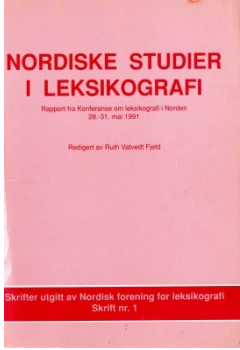


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Fra seddelsamling til database: Leksikografisk analyse af islandske verber	
Forfatter:	Asta Svavarsdóttir, Jón Hilmar Jónsson, Kristill Bjarnadóttir	
Kilde:	Nordiske Studier i Leksikografi 1, 1992, s. 390-402 Rapport fra Konferanse om leksikografi i Norden, 28.-31. mai 1991	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive">http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive</a>	

© 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.

## Fra seddelsamling til database: Leksikografisk analyse af islandske verber

Ved *Orðabók Háskólans* — Leksikografisk institut — i Reykjavík arbejder man på en stor historisk ordbog over islandsk. Hovedkilden er en stor seddelsamling der er baseret på traditionel excerpering af tekster samt andre mindre særsamlinger. Disse samlinger bliver siden kompletteret med materiale fra maskinlæsbare tekster når det viser sig at være brug for flere og/eller bedre eksempler for at få frem en fuldstændigere beskrivelse af ordenes mening og brug. Som første trin i redaktionen bliver der lagt hovedvægt på verbernes analyse og udarbejdelse. I det følgende beskriver vi analyseprocessen og redaktionsarbejdet og fortæller om vores erfaring af at forene ny teknologi og arbejdsmetoder med gammelt materiale.

### 1 Indledning

Efterhånden som databehandling vinder indpas i forbindelse med udgivelse af ordbøger, mister de traditionelle seddelsamlinger deres værdi og bliver kun vidnesbyrd om en gammel arbejdsmetode. De håndskrevne og ofte næsten ulæselige sedler afløses af databanker. I mange lande har man historiske ordbøger, og selv om de trænger til revision, nyder man godt af at have en gennemarbejdet ordbogstekst som grundlag.

Da databehandlingen holdt sit indtog var man i Island endnu ved at indsamle materiale til en stor historisk ordbog over islandsk, overvejende med den traditionelle excerpering af skrevne kilder. Redaktionsarbejdet var ikke påbegyndt, og der forelå ingen endelige planer om udformningen af den ordbog, hvis udgivelse man arbejdede hen imod. I det følgende vil vi prøve at gøre rede for, hvordan vi i Leksikografisk institut har forsøgt at forene nye metoder med gammelt materiale i det forberedende arbejde til ordbogen.

### 2 Leksikografisk instituts seddelsamlinger

Da man i 1983-84 tog datateknikken i brug på Leksikografisk institut, besluttede man at gøre et lille ophold og forsøge at få et overblik over det indsamlede materiale og undersøge mulighederne for at gå i gang med redaktionsarbejdet. Man havde indsamlet materiale i næsten 40 år, og resultatet af indsamlingen var nogle kæmpestore alfabetisk ordnede seddelsamlinger:

*Skriftsprogsamling:*

2.2 millioner sedler med eksempler på ord og ordbrug i skriftsprog fra 1540 og til i dag som er blevet samlet med traditionel excerpering, hovedsagelig af trykte bøger.

*Talesprogsamling:*

Ca. 190 tusinde sedler med kommentarer til ord og ordbrug i talesprog der er blevet samlet gennem forespørgsler, især i radioen.

*Andre mindre samlinger:*

Sedler fra utrykte ordsamlinger fra det 18., 19. og 20. århundrede.

Som forholdene var, forekom det indlysende at man ved indførelsen af edb i første omgang burde koncentrere sig om selve seddelsamlingerne, snarere end redaktionen af ordbogsteksten som sådan. Med dataregistrering af samlingerne sigtede man efter at skaffe sig viden om deres omfang og struktur, inden man gik i gang med selve ordbogen for derved at få en bedre bearbejdet tekst. Man besluttede at gå i gang med to opgaver: For det første udarbejdelsen af et fuldstændigt dataregister over ordforrådet i skriftsprogsamlingen, og for det andet en nærmere bearbejdelse af visse afgrænsede dele af hele samlingen.

Man gik først i gang med udarbejdelsen af førnævnte oversigtsregister. Foruden selve ordene har man også registreret fundamentale oplysninger om dem. Dette arbejde er nu færdigt, og det har bragt forskelligt for en dag vedrørende ordforrådet og samlingens karakteristiske træk. Selv om registeret kun giver begrænsede oplysninger, kaster det lys over visse punkter, som tidligere tiders ordbogsfolk ikke havde nogen mulighed for at gøre sig bekendt med og giver adgang til samlingen fra betydeligt flere sider en før. Vi har nu, bl. a., en ganske klar idé om skriftsprogsamlingens omfang m.h.t. ordforråd og dets inddeling i ordklasser (se figur 1 og 2).

### Skriftsprogsamlingen (1540-1991)

Antal eksempler	2.200.000
Antal ord	610.000
Eksempler pr. ord:	
1	62%
2-4	23%
≥ 5	12%

Figur 1: Skriftsprogsamlingens omfang.

