

Eiríkur Rögnvaldsson, Órnólfur Thorsson

Fornir textar í tölvubanka

I. Textar

Á undanföllum árum hafa allnaegir textar fornrita komið út í lestrarútgáfum sem ætlaðar eru almenningi, einkum á vegum bókaforlagsins Svart á hvítu en einnig hjá Máli og menningu. Þessar útgáfur hafa að sjálfsögðu verið unnar í tölvu og eru textarnir því aðgengilegir til frekari vinnslu á tölvutæku formi. Þeir textar sem um ræðir eru þessir:

A. Íslendinga sögur

1. Bandamanna saga (Möðruvallabók)
2. Bandamanna saga (Koningsbók)
3. Bæðar saga Snæfellsáss
4. Bjarnar saga Hítarlakappa
5. Brennu-Njáls saga
6. Droplaugarsona saga
7. Egils saga Skalla-Grímssonar
8. Eiríks saga rauða
9. Eyrbyggja saga
10. Ævi Snorra goða

- 11 Finnboga saga ramma
- 12 Fjötsdæla saga
- 13 Flóamanna saga (auk kafla úr styttri gerð)
- 14 Fóstbræðra saga (auk kafla úr Flateyjarbók)
- 15 Gísla saga Súrssonar (styttri gerð)
- 16 Gísla saga Súrssonar (lengri gerð)
- 17 Grænis saga
- 18 Grænlandinga saga
- 19 Grænlandinga þáttur
- 20 Gull-Þóris saga
- 21 Gunnars saga Keldugnópsfílls (AM 496 4to auk brota úr AM 551 i 4to)
- 22 Gunnlaugs saga ormsunga
- 23 Hallfreðar saga (Móðruvallabók)
- 24 Hallfreðar saga (úr Ólafs sögu Tryggvasonar hinnu mestu)
- 25 Harðar saga og Hólmverja (AM 556 a 4to auk brota úr AM 564 a 4to)
- 26 Hávarðar saga Ísfirðings
- 27 Heiðarvíga saga
- 28 Hrafnkels saga
- 29 Hænsna-Þóris saga
- 30 Kjalnesinga saga
- 31 Jökuls þáttur Búasonar
- 32 Kormáks saga
- 33 Króka-Refs saga
- 34 Laxdæla saga
- 35 Ljósveitinga saga (C-gerð auk kafla úr A-gerð)
- 36 Reykdæla saga og Víga-Skátu
- 37 Svarfdæla saga
- 38 Valla-Ljóts saga
- 39 Vatnsdæla saga
- 40 Víga-Glúms saga
- 41 Víglundar saga
- 42 Vopnfirðinga saga
- 43 Þórðar saga hreðu
- 44 Þorsteins saga hvíta
- 45 Þorsteins saga Síðu-Hallssonar
- 46 Ölkofra saga.

B. Íslendinga þættir

Arnórs þ. jarlaskálds, Auðunar þ. vestfirska, Bergþúa þ., Bollu þ. Bollasonar, Brandkrossa þ., Brandís þ. órva, Draumur Þorsteins Síðu-Hallssonar, Egils þ. Síðu-Hallssonar, Einars þ. Skúlasonar, Gísls þ. Illugasonar (efir Hulðu og Hrokkinskinnu, úr B-gerð Jóns sögu helga, úr A-gerð Jóns sögu helga), Gull-Ásu-Þórðar þ. (efir AM 518 4to, efir Morkinskinna), Gunnars þ. Þórandabarna, Halldórs þ. Snorrasonar hinn fyrri, Halldórs þ. Snorrasonar hinn síðari, Hrafn þ. Guðrúnarsonar, Hreiðars þ., Hrómundar þ. halta, Íslending þ. sögufróða, Ívars þ. Ingimundarsonar, Kamlbúa þ., Mána þ. skálds, Odds þ. Ófeigssonar, Orms þ. Sörolfssonar, Ótars þ. svarta (efir Flateyjarbók, Tómassskinna, Bæjarbók, Bergsbók), Sneglu-Halla þ. (efir Morkinskinna, Flateyjarbók), Stjörnu-Ódda draumur, Stáfs þ. hinn skemmi, Stáfs þ. hinn meiri, Svada þ. og Arnórs kerlingarnefs, Þérandi þ. og Þórhalls, Þórarins þ. Nefjólssonar, Þórarins þ. ofsa, Þórarins þ. stutfeldar, Þorgríms þ. Hallasonar, Þórhalls þ. knapps, Þorkéls þ. jarlaskálds, Þormóðar þ. (efir Flateyjarbók, efir Fóstbræðra

