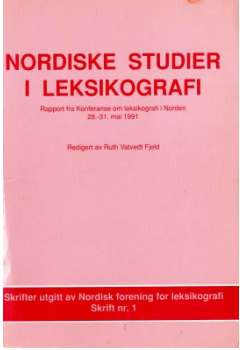


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Ordbogsartikler som skriver sig selv - ? Stednavne som kilde i Gammeldansk Ordbog	
Forfatter:	Merete K. Jørgensen	
Kilde:	Nordiske Studier i Leksikografi 1, 1992, s. 455-464 Rapport fra Konferanse om leksikografi i Norden, 28.-31. mai 1991	
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.

Merete K. Jørgensen

Ordbogsartikler som skriver sig selv - ? Stednavne som kilde i Gammeldansk Ordbog

I bestræbelserne på at beskrive det danske sprog i middelalderen så præcist og så fuldstændigt som muligt har der ved Gammeldansk Ordbog (GO) været gennemført et projekt med det formål at få afklaret, efter hvilke principper det ville være muligt at udnytte sproglige vidnesbyrd fra *stednavne* - en kildegruppe, som ikke tidligere har været systematisk udnyttet ved GO. Nærværende artikel beskriver, hvorledes ordbogsartikler automatisk kan genereres i GOs ordbogsredigeringsystem GestorLEX på baggrund af et maskinlæsbart stednavnemateriale, således at ord, om hvis eksistens kun stednavnebelæg vidner, også kan inddrages og behandles i GO.

Det følgende beskriver de erfaringer, som er opnået under forsøgene med automatisk dataudveksling mellem Gammeldansk Ordbogs redigeringsystem GestorLEX og en database indeholdende stednavnemateriale. De første arbejder med denne udveksling har behandlet de gammeldanske ord som kan ekstraheres fra stednavne, men som ikke findes belagt som gloser i andre gammeldanske kilder.

GO er en videnskabelig ordbog, hvis opgave det er at beskrive det danske sprog i perioden 1100-1515 så præcist og så fuldstændigt som muligt. For at opnå dette, udnytter GO i princippet hele den direkte overleverede tekstmasse fra middelalderen, overleveret enten i form af håndskrifter eller i form af trykte tekster. Ud over dette kernestof findes der imidlertid andre sprogkilder, som indirekte kan skaffe os viden om det middelalderlige sprog, og dermed kan supplere og nuancere den viden vi i forvejen har om ordforrådet i den gammeldanske periode. En af disse sprogkilder er stednavne.

I stednavne kan der findes dokumentation for ord, der ellers ikke er kendt fra middelalderlige kilder, og sproglige vidnesbyrd fra stednavne kan supplere den viden om et bestemt ords kronologi, lydlig udvikling, morfologi, geografiske distribution og betydningsindhold, der kan opnås gennem ordets ikke-proprie forekomst. For at få klarlagt under hvilke forudsætninger og på hvilken måde sprogstof ekstraheret fra stednavne kunne udnyttes i GO, blev der gennemført et projekt kaldet "Det gammeldanske ordforråd i stednavne - til brug ved Gammeldansk Ordbog". Der skal her redegøres for projektets overvejelser vedrørende anvendelse af ordstof fra stednavne som dokumentation for eksistens af ord, som ikke kendes fra periodens ikke-proprie ordforråd. Projektet samt dets konklusioner findes beskrevet i Jørgensen (1989).

Det må for disse navne først sandsynliggøres, at navnets dannelse ikke er sket efter 1515, som udgør GOs grænse. Dette krav kan honoreres fx ved middelalderlig overlevering. Når et navn opfylder dette krav og samtidig anses for at være dannet af et ord, som ikke ellers er kendt fra gammeldansk periode, så kan og skal dette ord inkorporeres i GO, under forudsætning af at ordet samtidig er kendt i ældre nydansk. Et sådant navnebelæg viser nemlig, at det pågældende ord på ét eller andet tidspunkt før 1515 har eksisteret i det danske sprog, og da det er bevidnet i det ikke-proprie kildemateriale i perioden efter 1515 må det

også have eksisteret i den gammeldanske periode. Det kan nemlig ikke godtages, at et ord eksisterer i sproget i en (før-gammeldansk) periode, derefter helt forsvinder i gammeldansk, for atter at genopstå eftermiddelalderligt. Den eneste acceptable antagelse er, at ordet har eksisteret i gammeldansk, men blot ikke er belagt i kilder fra perioden.

