

NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	En historisk ordbogs anvendelsesmuligheder	
Forfatter:	Guðrún Kvaran	
Kilde:	Nordiska Studier i Lexikografi 4, 1997, s. 131-138 Rapport från Konferens om lexikografi i Norden, Esbo 21.-24. maj 1997	
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.

En historisk ordbogs anvendelsesmuligheder

The discussion in this paper focuses on the possibilities a user can expect from a historical dictionary. The examples used for demonstration are from the current historical dictionary project at the Institute of Lexicography at the University of Iceland. The archives are already in use for various forms of research, although the dictionary project is still in progress. The computerized data available to users will be discussed, and its use in research on the changes in Icelandic vocabulary, inflectional and derivational morphology, and syntax, as well as in research on the modern language.

1 Indledning

I mit foredrag har jeg tænkt mig at diskutere hvilke anvendelsesmuligheder de som benytter sig af en historisk ordbog kan forvente. Mine eksempler henter jeg til det ordbogsværk som er i forberedelse ved Institut for leksikografi ved Universitetet i Reykjavik hvor man har arbejdet på en historisk ordbog i over femti år. Det er almindelig kendt at der går lang tid fra man kommer i gang med et historisk ordbogsværk til det er færdigt, og i virkeligheden bliver sådant et værk aldrig færdigt. Men brugeren kan ikke vente i årtier med at få adgang til materialet. Det er hans ret at kunne få de oplysninger han behøver hvis de allerede foreligger. Det kan han i vores tilfælde gøre til dels ved hjælp af internettet og til dels ved at besøge instituttet.

Men før jeg kommer til mulig anvendelse af materialet må jeg i nogle få ord beskrive instituttets materialsamlinger som de er i dag og som interesserede kan benytte sig af.

2 Instituttets materialsamlinger

Ordbogens samlinger er delt i tre:

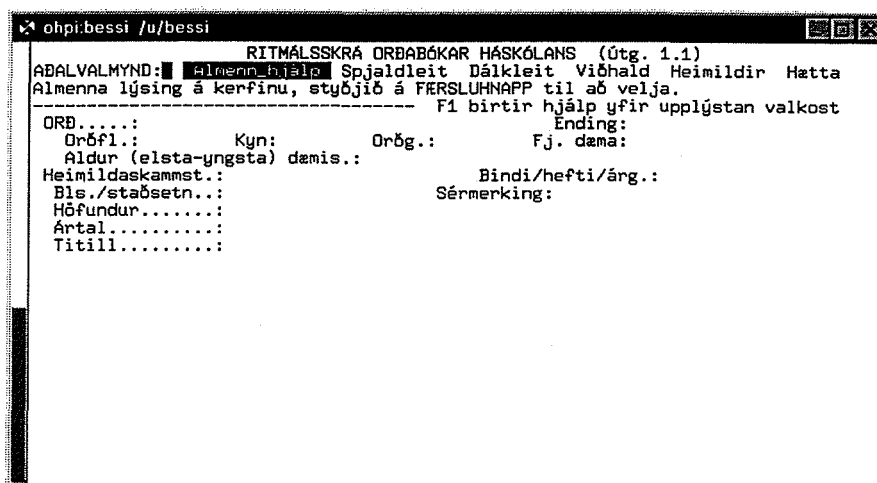
- Skriftsprogssamling med omkring 630.000 ord og 2,5 millioner sedler.
- Talesprogssamling med omkring 60.000 ord og 210.000 sedler.
- Tekstsamling med omkring 700 tekster, ialt med over 25 millioner løbende ord.

Skriftsprogssamlingen er den ældste af disse samlinger. I de første årtier og op til 1980 blev der lagt hovedvægt på excerpering af skrifter fra perioden 1540 op til nutiden, især trykte bøger, men som bekendt arbejder man i København på en historisk ordbog som omfatter norrøn prosa op til 1540. Senere begyndte man, parallelt med excerperingen, at indsamle materiale fra talesproget, for det meste gennem radioudsendelser.

