

Om registerværket til Hermes

Af Loránd-Levente Pálfi

I september i år udkommer *Hermes* nr. 39. Dette nummer vil ikke som de tidligere (og fremtidige) numre indeholde artikler, anmeldelser osv., men et registerværk med registre til samtlige danske, norske, svenske, engelske, tyske, franske og spanske tekster i *Hermes* nr. 1-38, som er udkommet i tidsrummet 1988-2007.

Loránd-Levente Pálfi
Research Assistant
Center for Leksikografi
llp@asb.dk



Det er dette registerværk, nærværende artikel, som dels er en præsentation af arbejdet og dels en værkstedsrapport, skal handle om. For en ordens skyld og for at belyse registerværkets betydning og vigtighed samt ikke mindst af hensyn til de læsere, som ikke er bekendte med tidsskriftet *Hermes*, skal der imidlertid først gives en kort præsentation af samme.

Tidsskriftet Hermes

Hermes – Journal of Language and Communication Studies er et internationalt, videnskabeligt tidsskrift for sprog- og kommunikationsforskning med peer review, som udgives af Institut for Sprog og Erhvervskommunikation ved Handelshøjskolen, Aarhus Universitet. *Hermes* udkommer hvert år i februar og september og har siden 1988 etableret sig som et vigtigt forum for international publicering af forskningsartikler inden for et bredt spektrum af sprogligt orienterede projekter og arbejder. Siden 2001 har *Hermes* desuden været tilgængeligt på internettet som *NetHermes* (<http://hermes.asb.dk>).

Tidsskriftets overordnede formål er at udbrede kendskabet til ny forskning inden for områderne diskurs- og genreanalyse, leksikografi, oversættelse hhv. tolkning, erhvervskommunikation, interkulturel kommunikation og almene sproglige emner. Der bringes bidrag inden for alle de nævnte områder i form af artikler, review articles, debatindlæg, præsentationer af nye ph.d.-afhandlinger samt anmeldelser og omtaler af bøger, tidsskrifter og konferencerapporter.

Hermes har bidragydere og læsere blandt forskere og undervisere, som er tilknyttet en bred vifte af forskningsinstitutioner i en lang række lande. Derfor bringer tidsskriftet indlæg på engelsk, tysk, fransk og spansk. Indtil 2005, hvor redaktionen søsatte en ny sprogpolitik gældende fra og med *Hermes* nr. 37, bragtes imidlertid også indlæg på dansk, norsk og svensk (om end de to sidstnævnte sprog har været meget svagt repræsenterede).

Registerværket

Arbejdet med registerværket, som blev afsluttet primo august i år, har strakt sig over 23 måneder. Der var én deltidsmedarbejder på projektet, nemlig undertegnede, som var ansvarshavende for hele arbejdet, og jeg fik assistance af Anne Bøggild Bennekou, som stod for excerpering og digital bearbejdning af de franske tekster og José Augusto Marín Quico, som stod for excerpering af den eneste spanske artikel i *Hermes*. Tidsforbruget fordelte sig på følgende måde: Gennemlæsning af og excerpering fra tekstmassen tog 14 måneder, den digitale bearbejdning, hvortil indekseringsprogrammet Macrex Version 7.18 blev benyttet, tog fem måneder, mens de sidste fire måneder blev brugt på korrekturlæsning.

Foruden registre til de danske, norske, svenske, engelske, tyske, franske og spanske tekster (grundet deres lave antal behandles de norske og svenske artikler dog som danske tekster) i de forgangne 38 numre af *Hermes*, indeholder registerværket to forord (det ene skrevet af chefredak-

tionen, det andet af undertegnede), en tegnforklaring, en tabel med et dertil hørende diagram over sprogfordelingen i *Hermes* gennem årene, en fortegnelse over temasektionerne samt en samlet alfabetisk indholdsfortegnelse, der dels fungerer som et forfatterindeks, dels som et anmeldelsesindeks og dels lokaliserer omtaler af internationale videnskabelige konferencer hhv. konferencerapporter, samt præsentationer af ph.d.-afhandlinger. I alt fylder registerværket 266 sider.