Et af projektets delformål var at afprøve metoder til (delvis) automatisk excerpering af stednavne, og som kildemateriale valgtes bind 12 fra serien *Danmarks Stednavne* (Bjerrum 1964). *Danmarks Stednavne*, som udgives af Institut for Navneforskning ved Københavns Universitet, behandler på videnskabeligt grundlag stednavne. De enkelte artikler har en fast og stringent struktur og de besidder tilstrækkeligt karakteristiske træk til at lade sig beskrive formelt; i projektet anvendtes DANLEX-gruppens taksonomi og formatbeskrivelse¹.

Skablund, lb [D *sgå·blæn*, EJ *skå·blæn*, M *sgå·blīn*, *sgå·blūn*, *sgo·blīn*]. 23/6 1438 Skoblingh; Jb 1544 Skobling; KancBrevb c. 7—12/12 1551 Skoubylund; Jb c. 1571 Skublinge; KancBrevb 26/12 1580 Skubling; Jb 1582 Skoubølling, Skowbølling; Regnsk 1582 Skowbølling, Skobling; KancBrevb 12/1 1582 Skobling; KronSk 4/4 1583 Skoblinggaard; KancBrevb 14/5 1584 Skublinggaard; Jb 1589 Skoblingh, Skoblinggaardt; Mandt 1610 Sekobelingh; M 1664 Schoblinggaard; M 1688 Schabling; DATl, VSK Orig Skabling; M 1844 Skrablund; Da. Byremser p. 171 Skabling; Gst, Postadr 1958 Skablund. — Navnet er sandsynligvis et opr. *Skōghbōling sammensat af subst. skov, gl̥da. skōgh, m. og det hyppigt forekommende efterled bōling 'beboer'; se iøvrigt p. 65 under Skåbling.

Figur 1

Fig. 1 viser, at den enkelte stednavneartikel kan indeholde følgende informationskategorier:

- opslagsnavn,
- angivelse af lokalitetens art,
- udtalekompleks,
- belægskompleks,
- tolkningsafsnit.

Typografisk er det især af vigtighed:

- at de enkelte kilders belæg med samt deres datering afgrænses af et semikolon
- at de(t) gammeldanske ord, som stednavnet anses at være dannet af i tolkningsafsnittet er markeret ved angivelse af *gl̥da.*, eller ved angivelse af *, eller ved hjælp af spatium (eller en kombination af disse muligheder)².

Det udvalgte bind blev gjort maskinlæsbart og lagret i en til projektet programmeret database. Indlæggelsen i denne base skete på baggrund af en nøje udarbejdet struktureret syntax-analyse over artiklernes format.

Med udgangspunkt i det nu strukturerede materiale som denne database udgjorde, kunne der foretages ordsøgninger med meget præcise søgekriterier, som bl.a. kunne indeholde krav til navnets overlevering og til det eller de ord, det ansås at være dannet af.

Opfølgende arbejde på projektets konklusioner var forsøg med direkte udveksling mellem projektets database og GOs editeringssystem GestorLEX. Til dette etableredes et samarbejde med TEXTware, det software-firma, som har udviklet GestorLEX.

På TEXTware parsedes en tekstfil, som bestod af en udlæsning af den gamle bases data. Denne parsing konverterede tekstens gamle præsentation til SGML-format³. Som det blev beskrevet ovenfor, besidder artikler fra Danmarks Stednavne en meget præcis struktur, der muliggør at opdele den i informationskategorier og lagre disse i en database. De samme fremtrædelsesformer blev nu benyttet til at mærke tekstfilen i SGML.

Formålet med denne første parsing var at få udskilt alle de artikler, som for det første indeholdt mindst et belæg før 1515, og som for det andet havde forekomst af tekststrengen *gl̄da.* og/eller *. Det første kriterium sikrede, at resultatet ville blive alle de artikler, som behandlede navne, der var dannede senest i middelalderen, og dermed også sprogstof, som ikke var yngre end den gammeldanske periode. Formålet med det andet kriterium var at få mærket de ord som navnet ansås dannet af; af disse ord skulle de med * mærkede senere danne opslagsord i ordbogen.