Ordklasse	Verbum	Substantiv	Adjektiv	Adverbium
Antal ord	16.000	500.000	84.000	6.000
% af ordforrådet	3%	82%	14%	1%
1 eksempel pr. ord	44%	64%	56%	50%
2-4 eksempler	22%	23%	25%	26%
≥ 5 eksempler	34%	13%	19%	24%
Grundord	35%	5%	12%	35%
Præfiks/sammensat	65%	95%	88%	65%

Figur 2: Skriftsprogsamlingens sammensætning.

Med hensyn til opgave nr. to, nemlig en nærmere bearbejdelse af en afgrænset del af samlingen, valgte man verberne. Man gik i gang med at dataregistere selve eksemplerne, samt at analysere dem med henblik på udarbejdelsen af en ordbogstekst. Dette arbejde er endnu langt fra færdigt, men ikke desto mindre har man allerede nu en langt bedre oversigt over samlingens karakteristiske træk, end man fik gennem oversigtsregisteret. Hvis man i korte træk vil gøre rede for samlingens almindeligste karakteristiske træk, kan de siges at være følgende:

- \* Et stort og mangfoldigt ordforråd.
- \* Righoldige vidnesbyrd om ord dannelse.
- \* Mangfoldighed m.h.t. betydning, bøjning, syntaks og ordforbindelser.
- \* Eksempler der er værdifulde langt ud over det leksikalske (fx forskellige kulturhistoriske udsagn).

De her nævnte træk er på samme tid samlingens stærke og svage side og indebærer visse mangler, som man må tage stilling til. Inden vi kommer nærmere ind på dette, er det dog nødvendigt at sige lidt om, hvordan man har grebet analysen an og om forbindelsen mellem denne og en påtænkt ordbogstekst.

### 3 Bearbejdelse af samlingerne

#### 3.1 Analyse af materiale

Ved bearbejdelsen af instituttets materiale har man hidtil udelukkende koncentreret sig om verberne, og ved udarbejdelsen af analyseprogrammer har man taget sigte efter dem. Analysen skal for det første være tjenlig i forbindelse med udarbejdelsen af ordbogstekster, og for det andet skal den være så nøjagtig, som det er muligt. Databasen skal senere være tilgængelig for såvel forskere som publikum. Det er derfor ønskeligt, at oplysningerne om ordforrådet er så udførlige som muligt, selv om de ikke alle kommer frem i ordbogen. Da analysen foretages i etaper, kan visse dele af den også komme til nytte ved redaktion, selv om de ikke

forekommer i den endelige tekst. Det er praktisk at kunne inddele eksemplerne efter forskellige kendetegn, der allerede er blevet analyseret for at gøre det nemmere og hurtigere at analysere på senere trin.

Vi vil ikke her komme ind på analyseprogrammerne som sådanne. Man arbejder i øjeblikket med ændringer af disse i forbindelse med en ny database, der skal tages i brug, og samtidig foretager man en omvurdering af selve analysen og dens tilrettelæggelse. Analysen forbliver dog i princippet uændret og kan derfor beskrives her i store træk.

### 3.2 Verbalanalyse

Det er meningen at eksemplerne på ordenes brug skal udgøre den vigtigste del af ordbogsteksten. Bearbejdelsen går derfor først og fremmest ud på analysen af eksemplerne. Hvert enkelt eksempel analyseres systematisk, og analysen af verber kan inddeles i fem hovedpunkter:

- \* *Eksemplets aldersbestemmelse.* Den kan udføres maskinelt efter kildefortegnelse.
- \* *Morfologisk analyse af verbets form i det foreliggende eksempel* (modus; form; person og tal; tempus).
- \* *Syntaktisk analyse af verbets tekstuelle omgivelser.* De sætningsled, der på en eller anden måde knyttes direkte til verbet, analyseres m.h.t. form og betydning. Der er her tale om subjekt, objekt — et eller to, samt adverbier og præpositionsled der er nært knyttet til verbet som en slags fast bestemmelsesled (se figur 3).
- \* *Semantisk analyse af verbum eller verbumsforbindelse.*
- \* *Andre punkter i nødvendigt omfang:*
  - faste ordforbindelser
  - ordsprog
  - kulturhistorisk beskrivelse, forklaring vedrørende ordbrug m.m.

Efter analysen sørger datamaskinen for inddeling af eksemplerne og ordner dem efter visse regler, hvorefter et ombrydningsprogram, der er tilkoblet databasen sætter ordbogsteksten i overensstemmelse med analysen. Figur 4, hvor man kan se en prøvetekst af verbet *péna* (tjene), viser resultatet. De syntaktisk analyserede afsnit "oversættes" maskinelt som hoved til vedkommende artikel, - d.v.s. at datamaskinen selv udformer redaktionel tekst, fx *péna e-m til e-rs, péna e-u* o.s.v. Tekstens udformning bestemmes især af verbets form og dets tekstuelle omgivelser. Beskrivelsens hovedafsnit tager sigte efter visse morfologiske og syntaktiske træk: aktiv, medium, (selvstændige) participier; argumenter (tilstedeværelse og type af subjekt og objekt(er)). Disse hovedafsnit kan så blive delt i mindre afsnit m.h.t. andre led der er nært knyttet til verbet (ordnet alfabetisk efter præpositioner og adverbier).