Registrene indeholder stikord, emneord og navne – sidstnævnte gælder såvel personer som institutioner, forskningsprojekter samt (titler på) ordbøger, leksika og (navne på) tidsskrifter. I alt indeholder registrene tilsammen over 15.000 opslagsord (inkl. underopslagsord) med i alt 42.300 sidehenvisninger (inkl. runs).

Indekseret er al tekst (inkl. fodnoter) på nær smudstitelblade, titelblade, kolofoner, indholdsfortegnelser, litteraturlister, abstracts samt "Publications Received"-sider og anden omtækt (fx reklamer, omslagstekst mm.) i *Hermes* 1-38. Den indekserede tekstmasse udgør ca. 7.850 rene tekstsider. Af disse er 2.788 sider skrevet på dansk (det drejer sig om 278 tekster; heriblandt artikler, anmeldelser osv.), 78 på norsk (fem tekster), 13 på svensk (én tekst), 3.380 på engelsk (245 tekster), 1.169 på tysk (91 tekster), 376 på fransk (25 tekster) og 20 på spansk (én tekst).

Hensigten med registerværket turde være indlysende: At lette tilgængeligheden af den viden hhv. den informationsmængde, som forefindes i den givne dokumentmasse. Vil en

Hermes har bidragydere og læsere blandt forskere og undervisere, som er tilknyttet en bred vifte af forskningsinstitutioner i en lang række lande

Ordene har naturligvis måttet ”gøre sig fortjent” til at få en plads i registrene. Som udgangspunkt har et vist mål af informationstyngde været kriteriet.

læser – det være en forsker, en studerende eller en helt tredje – vide, hvor (og evt. hvornår) der i *Hermes* er blevet skrevet noget om fx ordbøger eller mere præcist om ordbogsstrukturer eller om fx tidsskriftet *Budstikken*. *Nyt om dansk-nordisk sprog og sprogrøgt*, behøver vedkommende ikke at gennemlæse de 646 videnskabelige artikler i *Hermes* 1-38 (i alt 7.850 sider) men kan nøjes med at konsultere registerværket.

Principper og nyskabelser

Følgende tre principper har været retningsgivende under arbejdet: For det første har det været en primær bestræbelse at undgå såkaldte begravelser (”begravelse” er en betegnelse for, at et opslagsord i en bibliografi, et katalog eller et register henviser til så mange indførsler, at de pågældende indførsler ”begraves”, jf. *Informationsordbogen* (2002)). For det andet har det været tilstræbt, så vidt det overhovedet har været muligt at mindske muligheden for informationstab. For det tredje blev der – trods tanken om at minimere informationstab – lagt vægt på ikke at skabe regulære frekvensordlister, dvs. optegnelser over ordenes blotte forekomster uanset informationsstyngden i de pågældende tekstsekvenser.

Det er de to første principper, nemlig principperne om at undgå begravelser og at mindske informationstab, der har affødt markeringerne ”(fig)” og ”(def)”, af hvilke i hvert fald sidstnævnte, så vidt det er mig bekendt, er en nyskabelse inden for dansk indeksering: Med ”(fig)” henvises der til en grafisk fremstilling, dvs. en tabel, en figur eller anden form for illustration; med ”(def)” henvises der til en tekstpassage, hvor den pågældende term defineres.

Problemer og løsninger

Der har været forskellige problemer undervejs. De fleste kunne heldigvis løses, men nogle enkelte omstændigheder var ikke til at ændre. I det følgende vil jeg komme ind på nogle af de mest nævneværdige stridspunkter: Opsætningen, alfabetiseringen i det danske register, selektionen af opslagsord og problemet med begravelser.

Opsætningen var det hurtigst afviklede problem: Den lod sig ganske enkelt ikke ændre, idet den var givet på forhånd af det anvendte indekseringsprogram.