Det var imidlertid nødvendigt, før en kodning i SGML kunne foretages, nøje at overveje, på hvilken måde de enkelte informationskategorier fra stednavnematerialet skulle korrespondere med GOs grammatik.

Fig. 2 viser et udsnit af GOs grammatik. Som tidligere anført skulle de ord, hvoraf navnet ansås at være dannet, indgå som opslagsord, her benævnt *Opslag*; det som i GOs grammatik benævnes *Citat* skulle korrespondere med belægget fra stednavneartiklen; og *Kilde* til et sådant "citat" blev belæggets datering.

Følgende korrespondancer fastsattes således:

GO	DS
Opslag	tolket ordstof
Citat	belæg
Kilde	datering

Med udgangspunkt i en sådan konverteringstabel foretoges mærkningen af tekstfilen i SGML.

Fig. 3 viser et eksempel på en artikel kodet i henhold til SGML-standard. Under import i GOs grammatik genereres der et felt *Artikel*, hvortil der knyttes en note, som giver oplysning om det pågældende stednavns lokalisering. Der genereres et felt *Hoved*, hvortil der knyttes en note, som fortæller om ordet er valgt pga. forekomst af *gl̄da.* eller *. Der genereres et felt *Opslag*, der indeholder det gammeldanske ord; i eksemplet her står 'grim', med langt i, men koderne for karakterer med diakritiske tegn vil senere blive konverteret efter GOs konventioner. Efterfølgende genereres et tomt felt *Ordkl*, hvis indhold senere skal indtastes. Endelig genereres en *Krop* og en *Krop 1* og et tomt felt *Tyd*, hvor betydningsbeskrivelsen senere kan indføres, og sluttelig genereres feltet *DokTyd*, som består af *Citat* (her: Grynskiær) og *Kilde* (her: 9/3 1311 (Århusb 1428-c. 1450 SRD VI 481)). Hele belægskomplekset (belæg + datering) lægges ind både i *Citat* og i *Kilde*, og det overflødige fjernes så senere.

GOs grammatik

```

<!DOCTYPE Ordbog [
<!ELEMENT Ordbog          (Artikel+)>

<!ELEMENT Artikel        (Hoved, Hals*, Krop+, Hale?)>

<!ELEMENT Hoved         (HomNr?, Opslag, Udnyttelse?, Ordkl, Var*)>
<!ELEMENT HomNr          (#PCDATA)>
<!ELEMENT Opslag         (#PCDATA)>
<!ATTLIST Opslag         antal NUTOKEN      #IMPLIED>
<!ELEMENT Udnyttelse    (#PCDATA)>
<!ELEMENT Ordkl         (#PCDATA)>

<!ELEMENT Krop          (Krop1 | Krop2)>

<!ELEMENT Krop1        (Tyd, Henv?, Stil?, Brug?, DokTyd+)>
<!ELEMENT Tyd            (#PCDATA)>
<!ELEMENT Henv          (JfOrd, Refer+)>
<!ELEMENT Stil          (#PCDATA)>
<!ELEMENT Brug          (#PCDATA)>
<!ELEMENT DokTyd       (Citat?, DokParal*, Kilde, DokForl?)>
<!ELEMENT Kilde         (KildeK, KildeV+)>
<!ELEMENT KildeK       (#PCDATA)>
<!ELEMENT KildeV       (#PCDATA)>

```

Fig. 2. Udsnit af GOs grammatik; de understregede felter angiver de elementer som automatisk genereres.

```

</Artikel>
<#NOTE>ID=Grimskær    Naturnavn</> <Hoved><#NOTE>stjerne</>
<Opslag>gr<u>i</u>m</>
<Ordkl></></>
<Krop><Krop1><Tyd></>
<DokTyd><Citat>Grynskiær 9/3 1311 (Århusb 1428-c. 1450 SRD VI
481)</>
<Kilde><KildeK></>
<KildeV>Grynskiær 9/3 1311 (Århusb 1428-c. 1450 SRD VI
481)</DokTyd>

```

Fig. 3 Eksempel på SGML-kodet artikel.