## Syntaktisk og morfologisk analyse

A. eptir sem uppáfynding þessi þjenar fólkinu til virðingar og þrýðis. Kilde: JólGrvKlím. 157 Alder: 18m	B. Á ég engar skræður, sem þéna þessu guðspjalli? Kilde: JThSk. II, 52 Alder: 19m
---	---



Verbum:

	Form	Modus	Tempus	Tal	Person
A. þjenar	aktiv	indikativ	presens	ental	3
B. þéna	aktiv	indikativ	presens	flertal	3

Subjekt:

	Kasus	Betydning
A. uppáfynding þessi	nominativ	-levende AKT
B. [skræður]	nominativ	-levende BOG

Objekt:

	Kasus	Betydning
A. fólkinu	dativ	+levende MENNESKE
B. þessu guðspjalli	dativ	-levende TEKST

Adverbial:

	Type	Ord	Nominal
A. til virðingar og þrýðis	præp.led	til	genitiv/-levende
B. —	---	---	---




---

A. þéna e-m til e-rs  
B. þéna e-u

---

Figur 3: To eksempler på verbet *þéna* og hovedtrækkene i den morfologiske og syntaktiske analyse af dem.

Betydningen er underordnet i forhold til strukturen. Hver artikel får sin betydningsbeskrivelse, og hvor der er tale om mere end én betydning, indeles vedkommende artikel efter nummererede betydninger. Ordforbindelser placeres i artikler, hvor de anses for at høre hjemme m.h.t. struktur og betydning men får (som oftest) en særlig betydningsbeskrivelse. I hvert enkelt afsnit eller artikel ordnes eksemplerne efter alder således at de ældste eksempler kommer først og de yngste sidst (det er angivet foran hvert eksempel fra hvilket århundrede og hvilken trediedel af århundredet det stammer, således at et eksempel fra 1540 kendetegnes ved *16m* og et eksempel fra 1921 kendetegnes ved *20f*).

I verbets hoved gives oplysninger om bøjning foruden andre punkter, der vedkommer verbet i det hele taget.

### 3.3 Vurdering af materialet

Først efter analysen kan man danne sig et billede af det materiale der er indeholdt i seddelsamlingen og vurdere beskrivelsen, bl.a. i sammenligning med foreliggende kilder om det islandske sprog, især ordbøger. Man må tænke på, at edb-analysen gør det muligt at fremstille en tekst, der er mere systematisk og udførligere analyseret, end man tidligere havde mulighed for at gøre. Samtidig kommer der visse skævheder, lakuner og slagsider til syne, som skjuler sig nemmere i en mindre systematisk og mere kompakt tekst. Det kan dreje sig om:

*Tvivilsspørgsmål* der skyldes at de foreliggende eksempler ikke er klare nok. Mange af eksemplerne i seddelsamlingen er så knappe, at de ikke forslår til at vise syntaks eller betydning klart og utvetydigt. Eksempler (ikke mindst fra poesi) kan også være vanskelige at forstå eller måske helt uforståelige. Endvidere er nogle eksempler tvetydige m.h.t. syntaks og/eller betydning.

*Lakuner* i beskrivelsen, hvor der mangler eksempler på strukturer og/eller betydninger som man ved om. Det kan enten være sådan at vi slet ingen eksempler har eller at vi ikke har eksempler fra et bestemt tidsrum.

*Skævheder* af forskellig slags, der ser ud til at findes i beskrivelsen. Antallet af eksempler på enkelte strukturer og/eller betydningsvarianter kan se ud til at give et forvansket billede af deres stilling i sproget. For det første kan der være forholdsvis mange eksempler på det, der må anses for at være forholdsvis sjældent og få eksempler på mere almindelige strukturer og/eller betydninger. Eksemplernes aldersdistribution kan også virke unaturlig i forhold til, hvad man kunne forvente. Og til sidst kan forekomsten af bøjningsvarianter se ud til ikke at give det rigtige billede m.h.t. alder, hyppighed o.s.v., — oftest derved at der er flere eksempler på forskellige afvigelser, end man kan anse for at være naturligt.

Sådanne skævheder må så vidt muligt rettes, selv om ordbogstekstens fremstilling indebærer en vis fare for slagside.

Det er meget forskelligt fra det ene verbum til det andet, hvor fyldestgørende en beskrivelse seddelsamlingen indeholder. Det er også forskelligt, om der helt klart er tale om lakuner og/eller skævheder, eller om der blot er tale om mistanke om sådanne. Det ville være ønskeligt at efterprøve seddelsamlingens vidnesbyrd for at stadfæste det billede, som det

giver, eller for efter omstændighederne at gøre det mere fuldstændigt. Datateknikken har gjort dette muligt, i hvert fald indenfor visse grænser, med søgning i maskinlæsbare tekster. Vi vil til sidst komme lidt ind på vores erfaring m.h.t. en sådan søgning efter verber i den tekstsamling, der forefindes på instituttet.