Alfabetiseringen i det danske register kan godt give anledning til kritik. Ganske vist er de forskellige ordlister alfabetisk ordnede efter ord for ord-princippet, men da det benyttede indekseringsprogram imidlertid er af amerikansk ophav og primært tiltænkt til brug for indeksering af engelsksprogede tekster, var det nødvendigt at opfatte det danske alfabets tre særbogstaver (æ, ø, å) som hhv. ae, oe og aa. Kritiske røster kunne her indvende, at jeg vel kunne have ændret placeringen af de pågældende ord manuelt. Det er rigtigt. Men når jeg har valgt at lade være, skyldes det hovedsageligt den kendsgerning, at et sådant manuelt indgreb i den digitalt bearbejdede og af indekseringsprogrammet ordnede tekst ville have været forbundet med en betragtelig risiko for at producere fejlplaceringer. Fejl – de være af den ene eller anden slags – kan måske aldrig undgås helt, men en alfabetisering foretaget af et program må alt andet lige regnes for mere pålidelig, end en manuelt foretaget alfabetisering. I den alfabetiske indholdsfortegnelse følger alfabetiseringen dog normal dansk standard med æ/å, ø/ö og å/aa placeret sidst. Da det her drejede sig om en forholdsvis overskuelig tekstmængde (51 sider) og i øvrigt med forsvindende få forekomster af de danske særbogstaver, vurderede jeg det manuelle indgreb som værende ufarligt.

At indeksere så dybt, som det overhovedet er muligt for at hindre informationstab og samtidig passe på ikke at producere en ren og skær registrering af blotte ordforekomster, er en svær balancegang. Ordene har naturligvis måttet ”gøre sig fortjent” til at få en plads i registrene. Som udgangspunkt har et vist mål af informationsstyngde været kriteriet. Men naturligvis har det uundgåelige spørgsmål meldt sig: Hvad er ”et vist mål af informationstyngde”? Jeg har ikke til formålet udkastet en informationsteori eller efterforsket, hvorvidt der findes noget sådant i forvejen. Tilgangen til problemet har været pragmatisk motiveret med brugeren i midtpunktet. Jeg har taget stilling ved hver eneste tvivlsom indførsel ved at sætte mig selv i brugersens sted og spørge: Hvis jeg nu selv havde slået op i et register og havde fået den pågældende henvisning, ville jeg da have været tilfreds?

Princippet om at undgå begravelser har været den mest dominerende ledetanke under hele arbejdet. Hvad det egentlig drejer sig om, og

hvorledes anstrengelserne har båret frugt skal præciseres i det følgende. En begravelse i det danske register ville have set således ud :