Til konverteringstabellen skal således tilføjes:

GO	DS
[note t. art.]	lokalisering
[note t. Hoved]	* / gl.da.

Fig. 4. Ordbogsartiklen tegh, sb.; struktur- og tekstvindue.

LexDevelop - ok1sted (test)

Filer Redigér Udklip Søg Præsentation Design Valg Hjælp Afprøvning

Artikel

#NOTE ok

#NOTE ID=Tejbøl 2219

Hoved

#NOTE stjerne

Opslag te:gh

Ordkl sb.

Krop

Krop1

Tyd

#NOTE svarende til ænyda. tey, jf. Kalkar u. Teje og Skovteje; jf. ds 12 s. 171

jordstykke

DokTyd

Citat Teebel

Kilde

KildeK ds.

KildeV 8/10 1471 (LDipl)

DokTyd

Citat Teybøll

Kilde

te:gh sb.

jordstykke Teebel ds 8/10 1471 (LDipl), Teybøll ds 16/5 1486, Deybøll ds 3/2 1489 (c. 1550), theybøl ds 21/2 1492 (17. årh.), Tiebøll ds 18/5 1512.

(TH)o sb.

ler Thom ds 1338 (1558 ÆDA II 38), Thom sogen ds 1391 (1558 ÆDA II 37), 1438 (1558 ÆDA II 36), 1445 (1558 ÆDA II 41), 1446 (1558 ÆDA II 14), 1449 (1558 ÆDA II 37), Theem S. ds 6/4 1463, Thom ds 1467 (1558 ÆDA II 38), 8/6 1470, Thom S.

Én artikel fra Danmarks Stednavne kan godt afstedkomme oprettelse af flere artikler i GO ved import i GestorLEX, da det er antallet af ord mærket med *glda.* eller * i tolkningsafsnittet, der bestemmer antallet af oprettede artikler, jf. fig. 1.

I det ordbogsdokument, som blev resultatet efter import af SGML-filen i GOs grammatik, skulle nu de artikler udsøges, som var blevet oprettet, fordi der forekom en asterisk i tolkningsafsnittet, altså fordi navnet af stednavneforskere ansås at være dannet af et ord, som ikke findes belagt i middelalderlige kilder. Artiklerne udvalgte ved hjælp af en søgeprofil, udformet i et SQL-lignende sprog⁴, som medtog alle artikler, som indeholdt den ovenfor omtalte note til feltet *Hoved* med 'stjerne' som data. Resultatet af denne søgning blev derefter nøje gennemgået med det formål at sikre, at de kunne honorere kravene til indlemmelse i GO. Mange kunne sorteres fra, fx fordi * gjaldt en fællesgermansk, en gammelsvensk eller indoeuropæisk form eller fordi navnet mentes dannet af et andet sted- eller personnavn. Andre igen blev udeladt, da det ikke kunne godtgøres, at det pågældende ord fandtes eftermiddelalderligt. Som eksempel på en artikel i GO, som automatisk er genereret på baggrund af stednavnemateriale ses i fig. 4 artiklen *tegh*, sb.

Fig. 4 viser artiklen, som den fremstår på skærmen, i strukturvinduet til venstre, med artikelstrukturen synlig, og i tekstvinduet til højre, sådan som præsentationen tænkes at blive i det trykte medie. Ordklasse og betydningsbeskrivelse er blevet udfyldt, og kilde og citat adskilt. 'ds' foran kilden angiver belæggenes proveniens.

Antallet af nye artikler som blev oprettet under forsøgene med det enkelte bind af Danmarks Stednavne er hverken nedslående lille eller overvældende stort; og det skal sammenholdes med den forholdsvis lille mængde stednavne, som indgik i disse første forsøg. Afslutningsvis må det indrømmes, at ikke alle relevante oplysninger fra stednavnematerialet kom med. Grunden kan være mangelfulde oplysninger fra forlægget, som her i fig. 5, stednavneartiklen *Lømsgård*.