sögu), Þorsteins þ. Austfirðings, Þorsteins þ. forvitna, Þorsteins þ. Síðu-Hallssonar (efir Morkinskinna, efir Flateyjarbók), Þorsteins þ. skeiks, Þorsteins þ. stangarhöggs, Þorsteins þ. sögufróða, Þorsteins þ. tjaldsteðings, Þorsteins þ. uxafóts, Þorvalds þ. tasalda, Þorvalds þ. víðföria, Þorvarðar þ. krákuvefs, Ógmundar þ. dyttis.

C. Sturlunga saga

D. Aðrir textar

- 1 Árna saga biskups
- 2 Edda Soorra Sturlusonar
- 3 Hrafn saga Sveinbjarnarsonar hin séstaka
- 4 Íslendingabók Ara fróða
- 5 Leiðarvísir Nikulásar Bergssonar ábóta á Þverá
- 6 Ragnars saga loðbrókar
- 7 Tristrams saga
- 8 Veraldar saga
- 9 Völsunga saga.

(Að auki eru málfraeðingirbirnar allar tiltækar á sílva, verulegur hluti Heimskringla og fáeinir þættir aðrir en þeir sem taldir voru hér að ofan.)

II. Úrvinnsla

Nú er í gangi margvísleg úrvinnsla úr þessum textum, sem ekki hefði verið möguleg nema hafa aðgang að þeim í tölvutækri mynd. Þegar hefur verið gengið frá orðstöðuskrám Íslendinga sagna, bæði einstakra texta og eins hefur verið smíðaður lykkill allra sagna þar sem unnt er að bera saman tölmi orðmynda í textum. Í þessum skráum koma fram upplýsingar um fjölda orðmynda, lesmálsorða og lesmálsstákna þannig að aðvelt er að bera saman orðmyndir í nútímafrönsku og fornúáli. Jafnframt höfum við gert bakstöðuskrá allra Íslendinga sagna sem ætti að njúast vel við rannsóknir á orðmyndun og beygingarfræði. Heildarfjöldi orðmynda í Íslendinga sögum (þeim textum sem taldir eru hér að framan) er tæplega 35.000.

Þá höfum við undanfarið í félagi við þau Bergljótu Kristjánsdóttur, Braga Halldórssón, Guðrúnu Ingólfssdóttur, Jón Torfason og Sverri Tómasson unnið að gerð orðstöðulykils (KWIC-concordance) fyrir Íslendinga sögur. Verkefnið er vel á veg komið og stefnt er að því að lykkill verði fullbúinn fyrir framót. Um er að ræða lemnaða skrá tæplega 13.000 orða (fellið eru brott manna-nöfn og staðahetit) og heildarfjöldi dæma er 7-800.000.

Verkið er unnið þannig að öllum textunum er steypt saman í eina skrá (sem er tæplega 5 megabæri að stærð). Síðan er sérstaki forrit (Word Cruncher frá Electronic Text Corporation í Utah) látið skrifa í textaskrá (ASCII-skrá) öll dæmi um hverja orðmynd, með u.þ.b. 40 bókstafa umhverfi í hvora átt og tilvísun í sögu, kafla og blaðsíðu. Brot úr slíkrí skrá er sýnt hér að neðan.