## 4 Tilføjelse af materiale

### 4.1 Tekstsamlingen

Leksikografisk institut har en ganske stor tekstsamling, der ikke desto mindre er begrænset i mange henseender. Den består nu af 171 tekster med ialt godt og vel 15 millioner løbende ord (ca. 80 Mb). De fleste tekster er ganske nye (fra de sidste 6-10 år) men samlingen indeholder dog også én tekst fra det 16. århundrede, to fra det 18. og én fra det 19. århundrede, foruden nogle gamle tekstfragmenter. Teksterne er endvidere ensidige m.h.t. struktur og indhold, langt den største del af teksterne stammer fra nogle få emnekatogier (romaner, biografier eller erindringer men også Bibelen og den islandske lovsamling i sin helhed).

Det har kostet forholdsvis lidt besvær og få udgifter at tilvejebringe tekstsamlingen. De fleste tekster har man fået i maskinlæsbar form fra bogforlag og trykkerier, og det forklarer til en vis grad ensidigheden. Den søgning efter eksempler, som vi har foretaget indtil nu, har til dels været forsøgsvirksomhed, og man har planer om at tilvejebringe en mere systematisk og mere varieret samling, når man har høstet lidt flere erfaringer.

### 4.2 Søgning i teksterne

På Leksikografisk institut anvendes operativsystemet Unix og tekstsøgeprogrammet GNU-grep fra Free Software Foundation. Det søger bogstavrækker i tekstsamlingen og giver linien med det ord som man søger foruden de to forudgående og de to følgende linier. Man må søge alle ordets bøjningsformer, og af dem kan der være temmelig mange i islandsk. Man laver en liste over alle bøjningsformer, der forekommer i hvert ord og søgeprogrammet søger siden i overensstemmelse med listen. Dette er forholdsvis enkelt, selv om det er ret forskelligt hvor mange rækker der skal opgives. For et regelmæssigt verbum med svag bøjning må man søge efter 30 forskellige bogstavrækker, men varianter i ortografi og/eller bøjning kan forøge antallet betydeligt, — helt op til godt og vel 100 former.

Når der findes mange eksempler, er det et forholdsvis stort arbejde at gennemgå hele samlingen af disse, sortere dem og udtage eksempler på andre ord, end det man søger. Der er en hel del homografiske former, og i de fleste tilfælde kommer der en del uvedkommende eksempler frem ved søgningen. Det er nødvendigt at være på vagt over for dette, for nogle homografiske former er temmelig besværlige at søge af denne grund. Skal man fx søge verbet *eiga* (have), må man passe godt på, idet én af verbets former er *á*, og er derved homonym med en af de hyppigst forekommende præpositioner i sproget, foruden at *á* også optræder som visse former i to forskellige substantiver. I en lille del af tekstsamlingen viste der sig at være godt og vel 12.500 eksempler på *á*, og heraf var kun knapt 3% verbalformen *á*. Til trods for enkelte eksempler af denne art sker det forbavsende sjældent, at homografiske former volder besvær.



Resultatet af søgning er meget forskellig, efter hvilket verbum der er tale om, og tekstsamlingen viser sig at være nyttigst ved analysen af verber i gennemsnitsstørrelse. Når verbet er meget hyppigt forekommende koster det et enormt arbejde at gennemgå eksemplerne og resultatet svarer absolut ikke til anstrengelserne (i en lille del af samlingen var der fx godt 22.000 eksempler på verbet *vera* (være)) og hvis verbet er forholdsvis sjældent kan det være tvivlsomt, om det overhovedet findes i tekstsamlingen. Resultaterne af tekstsøgning kan være tilfredsstillende af to forskellige grunde:

- \* selv om den ikke giver noget nyt, kan den være nyttig fordi den giver bedre eksempler end dem, man havde i forvejen eller en mere fyldestgørende aldersdistribution end den, der findes i seddelsamlingen;
- \* den giver rene tilføjelser til seddelsamlingen, dvs. eksempler som viser sætningskonstruktioner, betydninger eller ordforbindelser som ingen eksempler fandtes på i forvejen.

Vi skal nu se på to konkrete eksempler på resultater af søgning. I første tilfælde kom der intet nyt frem, selv om søgningen ikke var helt uden gavn. I sidste tilfælde havde seddelsamlingen øjensynligt givet en meget mangelfuld beskrivelse, og her gav søgningen gode resultater.

#### 4.2.1 *péna*

Verbet *péna* (tjene) er forholdsvis sjældent i islandsk nutidssprog. Ikke desto mindre fandt man ialt 51 eksempler ved søgning i tekstsamlingen. Selv om disse ikke viste nogen strukturer eller betydninger, som der ikke var eksempler på i forvejen, fandt man eksempler, der gjorde seddelsamlingens beskrivelse mere fuldstændig m.h.t. aldersdistribution. I 7 af 19 artikler fik man nye eksempler fra tidsrum, som ikke var repræsenteret i seddelsamlingen. Det er bemærkelsesværdigt, at de færreste af de nye eksempler er fra det 20. århundrede, selv om de fleste tekster er fra dette århundrede, og de fleste af de ældre eksempler kommer fra tekster, der allerede var blevet excerperet. Endvidere kom der nye eksempler, som måtte anses for at gøre ordbogsbeskrivelsen bedre, uden at de direkte udfyldte lakuner. Af de 51 eksempler, der blev fundet i tekstsamlingen, er 8 benyttet i verbets beskrivelse (d.v.s. i prøveteksten i figur 4).