Bearbejdelse af en historisk ordbog er på mange måder anderledes end bearbejdelse af en ordbog over nutidssproget. Hvad den sidste angår så kan leksikografen i de fleste tilfælde støtte sig til sin egen sprogfølelse og han kan også spørge ældre og yngre sprogbrugere. Men leksikografen som skal arbejde med sproget som det var for mange århundreder siden og beskrive ændringer op til dagen i dag har ikke sprogfølelse at stole på. Han har kun foreliggende tekster. En traditionel måde at arbejde på er at excerperere fra trykte bøger og håndskrifter og ordne sedlerne alfabetisk men med alle de hjælpemidler som man i dag har kan man bøde på det utilstrækkelige ved den traditionelle excerperingsmetode. En stor tekstsamling som dækker ordbogsværkets tidsrum er helt nødvendig. I nittenhundredegottitallet blev en samling af tekster i maskinlæsbar form grundlagt på instituttet, og disse tekster bliver brugt til eftersøgning af eksempler på ord og ordforbindelser, både i tilfælde hvor seddelsamlingens oplysninger ikke er fyldestgørende og når seddelsamlingen helt mangler eksempler på ord eller ordforbindelser. De fleste af disse tekster er fra det 20. århundrede men vi arbejder på at maskinlæse al slags ældre tekster. Vi har i sinde at gøre tekster som der ikke hviler copyright på tilgængelige på internettet, yngre tekster må man bruge på instituttet.

I begyndelsen af ottitallet tog man på instituttet de første skridt imod nye arbejdsmetoder med indførelse af edb udarbejdelse af materialet. En traditionel seddelsamling har mange indskrænkninger og eftersøgning er tidskrævende hvis den ikke gælder opslagsordene i alfabetisk ordning. Derfor blev det besluttet at begynde med et oversigtsregister over skriftsprogssamlingen (ritmålskrå). I registret findes alle ord i samlingen analyseret ud fra nogle kriterier:

1. Ordet som man spørger om
2. Sidsteled eller suffiks
3. Ordklassen
4. Køn, hvis ordet er et substantiv
5. Morfologisk struktur, f.eks. g (grunnorð) for usammensat ord, s (samsett orð) for sammensat ord, f (forskeytt) for ord med præfiks, v (viðskeytt) for ord med suffiks.
6. Antal eksempler: foreløbig 1–5 og > (en pil) hvis de er flere end 5, men når citaterne er blevet koblet til registret opføres det eksakte tal.
7. Alder: Der gives oplysninger om det ældste og det yngste belæg. Hvert århundrede er delt i tre perioder, f (første tredjedel), m (anden tredjedel) og s (sidste tredjedel).
8. Kilden er til slut anført med en forkortelse af titelen på værket hvorfra det ældste eksempel stammer.



Til oversigtsregistret er der knyttet et kilderegister med oplysninger om alle exciperede skrifter.

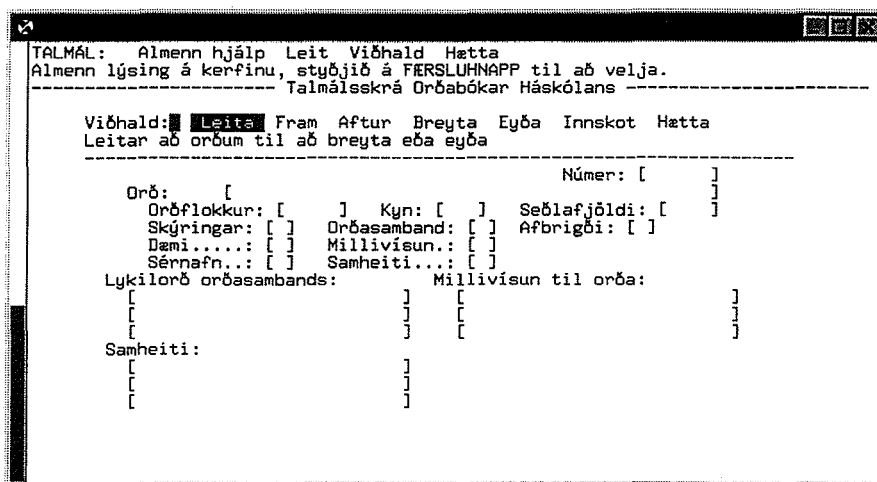
Vi var klar over at oversigtsregistret kun var det første skridt og at det næste måtte være at gøre selve eksemplerne tilgængelige i maskinlæsbar form. For et par år siden fik instituttet tildelt fondmidler til dette projekt, og nu er omkring 800.000 sedler blevet indtastet. Halvdelen af dem er allerede blevet koblet til oversigtsregistret som er tilgængeligt for alle ved hjælp af internettet.

Indtastning af citaterne giver lejlighed til al slags bearbejdelse af eksemplerne, som vil føre til nye maskinlæsbare registre i forbindelse med oversigtsregistret. F.eks. har min kollega Jón Hilmar Jónsson nu begyndt på en liste over ordforbindelser med materiale fra skriftsprogsamlingen. Her tages ordforbindelser i en vidforstand som ordsprog, faste ordforbindelser, såvel som kollokationer. Ordforbindelserne er fremsat på en standardiseret måde og der lægges vægt på at vise deres syntaktiske omgivelser for at gøre deres brug og betydning så klar som muligt. Denne liste indeholder nu omkring 13000 ordforbindelser og i den nærmeste fremtid vil den blive koblet til registret.