		I*þak (1)	
Laed	64:1634	stóðu út af ásendamir og var sinat I*þak	á húsinu og ekki gróð. Þá mælt
I*þakið (8)			
Njálá	77:213	eigi fyrir en þeir höfðu undið allt I*þakið	af skálanum Gunnar skýtur þá af þurt og bræddur víðurinn en nætum I*þakið
Egla	22:392	húsin að drýpa sem líklegi var er I*þakið	un mæð. Þóðitúur það menn sína
Gars	13:864	sva mikili að þegar tekur af skálanum I*þakið	tók að roma. Gali spratt upp
Gsil	18:917	og hver að dönum þar til er allt er I*þakið	dönum megin og hver að dönum þar til
Gsil	18:917	gull í skudina. Þetta herbergi var I*þakið	af húsinu. Vættali fylgdi hér og svo
Vgl	2:1957		biði og steint allt innan. Skúþgarður
I*þakka (14)			
Njálá	44:177	þá til nokkurs." "Núval skal eg I*þakka	þeim er þeir segja mér heim víg
Egla	16:384	öllu best varð er eg hef gert til I*þakka	yðvata." Konungur sagði að ekki
Finnv	31:684	þar hafa gefið sér, mág sínu. Þeir I*þakka	horum hæfða vel og ríða heim. Hæði
Fósth	5:786	þeir hafi með þeim Vemund. Þeir I*þakka	henni sit lílag er hún hafði þeim
Fósth	5:786	Og er þeir voru búin til ferðar þá I*þakka	þeir henni þarvíst sína og allan
Fósth	20:818	búin þá fara þeir á konungs fund og I*þakka	alla vingan þá er hann hafði þeim
GlaS	9:860	það var vart að þorkell var vartur að I*þakka	bróður sínum verkið en nú er hann
Gret	91:1093	ekki vísar um aturkomu sína "Vil eg nú I*þakka	yður öllum," sagði Þorsteinn, "Hversu
GurtV	11:1160	horum því að eg á horum líkjst að I*þakka."	Kat sagði "Eg vil lík að sjá
Hæðá	6:1205	þar nær sem konungurinn er. Vil eg I*þakka	horum hótivist." Þá kom það fram
Laed	40:1596	hann víðt. Þá mælt Kjetan: "I*þakka	vil þú sér yður konungur er þér gefið
Svaf	8:1789	og ríðu í braut. Nú I*þakka	menn Þorsteinn fyrir er hann sókst
Vgl	4:1911	þetta vil eg eigi veita." Þeir I*þakka	horum vel og þótti þeim þó ekki í
I*þakkað (1)			
Egla	65:468	vonu. Var þetta verk horum allvel I*þakkað.	Þá kvæð Egil:

Að þessu loknu er textaskráin tekin inn í WordPerfect-ráttvinnsluforritið og búið undir lemmun. Í því felst m.a. að fleiri orðid er feitleitnað og dæmum um hverja orðmynd er raðað í stafrófsröð þess sem á eftir kemur. Að þessu loknu lítur dæmin hér að ofan svoa út:

þak (1)		þak	
Laed	64:1634	stóðu út af ásendamir og var sinat	á húsinu og ekki gróð. Þá mælt
þakið (8)			
Gsil	18:917	og hver að dönum þar til er allt er	af húsinu. Vættali fylgdi hér og svo
Njálá	77:213	eigi fyrir en þeir höfðu undið allt	þakið
Vgl	2:1957	gull í skudina. Þetta herbergi var	þakið
GlaS	13:864	húsin að drýpa sem líklegi var er	þakið
Egla	22:392	þurt og bræddur víðurinn en nætum	þakið
Gsil	18:917	sva mikili að þegar tekur af skálanum	þakið
þakka (14)			
Fósth	20:818	búin þá fara þeir á konungs fund og	þakka
GlaS	9:860	það var vart að þorkell var vartur að	þakka
Fósth	5:786	þeir hafi með þeim Vemund. Þeir	þakka
Hæðá	6:1205	þar hafa gefið sér, mág sínu. Þeir	þakka
Finnv	31:684	þeir hafi með þeim Vemund. Þeir	þakka
Hæðá	7:1235	þar nær sem konungurinn er. Vil eg	þakka
Vgl	4:1911	þetta vil eg eigi veita." Þeir	þakka
GurtV	11:1160	horum því að eg á horum líkjst að	þakka."
Svaf	8:1789	og ríðu í braut. Nú	þakka
Laed	40:1596	hann víðt. Þá mælt Kjetan:	þakka
Gret	91:1093	ekki vísar um aturkomu sína "Vil eg nú	þakka
Egla	16:384	öllu best varð er eg hef gert til	þakka
Njálá	44:177	þá til nokkurs." "Núval skal eg	þakka
Fósth	5:786	Og er þeir voru búin til ferðar þá	þakka
þakkað (1)			
Egla	65:468	vonu. Var þetta verk horum allvel	þakkað.
Þá kvæð Egil:			