**béna** -tí/-aðl (nt. alltaf -ar), -t/-að  
Sjá Orðabók Fritzens.

**ÞÉNA** = 1. *þjóna* [e-u, e-m] 16m og hefði gjarnan kent þeim huernum þat Laugen ætti at þjéna. CorvPost., II 11 18s "hvær þénti þýlgjarr; / enn þú? Satan! MiltPar., 122 19f Sá kann ekki að stjórna, sem ekki hefir lært að þéna. CJ., 286 19f Sælir eru þeir sem vel hafa þént (af hafa lokið). CJ., 332 #2. *vinna, starfa; gegna starfi d áku, stað e. í áku embætti, vera í vísit e-rs staðar* 18f þann tíma Jón Magnússon þénti í því gestlega standi prestsemþættinu. Alþ. IX, 524 (1709) 18s 19f hann má ei þéna í réttinum nema í forðillum bróður sína. Fjalik. 1885, 79 (um 1800) 19m [hann] þénti í þeim stórum til dauðadags. CJ80ist., 49 (1863) 20m Hún hafði... þénað í húsum hjá fyrirmyndarfolki. VSVFimm., 52 20m Hjá honum hafði Magnús þénað og numið um 6 ára skeið. Arsrtsf. 1959, 72 #3. *nýast, koma að gagni; eiga við* 18m uppi Framkyrkjuni eru alls 16 bord sem þjéna fyrir Loftf. Þps. Áll 21, 6 (1758) 16m Til er eirnenn 1¼ af raufþryktu Cartune sem þéna kann í alltaresdúk. Þps. Áll 21, 214 (1761) 19m sú aðkomandi stúlka fór höndum um konna er á gólfu lá eftir því er þénaði og hún vissi að með þurft. JÁBj. 1, 17 20 mætti sumt í þessum ritdóm! þéna sem spegill, að vísu mjög skjáður, er aðrir ritdómar gætu séð í brot af sjálfum sér. ÞórbþEin., 175 20m [glæraugun] hafa þénað betur en glæraugun mín, og nú er ekki til neins að leita. ÞórbþEið., II, 342 20s En „þó hér séu litlar framfarir þá hef ég þó séð ýmislegt er ég hefði ekki áður séð. Hefur mér þá alltaf orðið að spyrja hvort þetta gæti ekki þénað heima!“ SvipAthn., 183 #4. *afila e. græða peninga* 20m ég vona að þú þénir vel í vetur, svo að þú getir sent mér aura. ÞjómMát., 251 20s Þegar ég hefði lokið verkefni mínu í nóvemberlok hafði ég þénað mjög vel. JóhSvEisk., 207 20s andskotinn og einstaklingurinn eiga að þéna en alþýðuræfilinn fær ekki neitt! LÁbergBlatt., 188

málsh.  
málsh.

væri á smávörðinni. Austri. 1887, 13 20m Það eru meiri lifandis ósköpin, sem allir geta þénað á þessu ströð. StSteinGlugg., 27 20s Það var sumaríð '52, ég... ætlaði að þéna einhver lifandis ósköp á því að salta síð. VéstLÉfir., 22 20s Mennirnir veir... státu hæglaftir við ódýran drykk... Það væri ekkert hægt að þéna á þeim. 80napt., 161

**ÞÉNA E-U**: #1. *þjóna e-u, stjórna e-ð, eiga við e-ð e. falla að e-u* 17 "þó vil eg gjarnan þéna hvers manns sinni / vísit í góðu, en vöndu ei. JSamsKD. II, 52 (17. öld) 17s Noturnar og Tomarnar eiga að þjéna Orðunum. Grad. A, III / 18m stundabe mitt eggi gagn undir þeim tilslætte, at eg vilde þar með þjéna almúgans tilstande. JÓGrvKlím., 136 19m Á ég engar skræður, sem þéna þessu guðspjallri JThsk. II, 156 20m toppfallsblakkir, svo og allt fast í reidunum, sem daglega þénaur lausum reidunum, er tekið niður... vor og haust. Verksjóv., 101 20m lengi hafði verið svipast um eftir drottningarefni, sem þénað gat tilgengi frúarinnar. BergjMann., 311 #2. *starfa e-rs staðar, þjóna e-u (með starfi að e-u embætti e. viðfangsefni)* 18f að lögmáðurinn Lauritz Gottrup dragi sig aldeilis undan að þéna réttinum í morðsök þeirri, sem víðvikur Magnósi Benediktssyni. Alþ. IX, 378 (1705) 18f [svo] presturinn, sem þénti Kolbeinsstaðakirkju [mætti tveim nóttum hvílast, þá hann fremur þar þjónustu vegna fólkins nauðsynja]. JHðsk. I, 132 #3. *koma að gagni e-rs staðar, nýstas e-u* 18f Reifingurista þjargleg, brúkast so sem heimilinu þjéna. Jarðab. XI, 27

