontrol.

- Felter med indhold, der tjekkes af programmet, fx ordklassebetegnelse og stilværdi. Er der nogle, der bør nedlægges eller slås sammen?
- Felter med en vis uniformering, som har et begrænset antal indhold. Det er fx homografnumre, bøjningsendelser, grammatiske oplysninger samt henvisningsformler.
- Tegnene i udtalefeltet, der skal sammenholdes med listen over lydtegn i vejledningen.
- Små ordklasser som præpositioner og pronomener, der bør gennemgås samlet.

Forkortelser uden for de automatisk tjekkede felter bør man lave statistik over og sammenholde med brugervejledningens forkortelsesliste. Har man eksempelvis brugt *osv.* eller *etc.*, eller er nogle forkortelser så sjældne, at man vil skrive dem helt ud og dermed spare en forkortelse?

5. Stavekontrol af alle danske ord

Vi kontrollerede stavningen af alle 715.000 løbende danske ord i ordbogen ved at køre dem op mod en standardliste med 178.000 grundord og bøjningsformer, der især er udvalgt på grundlag af frekvens. 72.000 ord, altså ca. 10 %, blev ikke genkendt. Det faldt i min lod at kontrollere de ikkegenkendte ord manuelt. Fidusen ved den manuelle stavekontrol er, at det er nemmere at finde fejl, når man tvinges til at se på ét suspekt ord ad gangen, end når man læser løbende tekst.

Jeg burede mig inde, rullede gardinet ned og bevæbnede mig med ordbøger og faglige opslagsværker. På en rigtig god dag kunne jeg kontrollere 2.400 ord, så det tog i alt omkring 35 dage. Jeg gjorde brug af makroer for at bevæge mig hurtigt frem gennem filen og inddelte ordene i tre typer (se figur 3):

- åbenlyse fejl eller eventuelle fejl (dvs. slåfejl, stavfejl, forkert felt, ikkeeksisterende ord, forældet ord eller ord, som jeg ikke kunne få verificeret), fx **afdelig*;
- de "måske egnede" ord, fx sammensætninger eller ad hoc-dannede afledninger, som vi har brug for til en oversættelse, fx *baseler*, *basistysk*, *besættelsesstatut*, *opfyldelighed*, *opinionsforskning*, *opsejlet*, *opstigningssted*;
- gode gængse ord, fx *advokatfirma*, *affaldstræ*, *afisolere*, *aktiehandel*, *aktiens*, *orddannelse*, *ordgåde*.

suspekt ord	TC's kommentar	feltnavn + indhold	opslagsord
afdelig	-ING	TSL neurologisk afdelig	<i>Neurologie</i>
afdragelse	FORFATTER TJEK	TSL afdragelse	<i>Tilgung</i>

afdragning	DO.	TSL afdragning	<i>Abtragung</i>
afdæmpetr	-PET	EXB hans kritik var særdeles afdæmpetr	<i>zahn</i>
affaldsafgit	-IFT	TSL affaldsafgit	Abfallabgaben
affejelse	TC RETTER TIL AFFEJNING (GOOGLE 78, MOD AFFEJELSE 1)	TSL affejelse	<i>Abfertigung</i>
affileret	AFFILIERET	TSL affileret selskab	<i>Konzerngesellschaft</i>
affindelses-godtgørelse	IKKE PÅ GOOGLE. AFFINDELSESSUM FINDES, SKAL DET BRUGES?	TSL affindelses-godtgørelse	<i>Abfertigung</i>

Figur 3. Stavekontrol. De suspekter ord, der er vist både isoleret og i sammenhæng, er forsynet med forlagsredaktørens kommentarer.

Resultatet af gennemgangen var følgende:

- Ca. 10.000 ord kunne umiddelbart rettes eller sendes videre til forfatterne til kontrol. Der blev rettet eller ændret i ca. 4.000 artikler.
- 38.000 ord var "måske egnede". Dem gjorde vi ikke mere ved.
- Og endelig var der 24.000 gode gængse ord, som blev lagt til standardlisten. Vi passer på ikke at putte for mange ord ind i listen, for at den ikke skal miste sin skarphed.

6. Artikelinterne tjek

I en tysk grammatik kan man finde regler for sammenhæng mellem substantivers endelser og deres køn. Eksempelvis skal tyske substantiver, der er afledt med *-heit*, være femininum. Jeg søgte derfor på opslagsord på *-heit*, der ikke stod som femininum, og fandt bl.a. *Birkenscheit* (neutrum). Men det er faktisk ikke en fejl, da ordet er sammensat med *Scheit* (neutrum). Hvis vi havde oplysninger om, hvilke morfemer der indgår i opslagsordene, ville *Birkenscheit* være klassificeret som en sammensætning. Moralen er, at jo flere, evt. skjulte oplysninger man har om sine opslagsord, jo mere præcist kan man søge.