Oversigtsregistret har vist sig at være til stor gavn for såvel instituttets medarbejdere som for alle andre som benytter sig af materialet. På instituttet gik man derfor i gang med at lave et tilsvarende register over talesproget. Opslagsordet angives samt ordklasse, substantivernes køn og antal sedler lige som i skriftsprogsregistret. Men på grund af samlingens natur måtte resten være lidt forskellig fra det andet register. Som jeg før sagde har instituttet samlet det meste af talesprogs-materialet ved hjælp af radioudsendelser, men en del stammer fra f.eks. ordlister, ordsamlinger og dagbøger som Ordbogen har fået tilsendt. Samlingen er derfor meget varierende og samtidig uensartet, med materiale fra forskellige kilder. I talesprogsregistret var det derfor nødvendigt at give flere oplysninger som vises som plus eller minus:

1. Er der forklaringer på sedlerne?
2. Forekommer ordet i bestemte ordforbindelser?
3. Findes der variationer af ordet?

4. Er der citater på sedlerne?
5. Findes der henvisninger til andre ord?
6. Er ordet et proprium?
7. Er ordet et appellativ?



Hvis ordet forekommer i en ordforbindelse får man oplysninger om hovedordet, også om ordene som der henvises til og om synonymer.

Dette register er stort set færdigt. Det kan allerede anvendes på instituttet og vi arbejder på også at gøre det tilgængeligt på internettet.

3 Brugeren og materialet

Jeg har nu givet en oversigt over vores samlinger og vil gå over til at gøre rede for den nytte som brugeren kan have af ordbogsværket selvom det endnu er igang.

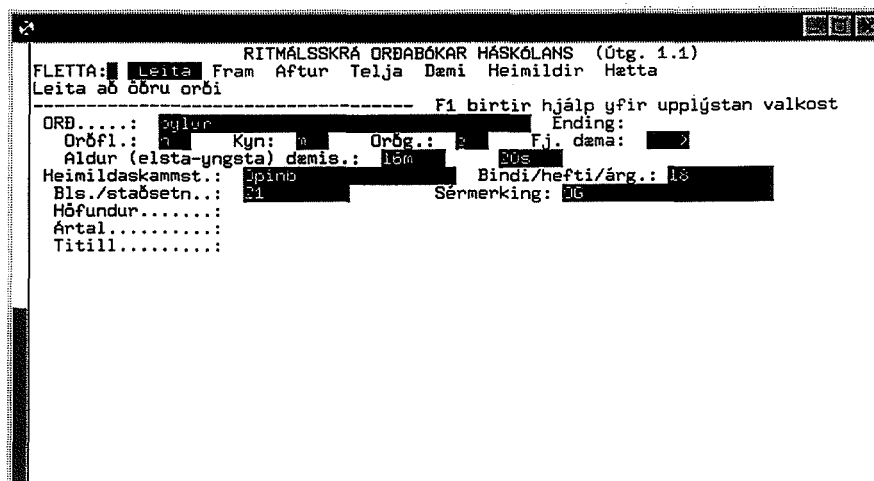
Hvad brugere angår må man skelne imellem på den ene side den almindelige sprogbruger og på den anden side den som arbejder med sprog og sproglige undersøgelser. Den første slår enten op i en monolingual ordbog eller en bilingual ordbog ud fra sit kendskab til nutidssproget. Han henvender sig kun til en historisk ordbog hvis han støder på ord eller betydning som han ikke kender og som ikke findes i en ordbog over nutidssproget. Den anden søger forskellige oplysninger vedrørende ordforrådets historie, historisk morfologi og ordannelseslære, historisk syntaks såvel som undersøgelser af samtidssproget. Begge to kan benytte sig af oversigtsregistre.

Men hvad er det som brugeren ønsker at finde eller forventer at finde i en historisk ordbog?