Fram að þessu hefur vinnslan verið að mestu vélræn, en ná tekur við handvinnan: sjálf lemmunin. Í henni felst það að steypt er saman öllum beygingarmyndum sama orðs, og samhjöfða orðmyndir sem tilheyra mismunandi flectingum (lexemum) eru greindar sundur, auk þess sem dæmin eru öll greind í orðflokk. Þetta er gífurlega mikil vinna, og ekki alltaf auðveld, því að ofi leikur vafi á því hvernig greina skuli einstakar orðmyndir (hvar mörkin liggja milli lýsingarorða og atviksorða, hvenær greina beri stífröð föll nafnorða sem atviksorð o.s.frv.). Dæmin sem áður var tekið sem dæmi lítur svoa út að lokinni lemmun:

þak hk; þak (1); þakið (4)		þak	
Laed	64:1634	stóðu út af ásendamir og var sinat	á húsinu og ekki gróð. Þá mælt
Gsil	18:917	og hver að dönum þar til er allt er	þakið
Njálá	77:213	eigi fyrir en þeir höfðu undið allt	þakið
GlaS	13:864	húsin að drýpa sem líklegi var er	þakið
Gsil	18:917	sva mikili að þegar tekur af skálanum	þakið
þakka sv; þakka (13); þakkað (1)			
Fósth	20:818	búin þá fara þeir á konungs fund og	þakka
GlaS	9:860	það var vart að þorkell var vartur að	þakka
Fósth	5:786	þeir hafi með þeim Vemund. Þeir	þakka
Hæðá	6:1205	þar hafa gefið sér, mág sínu. Þeir	þakka
Hæðá	7:1235	þar nær sem konungurinn er. Vil eg	þakka
Vgl	4:1911	þetta vil eg eigi veita." Þeir	þakka
GurtV	11:1160	horum því að eg á horum líkjst að	þakka."
Svaf	8:1789	og ríðu í braut. Nú	þakka
Laed	40:1596	hann víðt. Þá mælt Kjetan:	þakka
Gret	91:1093	ekki vísar um aturkomu sína "Vil eg nú	þakka
Njálá	44:177	þá til nokkurs." "Núval skal eg	þakka
Fósth	5:786	Og er þeir voru búin til ferðar þá	þakka
Egla	65:468	vonu. Var þetta verk horum allvel	þakkað.
þakka sv; þakka (2)			
Vgl	2:1957	gull í skudina. Þetta herbergi var	þakið
Egla	22:392	þurt og bræddur víðurinn en nætum	þakið
þakka kv; þakka (1)			
Egla	16:384	öllu best varð er eg hef gert til	þakka
yðvata." Konungur sagði að ekki			

Hér hefur myndunum þak og þakið af no. þak verið steypt saman, og eins þakka og þakkaði, af so. þakka. Þakið getur hins vegar einnig verið beygingarmynd af so. þekja, og þakka getur verið mynd af no. þökk.

Síðan eiga auðvitað eftir að bætast við fleiri beygingarmyndir af flestum orðanna, og önnur orð að koma þarna inn á milli, því að ekki er svo vel að allar myndir séu í stafrófsröð. Til að sýna dæmi um hve beygingarmyndir sama orðs geta orðið margar og fjölbreyttar má taka dæmi af sögninni fara:

fara so; far (135); fara (2308); farast (13); farðu (8); fari (124); farið (554); farin (8); farinn (72); farir (76); farist (10); farnar (1); farnir (33); fer (822); ferð (41); ferðu (1); ferst (9); fór (2117); fóst (36); fórstu (13); fóru (988); fóruð (8); fórum (3); fórust (3); fær (48); færð (4); færum (7); fari (133); ferir (44); fernum (4); ferum (68)