Lømsgård, gd [D *lœns'go-r*, M *lœms'go-i*, *lœns'go-*]. Dag? 1474 (18. årh. efter Dyrskjøet Rep 3574) Lousgaard; 20/6 1485 (HdDb 1599) thoe gaarde kallis Lømsgaarde; dag? 1492 (18. årh. efter Dyrskjøet Rep 7333) Lousgaard; KancBrevb 23/6 1593 Lømsggaard; M 1664 Lømsggaard (en liden Hgd.); M 1688, DAtl Lømsggaard; VSK Orig Lynsgaard; Gst 1873 og 74, Trap⁴, Gst Lømsggaard; Postadr 1955 Lehnsgaard. — Forleddet er antagelig gen. af trænavnet *løn*, oldn. *hlynr*, m. 'ahorn'. Udtalen med [œ], [e] er i så fald vidnesbyrd om, at det østjyske område, der i nutiden har æ, e svarende til *glda.* y foran n f. eks. i subst. *søn* 'søn', ifl. KortFm p. 65 „fra Skanderborg til Skagen“, tidligere har strakt sig længere mod syd.

Fig. 5. Artiklen *Lømsgård*, Danmarks Stednavne, bd. 12: 168.

Det angives her, at navnet er dannet af trænavnet *løn* og derefter meddeles den oldnordiske form. Ordet 'løn' er ikke overleveret i *glda.* - ældste belæg i Kalkar er fra 1561 - men dette til trods mangler *, og derfor medtages denne stednavneartikel og ordet 'løn' ikke ved den før omtalte søgning. Også andre mangelfulde oplysninger i forlægget kan få indflydelse på resultatet, men arbejde med især de nyeste udgivelser af Danmarks Stednavne vækker håb om større konsekvens i angivelser og i artikelstruktur. Et andet fortrin som disse nyere udgivelser besidder frem for de ældre, er at de, som et resultat af moderne bogproduktion, foreligger i maskinlæsbar stand.

Forsøgene indtil nu har omhandlet de gammeldanske ord som kun er overleveret til os via stednavne. Men stednavnematerialet rummer - som ovenfor omtalt - også oplysninger om i forvejen kendte gammeldanske ords kronologi, lydlig udvikling, morfologi, geografiske distribution og betydningsindhold. Kommende forsøg skal vise, hvorledes også sådanne oplysninger kan optages i GOs korpus. Materialet må udsøges på basis af andre søgekriterier, og det program, som mærker datafilen og siden kan udskrive den i SGML-format, må defineres derefter.

Noter

1. Se DANLEXgruppen 1987 for yderligere oplysning om klassifikation af ordbogsdata og beskrivelse af datastrukturer.
2. En lignende analyse af et andet bind af Danmarks Stednavne er foretaget af Ebba Hjort (Hjort 1985).
3. SGML er det formelle sprog ved beskrivelse af ordbogens struktur (ISO 8879).
4. SQL står for Standard Query Language, se fx Date 1986.

Litteratur

- Bjerrum, Anders. 1964. *Stednavne i Århus og Skanderborg Amter*. Danmarks Stednavne bd. 12. København.
- DANLEXgruppen. 1987. *Descriptive Tools for Electronic Processing of Dictionary Data*. Lexicographica, Series Maior 20. Tübingen.
- Date, C.J. 1986. *An Introduction to Database Systems* I. 4. Edition.
- Hjorth, Ebba. 1986. Leksikografisk onomastik og onomastisk leksikografi. I: *Mange bække små. Til John Kousgård Sørensen på tresårsdagen 6.12.1985*: 95-107. København.
- ISO 8879: *Information Processing. Text and Office Systems. Standard Generalized Markup Language (SGML)*. International Standard Organization, 1986.
- Jørgensen, Merete K. 1989. *Det gammeldanske ordforråd i stednavne - til brug for Gammeldansk Ordbog*. Rapport til Statens Humanistiske Forskningsråd. J.nr. 15-6120. Utrykt.
- Jørgensen, Merete K. 1990. Projektet "Det gammeldanske ordforråd i stednavne - til brug for Gammeldansk Ordbog". I: *Selskab for nordisk filologi. Årsberetning 1987-1989*: 106-114. København.