Nogle bøjningsformer kan være uforenelige med bestemte bogstaver sidst i opslagsordet. Eksempel: Hvis et opslagsord, der ender på *-e*, har en bøjningsendelse, der starter med *-e*, kan det være en fejl. Jeg fandt ca. 40 artikler, der skulle rettes, fx **Farbenprobe die** - *-en* i stedet for *-n*. Det samme kan man søge efter i ordbøger med danske opslagsord.

7. Semantiske områder og over-/underbegreber

Vi har arbejdet med ca. 50 semantiske områder og over-/underbegreber, i alt knap 3.000 opslagsord. Hver gruppe indeholder et kvantitativt begrænset antal ord, der kan være alt lige fra ugedage, grundstoffer, græske bogstaver og landnavne til den tyske karakterskala. Formålet med denne redigering er dels at sikre, at vi har optaget alle de relevante opslagsord, dels at redigere dem ensartet. Til klassifikation af disse opslagsord har jeg brugt et skjult felt, hvor vi frit kan skrive et overbegreb (fx 'karakter'), så vi kan søge på artiklerne, printe dem og behandle dem samlet.

Et eksempel: Den tyske karakterskala går fra *Eins, Einser* op til *Sechs, Sechser*. Vi skal her sikre os en overensstemmelse med trinnene på den danske skala. Desuden skal vi være opmærksom på, at eksempelvis *Eins* oversættes "elleve, ellevetal; tretten, trettental", mens *Einser* oversættes "ellevetal; trettental".

Denne ordbog skal indeholde alle de græske bogstaver. Vi fandt de relevante tyske betegnelser ved at søge på indhold i *Duden – Deutsches Universalwörterbuch* (se figur 4). En søgning viste imidlertid, at nogle af dem simpelthen manglede som opslagsord i vores ordbogsbase, og dem, der var optaget, stod en lille smule forskelligt, idet der ved nogle stod "græsk bogstav", ved andre blot "bogstav" eller slet ingenting. Jeg vil påstå, at hvis man laver den slags søgninger i ældre ordbøger, så vil man finde mange ting, der ikke er redigeret konsekvent.

OMR12 Græske bogstaver

Status: o.k.

Baggrund

TC har konstateret, at vi har Alpha og Beta. Vi skal sikre 1) at vi har alle de græske bogstaver, 2) at de er opstillet ensartet.

TC har indsat CLC:

CLC .græsk.bogstav.

Vedlagt udtræk fra Dansk Fremmedordbog.

** griechischen Alphabets" gefunden: 24

Sucheinstellungen:

Suchbereich: Buch "Duden - Deutsches Universalwörterbuch"

Suchoptionen: Volltext (inkl. Stichwörter)

*Alpha

*Beta

*Chi

*Delta

*Epsilon

*Eta

*Gamma

*Jota
 Kappa
 *Lambda
 *My
 Ny
 *Omega
 *Omikron
 Phi
 *Pi
 Psi
 Rho
 Sigma
 Tau
 Theta
 *Xi
 *Ypsilon
 Zeta

* = er med.

Mangler: Kappa, Ny, Phi, Psi, Rho, Sigma, Tau, Theta, Zeta.

Print til HBN 11.12.2003 (15 artikler). Print retur 5.1.2004, TC har rettet 13.1.2004.
 Modtaget print 4.2.04, rettet 6.2.2004.

Figur 4. Uddrag af logfilen: kontrol af de græske bogstaver.

Samme behandling fik sammensætninger, der starter med tal, fx *-stöckig* (‘-etages’/‘etagers’). Vi skulle her fastsætte et rimeligt tal-interval (i dette tilfælde medtog vi fra 1 til 10 etager), der måtte desuden ikke være nogen umotiverede huller, og artiklerne skulle være ensartede.

8. Tjek ved hjælp af tekstbehandling

Et ordbogsprogram vil aldrig kunne tage højde for alle de tjek, som man kan finde på at foretage. Derfor bør man kunne eksportere sine data i en form, hvor man får feltets navn plus indholdet, så man kan massere indholdet på alle mulige måder. Et eksempel:

Foden i artiklerne indeholder idiomer, der kan have et fremhævet ord (“blikfang”), som idiomene skal være alfabetiseret efter. Da ordbogsprogrammet ikke kan alfabetisere efter blikfangene, trak jeg idiomene ud. Dernæst stillede jeg blikfangene op under hinanden. I artiklen **aber** stod to blikfang i forkert alfabetisk rækkefølge:

... *sehr* ...
 ... *schon* ...

Jeg alfabetiserede blikfangene i tekstbehandling, sammenlignede de to filer automatisk og fik derved en markering af forkerte rækkefølger.

Torben Christiansen
cand.mag., redaktør, f. 1955
Børglumvej 12
DK-2720 Vanløse
tc@post.tele.dk