**ÞÉNA E-M**: #1. *þjóna e-m, koma e-m að gagni* 17m "Loddurum þénti listin sú, / að láta folana gæta. JGubnÁrm., I, 43 17s "allflestar frúr um Danmerkið rídi / voru sestur at henni þénti. lík. II, 40 18 "Læzt ég vanda legstaðinu, / — lengi klárin þénti mér. HérabSgf. II, 276 (18. öld) 18m Þeir voru eingvum lögum under-örpner, heildur þjénaði þeim sé eintóme keisarans vilje fyrir Jög. JÓGrvKlím., 189 19f með hvörju megum við hafa þá eru að þéna Ydur? MSIGláv. I, 84 19f Sá, sem ofmjög þénað öðrum, fortaerir sjálfum sér. CJ., 296 19f Sá þénað öðrum nóg, sem gjörir það af alvöru. CJ., 302 20 til hvers er verið að þessu? Hverjum þénaur það? ÞórbþEin., 158 20m Ein stór og mikil líkurnir, vandlega girt og vel sterk, þénaði öllum vinnuöðnum. GHagaVrk., 23 #2. *hafja e-m, koma e-m vel* 18m so hefði [minn lereimastari] að vísu kendt mjer þat sem þjénaði einum áhöndandi lýkjúþjénara. JÓGrvKlím., 234 19m að enginn býði sér... neins konar hægingdi sem ferðamanni þénaði. JÁBj. II, 227 19s Þú erst svo rúðsettur og gæðinn og sér strax allt hvað höndum þénaur. JGubnSKth. I, 34 (1879) e-m er þént með e-ð: Tms Hér er nú ekki þént með það, ég er til ekkert hrifin af þessu, langar ekki til einhvers. S-Múli. e-m er þént með e-u: e-ð kemur e-m vel, e-ð hafir e-m 17m "Samt er þér ekki þént með þú, / þó eg þig auka villdi. HPPass. XXXI, 12 18m Hvat er okkur þá þént með því að binda öðrum þjént með þessu, því eg treysti mér alls ekki til að gjöra það svo, að þinginu verði þént með því. Alþ. 1855, 794 19m hvermin kirkjum góð prestum gesti nú leingur verið orðið þént með eður hagkvæmara, að fá greiðslur sínar í ull og tóg. Alþ. 1849, 361 e-m er þént með e-m: e-r gagnast e-m 19s mér er þetta bending um, að stjórnmíni sé ekki þént með mér. ÚrtjÁ. II, 199 (1873)

málsh.

málsh.

þéna að e-u: *valda e-u, stuðla að e-u* 19s Það sem þjénað er hreifingunni [þ.e. í vatnsmyllu]. Nf. XI, 60

þéna á e-u: *græða (f) á e-u* 20m menn eru nú vanir að gefa fremur með sér, heldur en þéna á því. ÞorEinnKinn., 89

A+

A+

þéna til e-rs: *koma að gagni e. nýstas við e-ð; stuðla að e-u* 16m með þú / þat í holddino at lífa þjéna mér til a usgatar at af þa veit eg eigi huort eg skal kiosa. Fil. I, 22 (OG) 16m Enn andi er sa hinn innre madur og yre / sem lífver og verkar það sem til andgígra hluta og epter komandi lífdaga þjénaur. Rómform., a II / (OG) 17f að því Leite sem [voinin] þjénaur til Syndnanna fyrrengingur. VidPost. II, 303f 18m í þessu lande var yferljóstanligt af öllu því sem þjénaði til mannsins gagns og hæginda. JÓGrvKlím., 188 18ms það þjénaur og til að gjöra greinarmun milliun dyggða og ódyggða. ÞPStJÁh., 4 19 "að þér streymj jafnan / all, hvað þéna yrdit til. GPHSKáttk. II, 170 (19. öld) 19m Sér hún þá til hvers [hrifurinn] þénaur og tekur feginlega á móti honum. JÁBj. 1, 261 19m Flest vilji þið mér fyrirnuma sem til skemmtunar þéna mætti. JÁBj. V, 90 19s gaf [hann] honum fullnagt til að taka sér fyrir hendur það hann áliti þéna til að málinu yrði framgengt. FréggFyig., I, 313 20 efturfarandi ráðleggingar munu þó geta þénað til að lína nokkuð þrautir sjúklingsins. ÞórbþEin., 9 20m Enginn vissi til hvers þetta næstum útmáða mannvirkí hafði þénað í lífinu. ÞórbþStein., 224 20m allt sem þénaur til upp lýsingar um síða og kristniháld er dyggð. Þjómjómf., 61 20m Ég hef hugsað mér að bjóða dönsku stjórnmíni mína uppfinningu... en ég er þó hræddur um, að það þénti ekki til neins. JónHelgMann. II, 138

A+

A+

þéna til e-s: *þjóna e-m, starfa hjá e-m* 20m hann þénti fyrtrum til Alfeigs þess er réð fyrir kristni á Einglandi. HKLCerpla., 214