Opslagsordene

For det første er det selve opslagsordene. Findes ordet overhovedet? Hvor gamle belæg har man? Findes der variationer o.s.v.? Lad os f.eks. tænke os at nogen slår op på ordet *kónguló* som er det vanlige ord i islandsk for 'edderkop'. Det vil han finde eksempler på og det ældste fra den anden tredjedel af det 18. århundrede. Nu bliver han måske interesseret i at vide om der er andre ord om edderkoppen som har *-ló* som sidsteled. Så finder han formerne *kónguló* som vi har lidt ældre belæg på (17f), *kóngulló* (19s), *kóngurló* (20m) men også *gönguló* (18s). Hvis han slår op på *köngu** plus stjerne finder han *kóngulófa* (19f), *kóngurófa* (18s), *kóngurváfa* (19s) og ser han på *köngu** plus stjerne finder han *kóngulvofa* (18m), *kóngurvofa* (17f), *kóngurváfa* (20f). Og hvis han søger videre finder han *göngurófa* som er det ældste i vores samling med belæg fra sidste tredjedel af det sekstende århundrede. Her, som iøvrigt, må brugeren have i sinde at vores excerpering ikke dækker perioden før 1540, således at han er nødt til at slå op i ordbøger over oldsproget for at søge efter eventuelle ældre belæg. (I edderkoppens tilfælde går belæggene på forskellige former af det her omtalte ord helt tilbage til 1200).

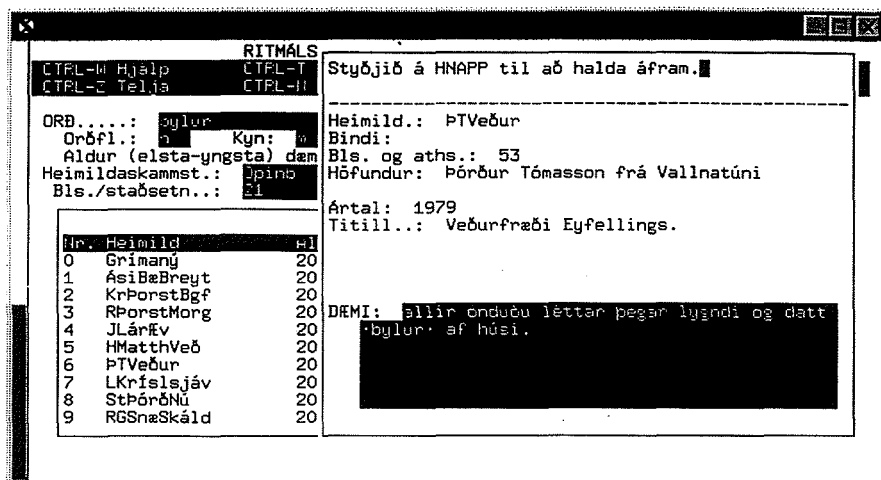
Tager vi et andet eksempel og slår op ordet *bylur* 'snestorm', så får vi de oplysninger at det er et substantiv (n), maskulinum (m), det er usammensat (g) og i samlingen findes der flere end fem eksempler. Det ældste citat er fra midten af det 16. århundrede, det yngste fra slutningen af det tyvende århundrede og den ældste kilde er det Ny testamente 1540.



Derefter kan brugeren kalde frem en liste over de kilder som vedkommende ord er excerperet fra og siden selve citaterne. Han kan også f.eks. undersøge om *bylur* findes i flertal som første led i et sammensat ord: *kafaldsbylur*, *umsvifabylur*, *öræfabylur* o.s.v. I det tilfælde ordnes første led finalfabetisk.

På instituttet kan han slå op i talesprogsregistret hvor han får at se at *bylur* forekommer i ordforbindelser hvor hovedordene f.eks. er *standa*, *hús* og *detta*. I listen over ordforbindelserne vil han finde: *standa af sér byl* i betydningen 'vente til snestormen hører op', *hann stendur með byl*

‘det ser ud til snestorm’, og i overført betydning *það dettur bylur af húsi* ‘pludselig er alt roligt igen efter larm og støj’.



Mange andre muligheder findes. Litteraturforskeren er måske interesseret i en forfatters ordforråd. Han kunde som det første skridt i sin studie søge efter ord i forfatterens værker hvorfra Ordbogen har de ældste belæg. Seddelsamlingen selv hjælper ikke her men derimod oversigtsregistret. Her har vi f.eks. de ord som først, ifølge seddelsamlingen, forekommer i et af Halldór Laxness værker.

Betydning

Hvis brugeren er interesseret i et ordsbetydning forventer han i en historisk ordbog at finde oplysninger om eventuel betydningsændring gennem tiderne. Som eksempel kan nævnes ordet *ffjölskylda* som nu bruges i betydningen ‘familie’. Det er et sammensat ord som består af ordledene *ffjöl-* ‘poly-, mange’ og *skylda* ‘forpligtelse’. I oldsproget (Fritzner I:433) betyder ordet: ‘hvad der påligger nogen som forpligtelse, der skal opfyldes, en gerning som ikke må forsømmes’. Dette er også den betydning som ordet har i de ældste citater i Ordbogens samlinger. Omkring 1800 har ordet fået den betydning som det har i dag, d.v.s. ‘familie’ men i det 18. århundrede ses at overgangen fra den ene betydning til den anden skete på den måde at forpligtelsen blev indskrænket til familien alene.