Síðan er skráin prentuð á geislaprentara og prófarkalesin. Prófarkalesurinn er tímafreakasti hluti vefsins, enda hefur verið reynt að vanda til hans svo sem kostur er. Sú greining sem unnin var við lemmunina og áður er lýst er öll endurskoðuð, og stuðst við tiltekar fornalds- og nútímamálforða-bækur. Hér koma upp ótal vafaatriði sem leysa þarf úr. Hvernig á að fara með tvímyndir (t.d. aðfangagadagur/afangagadagur, -igur/-ugur, áðláxi/lexi, o.s.frv.)? Hvernig á að fara að þegar aukafallsmynd getur tilheyrt tveim uppflectingum (t.d. afturgangur, en bæði er til afturganga kv. og afturgangur kk.)? Á að flokka lýsingarhætti undir sagnir, lýsingarorð eða hafa þá sér? Á að gefa upp (búa til) nefniföll einöflu af nafnorðum sem aðeins koma fyrir í fleiritölu eða aukafalli? Þetta er

aðeins brot af þeim vandamálum sem leysa þarf úr. Í mörgum tilvikum er ónægulegt að leysa þau svo að óvirkni sé; en reynt hefur verið að koma upp skýrum vinnureglum þannig að væntanlegir notendur geti auðveldlega áttað sig á því við hvað er miðað.

Við teljum að orðstöðulykill að Íslendinga sögum eigi að geta orðið ómetanlegt hjálparteki málfræðinga, bókmenntafræðinga, sagnfræðinga, þjóðfræðinga, þjóðháttafræðinga og annarra sem fást við rannsóknir á íslenskum miðaldatextum. Textinn er að sönnu ekki stafréttur en það kemur lítt að sék við aþugun á einstökum orðum og stöðu þeirra í setningum, fallstjórn sagna, notkun forsetninga, tíðni orða og orðflokka, merkingarbreytingum, o.s.frv. Má sem dæmi nefna að nú er hópur nemenda á kandidatsstigi í íslenski málfræði að vinna að aþugunum á setningagerð forn máls, þar sem orðstöðulykillinn er aðalhljálpargagnið.

Fyrir bókmenntamenn eru lykjar af þessum toga mákið þarfapung. Til að mynda má finna á skjótu augabragði öll dæmi um orðhluta, orð eða orðasambönd og tíðni þeirra í einstökum textum. Þá er unnt að glöggva sig á þeim orðaforða sem sérstakur er fyrir hvern texta (t.d. sérórðaforða Egils sögu eða Þórðar sögu kakala, Víglundar sögu eða Vatnsdæla sögu) og jafnframt er hægt að gera tæmandi skrár yfir þann orðaforða sem sameiginlegur er t.d. einhverjum tveimur sögum (sameiginlegur orðaforði Finnboga sögu og Hávarðar sögu, Grettis sögu og Gísla sögu, o.s.frv.) eða fleiri (orðaforði Gísla sögu, Harðar sögu og Grettis sögu, o.s.frv.). Auðvelt er að finna formálur og ýmiss konar tengsl í orðafari.

Orðstöðulykillinn verður prentaður út í einu eða tveimur eintökum og varðveittur á Stofnan Árna Magnússonar til afnota fyrir þá fræðimenn sem þar sitja. Hann verður jafnframt varðveittur á tölvutæku formi og stefnt að því að hann verði tiltækur dhugamönnnum með einföldum hætti. Útlökab er að gefa hann út prentaðan, því að hann yrði a.m.k. 8 þúsund blaðsíður; og útgáfa á disklingum kemur ekki heldur til greina, því að lykillinn er a.m.k. 80 megabæti. Því kemur helst til greina útgáfa á geisladiski eða fisjum. Ákvörðun hefur þó ekki verið tekin um þetta, og verður ekki gert fyrir en verkinu er að fallu lokið.

Þá verða helstu niðurstöður þessarar vinnu kynntar í bók sem Árnastofnun hyggst gefa út á næsta ári.