þéna undir e-n (e. e-ð): *þjóna e-m (e. e-u), vinna undir stjórna e-s* 18f því under þenna Skap-Lost þjénaur svo mikeld Illþyde annarra Lasta. VidPost. II, 168 18s Löwenhielm sáráfatekur, sem svo gott sem þénaði undir þá frændur. HFinnsAnöv. LIX, 33 19f kvenmann, sem ekki gjörir annað en þéna undir skólakokkinn. JHúsf., 56 (1818) 19s Friðrik þénaði að vanda undir Jón, mokaði móldina og rétti hnausinn. FréggFyig., II, 11

þéna e-ð: *græða -ft-, vinna sér inn* :peninga- 19s hafði þá nauðg þénað um daginn fullar 700 krónur. Erióni., 188 19s Ég þénaði þó þrjúfú og sjö í anledning af kosningunum til rúdsdagsins í vor. JGubnSKth. I, 69 (1884) 20m Ég þekki skiptsjóra, sem síðastliðið ár þénuðu um 150 þúsund franka. Ágri. 1940, 155 20s til gamans atþugaði ég hve mikið ég mundi þéna í dag fyrir svona tórn. Múlap. 1974, 60

þéna e-ð e-u (e. e-m): *græða -ft- á e-u (e. e-m), vinna sér inn* :peninga- fyrir e-ð 19s það líða sem [kaupmenn] „þénuðu“

A+

A+

A+

þéna e-m til e-rs: *koma e-m að gagni við e-ð; stuðla að e-u* í þágu e-s 17m "hvad þjént til frelses þjéna kann: / það er fullkonnad seigir hann. HPPass. XLII, 14 18m eptir sem uppáfynding þessi þjénaur fólkinu til virðingur og þrýðis. JÓGrvKlím., 157 19m ef „filgjandi“ sögukonur „mætti þéna“ iður til skjemtanur. Fjóni., II, 2, 57 19m Ég meina að Edda miður eigi að þéna skáldinu til létts en til að þrýða hér og hvar verk hans. Rætur., 235 (1835) 19m því allir hlutir þjéna þeim til góðs, sem elska guð. MfTækif., 97 19s hagleiddi jafnframt, hvernig hin líftilmótelegustu arvik geta þénað mönnum til góðs. FréggFyig., I, 117

þéna sér e-ð: *vinna sér e-ð inn* 18m að dimissi frá skólanum mættu ekki þjéna sér braud hjá verðslögum mönnum skikkantlegum. ÞorkÆt. I, 26 (1740)

ÞÉNAST: e-m þénast e-að: e-r græðir e-ð 20fm svo mikið þénað aðist honum af alls konar sveitavöru á flakki sínu. Raubók. IV, 92

ÞÉNANDI: sem kemur (e-m) að gagni 19f [hún er] samilega kunnandi til vellestra almúga þénandi og nauðsynlegra kvenmannsverka. JónHelgMann. III, 91 (1818)

Figur 4: En þröng þvætten dær viser verbet þéna sådan som man tænker sig at teksten i den historiske ordbog kommer til at se ud.

#### 4.2.2 *ljúka*

Det andet eksempel på søgning i tekstsamlingen angår verbet *ljúka* (blive færdig med, slutte). Det var helt klart, at seddelsamlingens beskrivelse var langt fra at være fyldestgørende, selv om der var 126 eksempler, bl.a. manglede der eksempler på almindelige sætningskonstruktioner.

Verbet er forholdsvis almindeligt i sproget, og ved søgningen fandt man knapt 4.700 eksempler af forskellig art. Figur 5 viser eksemplernes inddeling m.h.t. sætningskonstruktion sammenlignet med eksempler fra seddelsamlingen. Ved søgning i tekstsamlingen fandt man 41 sætningskonstruktioner — heraf 21 der ikke forefandt i seddelsamlingen. Seddelsamlingen har derimod eksempler på 6 strukturer, der ikke kom frem ved søgning. Nogle af de strukturer der kom fra tekstsamlingen, fandtes i gamle tekster, som allerede var blevet exciperet, uden at de blev taget med. Eksempler på *ljúka við e-ð*, der er en meget almindelig sætningskonstruktion, fandtes bl.a. i Det nye Testamente fra 1540.

Foruden de nye strukturer gav søgningen mange ordforbindelser, der ikke fandtes i seddelsamlingen, og de eksempler der fandtes ved søgningen viser også betydeligt flere betydningsvarianter end man havde i forvejen.

Disse resultater viser, hvilket gavn man kan have af at bruge tekstsamlingen sammen med seddelsamlingen, idet man i hver kilde for sig finder materiale, der ikke findes i den anden.

### 5 Konklusion

Til sidst vil vi sammenfatte nogle hovedpunkter og gøre opmærksom på, hvilken lærdom man kan drage af dem.