Figur 1

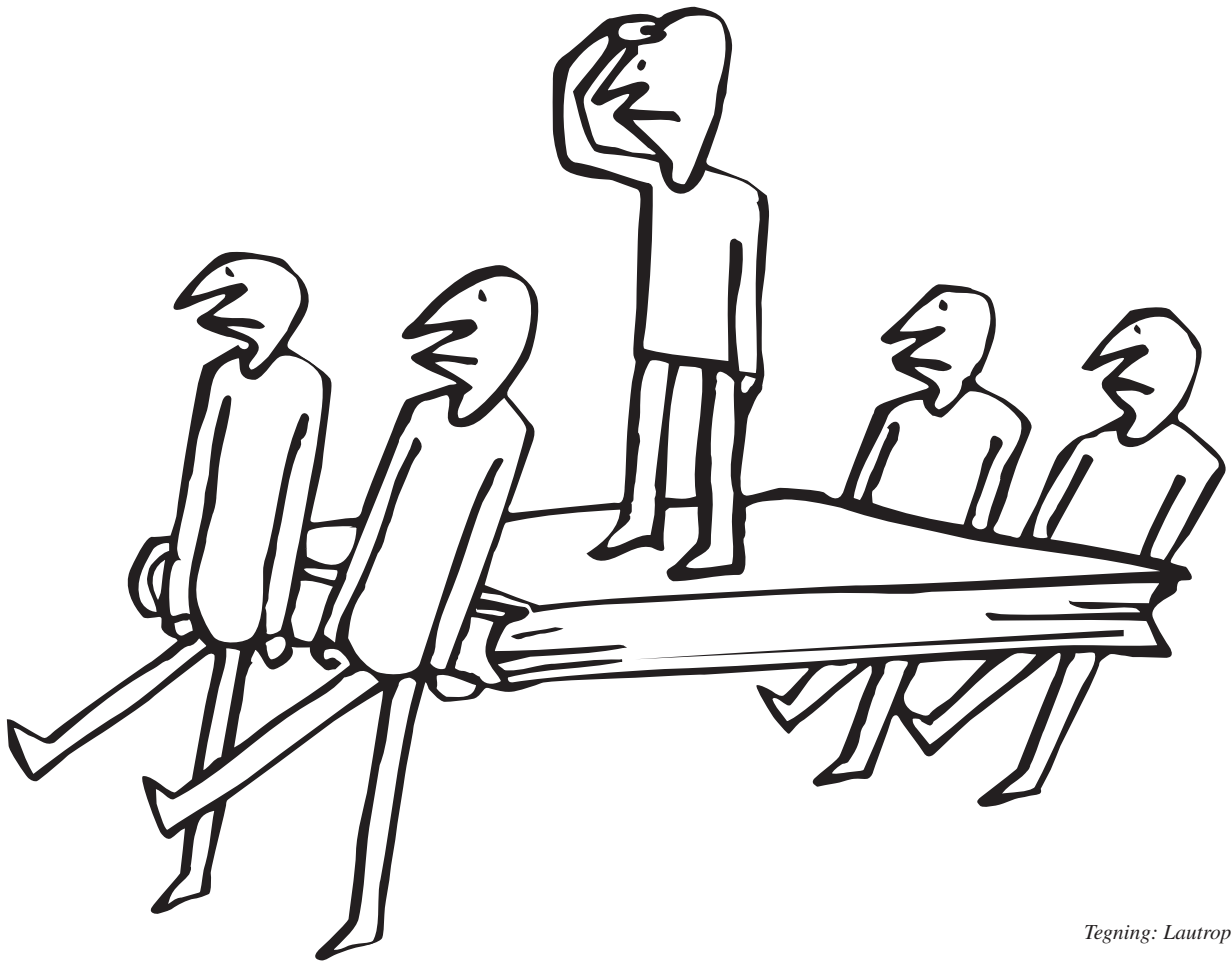
ordbøger **1** 27, 57, 239-244, 252, **2** 55-74, 143, 217-220, **4** 31-46, **5** 127-130, **7** 155, **8** 111, **10** 71-93, 135, 144, 154, **13** 292, 366, **14** 237-238, **15** 193, **16** 267-274, **18** 157-172, **20** 184-186, 195-198, **21** 95-116, 121-136, **25** 171-193, **26** 153, 156, **29** 26-29, 197-215, 327-336, **30** 96, **33** 117-147, 186-188, 190

Tal trykt med fed skrift henviser til det givne nummer af *Hermes* og efterfølges af sidehenvisning.

Når den i figur 1 illustrerede praksis er uhensigtsmæssig og lidet brugervenlig, skyldes det, at det for brugeren er umuligt at vide, hvad der mere præcist gemmer sig bag de enkelte henvisninger. Hvilke af dem er relevante? Måske

Figur 2

ordbøger
ækvivalens **2** 69-74, 69 (def)
-angivelse **2** 69 (def)
-ansættelse af idiomer **25** 189-193
falsche Freunde **25** 189f
pseudoækvivalens **25** 189
-grad **2** 70
betydningsrestriktioner **2** 70f
nul- **2** 70
afskrift i **2** 143
aktiv-passiv-teorien **26** 156
aktive **2** 72 (def), 72f
al- **2** 68 (def), **13** 292 (def)
betydningsforklaringer → definitioner
branche- **16** 267-274, 269 (def)
definitioner i **16** 272
fagsystematik i **16** 271f
flerordsforbindelser i **16** 272f
grammatiske oplysninger i **16** 273
illustrationer i **16** 272
lemmasektion i **16** 270f



Tegning: Lautrop

vil brugeren vide noget om Gyldendals røde ordbøger? Eller om ordbøgers funktioner? Ikke desto mindre er det desværre denne model, der er den absolut mest brugte i Danmark.