Et andet eksempel er substantivet *smali* som i oldsproget havde kun betydningen ‘kvæg’ og oprindeligt vel er en forkortelse for *búsmali*. I det 17. århundrede finder man i seddelsamlingen de første eksempler på betydningen ‘den som vogter kvæg, færehyrde’ som er den almindeligste i dag. I denne betydning brugte man før ordet *smalamaður*.

Der findes ganske mange eksempler på adjektiver som har ændret deres betydning. Et af de almindeligste positive adjektiver hos ungdommen i dag er *frábær* som kan bruges om næsten alt som er ‘fortrinligt, fremragende’. Denne betydning er almindelig i oldsproget men i det 17. og 18. århundrede er ordet også brugt i betydningen ‘sjelden, usædvanlig’. De fleste i dag ville undre sig

over en fortælling fra det 17. årh. hvor det siges at et barn var blevet fundet druknet i en brønd. Det var „frábært tilfelli“. Et ‘meget usædvanligt tilfælde’ var betydningen den gang men derimod ‘et fortrinligt tilfælde’ hvis læseren forstår det ud fra nutidssproget.

Et eksempel på et adjektiv som har ændret sin betydning er *sæmilegur*. Det er lavet af samme rod som substantivet *sómi* ‘ære som bevises’ og betydningen af adjektivet *soemiligr* i oldsproget var i overensstemmelse hermed. I begyndelsen af det 18. århundrede findes der eksempler på betydningen ‘anstændig, hæderlig’ men i dag bruges det næsten kun om noget som er ‘nogenlunde, ikke noget specielt’.

Grammatik

I et sprog med levende og indviklet flektionssystem må man antage at forskellige forandringer har fundet sted gennem tiderne. Sprogforskeren kan i Ordbogsmaterialet f.eks. finde nyttige eksempler som belyser den historiske morfologi. Han kan læse på sedlerne hvorledes bøjningsklasser har haft indbyrdes indflydelse på hinanden og hvordan flektionsendelser har ændret sig. Det findes der ganske mange eksempler på. Men det er almindelig kendt at den traditionelle excerpering giver under tiden falske oplysninger om brugen af et ord eller dets flektion. Den som exciperer har tendens til at vælge eksempler som på en eller anden måde er usædvanlige, og ofte findes der flere eksempler på uregelmæssige former end de normale. Her er det ikke nok at have oversigtsregister og seddelsamling. Leksikografen, såvel som sprogforskeren, må støtte sig til en tekstsamling og en konkordans for at være i stand til at skrive en leksikografisk definition eller ensproglig analyse. Det ville være til stor gavn hvis informationer om flektion var koblet til oversigtsregistret og det har vi planer om at gøre. Brugeren kunne så finde et mønsterparadigme tilknyttet det ord han slår op men også flektionsvarianter.

Oversigtsregistret alene kan give sprogforskeren nyttige oplysninger om brug af forskellige suffikser som f.eks. *-legur*, orddannelse, f.eks. om førsteled er roden, genetif singularis, eller genetif pluralis f.eks.:

afskræmi(s)legur
blástur(s)belgur
bygginga(r)vörur
eigna(r)upptaka
frámuna(r)legur
kosninga(r)réttur
rannsókn(a)r ráð
sál(ar)fræði

Endvidere kan man se brugen af køn, som kan være forskellig:

lasting(ur) m./f.
launsát(ur) n./f.
hveiti n./f.
smíð(i) n./f.
sykur n./m.

I mit foredrag har jeg beskrevet nogle af de anvendelsesmuligheder som et leksikografisk institut kan forvente at brugeren er interesseret i. Med stadig voksende edb udarbejdelse af alslags materiale og tilknytning til internettet må man være klar over at de krav som brugeren stiller ændrer sig. Han er ikke længere tilfreds med en traditionel ordbog, selv en ordbog som foreligger på CD-Rom. Han vil have mere oplysninger og de skulde man forsøge at give ham efterhånden som de foreligger under udarbejdelse af en historisk ordbog.

Bibliografi

Fritzner, Johan. *Ordbog over det gamle norske Sprog*. Omarbejdet, forøget og forbedret Udgave. I–III. Kristiania 1883–1896.