- \* Leksikografisk instituts seddelsamlinger indeholder rigt varierede eksempler på brugen af et stort antal ord, og her findes materiale, som det er vanskeligt at komme til på en hurtig måde.
- \* Edb-analysen gør det muligt for os at fremstille en udførligt analyseret og systematisk ordbogstekst, der giver et godt billede af seddelsamlingens karakteristiske træk.
- \* Forskellige punkter må underbygges med flere kilder, og der mangler især eksempler på herskende sprogbrug.
- \* Tekstsamlingen har til trods for sin ufuldkommenhed vist sig at være gavnlig til at udfylde lakuner og rette de beskrivelser, der findes i seddelsamlingen, især når det drejer sig om de mere almindeligt forekommende fænomener.
- \* For at have virkeligt gavn af søgning i tekstsamlingen uden at arbejdet bliver alt for overvældende er det vigtigt, at man gør sig klart, hvilke fordele og ulemper og begrænsninger både metoden og selve samlingen har.
- \* Skal man bruge en tekstsamling målbevidst ved redaktionen af ordbogen, sådan som vi her har beskrevet, må samlingen være fornuftigt sammensat:

- den må ikke være for stor;
- teksterne må spænde over hele det tidsrum, der tages med i ordbogen og være så varierede, som det er muligt m.h.t. emner, stil, m.m.;
- samlingen burde være klassificeret efter forskellige træk, der er karakteristiske for teksten: alder, emne, stil, forfattere (alder, afstamning, samfundsklasse), m.m.;
- det skulle være muligt at søge såvel i hele samlingen som i visse afgrænsede dele af den (dele af teksterne, tekster indenfor visse kategorier, tekster af nord-islandske forfattere fra det 18. århundrede o.s.v.).

- \* Det skal nævnes, at tekstsøgning af den art, som vi har arbejdet med, efter al sandsynlighed kan give bedre resultater, når det drejer sig om verber, end når man har med andre store ordklasser at gøre. Verberne er forholdsvis få (knap 3% af ordforrådet i skriftsprogsamlingen og ca. 7% i en ny frekvensordbog over islandsk nutidssprog (Jörgen Pind et al., 1991)), men de fylder dog forholdsvis meget i tekster (20% af løbende ord i Frekvensordbogen). Man må derfor gå ud fra, at en tekstsamling af en nogenlunde størrelse indeholder eksempler på de fleste verber. Det kunne derimod vise sig at være vanskeligere at finde eksempler på visse andre substantiver end de hyppigst forekommende.
- \* Der er sikkert også brug for flere og mere varierede eksempler med verber end med de fleste andre ord, idet de forskellige sætningsomgivelser er så vigtige i forbindelse med beskrivelsen af deres betydning og brug. Det kan også tænkes, at lakunerne i de andre ordklasser er af en anden art, — at der fx snarere mangler nye ord end flere eksempler med de ord, man har i forvejen — og da er det ikke sikkert, at de her beskrevne metoder kan anvendes.
- \* Konklusionen er, at både gammeldags seddelsamlinger og maskinlæsbare tekstsamlinger har deres svage og stærke sider i forbindelse med den foreliggende opgave og kan derfor supplere hinanden. Samtidig skal det understreges, at seddelsamlingerne er det grundlag, som vi bygger på, og erfaringen har vist os, at de er uvurderlige, under de nuværende omstændigheder.

	<i>Skriftsprogsamling</i>	<i>Tekstsamling</i>
ljúka	1	2
+ af e-u	-	1
+ aftur	-	1
+ um e-ð/e-n	-	2
+ upp	2	32
+ upp fyrir e-u/e-m	-	11
+ úr e-u	2	21
+ út af við e-ð	1	-
+ við e-ð	-	396
ljúka e-ð	2	2
+ aftur	1	1
+ í e-u	-	1
+ um e-ð	1	-
+ upp	3	4
ljúka e-u	15	1.295
+ af	4	36
+ aftur	-	9
+ á e-ð	1	10
+ á e-m	1	-
+ saman	-	1
+ (í) sundur	6	10
+ til	-	1
+ upp	20	270
+ upp fyrir e-m	-	24
+ utan um e-ð	-	1
ljúka e-m e-ð	2	-
+ upp	-	1
ljúka sér	16	37
+ upp	-	2
e-u lýkur	5	2.285
+ á e-u	-	2
+ með e-u	-	50
+ upp	-	4
e-m lýkur	7	31
... lýkur	7	5
... yfir lýkur	-	61
ljúkast	2	4
+ að e-m	1	-
+ aftur	1	5
+ sundur	1	2
+ um e-ð/e-n	-	6
+ upp	6	52
+ upp fyrir e-u/e-m	1	6
+ utan um e-n	-	2
+ yfir e-ð	-	1
+ yfir e-n	-	1
e-u lýkst	1	-
lokinn	6	6
	126	4.697

Figur 5: Antallet eksempler på forskellige sætningskonstruktioner med verbet *ljúka* fra tekstsøgningen, sammenlignet med seddelsamlingen.

**Litteratur**

- Friðrik Magnússon. 1988. Hvað er títt? Tíðnikönnun Orðabókar Háskólans. *Orð og tunga* I:1-49.
- Jón Hilmar Jónsson. 1988. Sagnorðagreining Orðabókar Háskólans. *Orð og tunga* I:123-174.
- Jón Hilmar Jónsson. 1990. A Standardized Dictionary of Icelandic Verbs. I: Jörgen Pind & Eiríkur Rögnvaldsson, red. *Papers from the Seventh Scandinavian Conference of Computational Linguistics, Reykjavík 1989*: 268-296. Institute of Lexicography, Reykjavík.
- Jörgen Pind, Friðrik Magnússon og Stefán Briem. 1991. *Íslensk orðtíðnibók*. Orðabók Háskólans, Reykjavík.