Efter råd fra bibliograf og registerfagmand Svend Bruhns – som dog ingenlunde skal stå til ansvar for resultatet – har jeg i stedet konciperet en model (et trappesystem), som optimerer udbyttet af brugerens informationssøgning. Da indførslen med ”ordbøger” imidlertid fylder omtrent to sider, kan der her kun gengives et lille udsnit. I figur 2 ser man således kun begyndelsen af indførslen ”ordbøger”; de forskellige angivelser, som er anført trappevis under indgangsordet, er underopslagsord.

En sidetalsangivelse efterfulgt af ”f” betyder, at omtalen af det pågældende emne eller den pågældende person strækker sig over den angivne side med samt den efterfølgende. En angivelse efterfulgt af ”ff” betyder, at omtalen strækker sig over den angivne side til og med den givne artikels (eller det givne kapitels) slutning. Forskellen på en ”ff”-henvisning og en henvisning af typen ”25 189-193” består i, at informationen om den givne sag i sidstnævnte tilfælde er konstant og findes på alle de angivne sider, mens den i førstnævnte tilfælde er spredt over adskillige sider uden at være sammenhængende. Pilen (→)

henviser til opslagsordet, hvor referencen gives. Der gives endvidere *se også*-henvisninger.

I figur 3 ses underopslagsordet ”idiomatik”, som indgår i hovedindførslen ”ordbøger”. Iblandt henvisningerne finder man, foruden den overordnede og såvel implicite som eksplicite henvisning til ”idiomatik”, mere præcise henvisninger til fx ”opstilling” (25 180-183) og ”lemmatisering” (25 183-189). Bemærk hvordan de to enheder (”opstilling” og ”lemmatisering”) bliver skelnet fra hinanden. Et mindre fintfølede register havde vel blot anført ”idiomatik 25 180-189”. Ved et mindre fintfølede register ville brugeren også have måttet give afkald på henvisningen til ”ordsprogsdominans” (25 188), da den ville have været underforstået i og med henvisningen til ”idiomatik”. Indførslen ”juridiske” er ligeledes et underopslagsord under ”ordbøger”. Et mindre fintfølede register havde blot anført ”juridiske [ordbøger]” uden at vise, at der i den selv samme tekst og på samme sted (2 55-74) gives information om såvel danske som engelske juridiske ordbøger.

Registerværket er sikkert ikke fejlfrit og ukritisabelt, men jeg har dog tilstræbt – inden for såvel de tidsmæssige som de pekuniære rammer – at skabe det bedst mulige værktøj. Om det er lykkedes må være op til andre at afgøre.

Under alle omstændigheder har jeg høstet en del erfaring, som vil komme mine fremtidige indekseringsprojekter til gode, bl.a. indeksbindet til tidsskriftet *LexicoNordica*, som er et nordisk-internationalt, videnskabeligt tidsskrift om leksikografi udgivet af Nordisk Forening for Leksikografi. Projektet befinder sig endnu i planlægningsfasen, og et egentligt indeksværk vil formentlig ikke se dagens lys før tidligst 2010.

Figur 3

idiomatik i 25 171ff
 ækvivalensansættelse 25 189-193
 alfabetisering af polyleksikalske syntagmer 25 173-175
 idiomatiske arkaismer 25 186
 lemmatisering 25 183-189
 markering af idiomacitet 25 175-180
 opstilling 25 174, 180-183
 ordsprogsdominans 25 188
 juridiske
 dansk 2 55-74
 engelsk 2 55-74
 lemmata og eksempler i 18 157-172