

Baldur Jónsson:

Stærð orðaforðans í orðabók Blöndals

1

Fyrir rúmum 30 árum skrifaði ég ritdóm um viðbættinn við orðabók Blöndals, sem út kom 1963, og reyndi þá að gera mér grein fyrir stærð orðaforðans með því að telja flettiorðin á allmörgum blaðsíðum og álykta síðan út frá því. Ég sagði þá m.a. (*Skímir* 138. ár 1964, bls. 246):

Talið hefir verið, að í Bl. væru alls um 120 þús. orð, en samkvæmt minni talningu eru þau milli 110 og 115 þús. Viðbættirinn nýi er 200 bls. eða tæplega fimmtungur á við Bl. að blaðsíðutali, en mér telst til, að í honum séu um 41000 orð, svo að hann jafngildir þá meira en þriðjung Bl. að orðafjölda.

Þetta var löngu áður en við fórum að nota tölvur við orðabókargerð. En nú hefi ég einnig gert vélræna talningu á orðaforða Blöndalsbókar og getur verið forvitnilegt að bera niðurstöður saman. Áður en að því kemur skal gerð ofurlítill grein fyrir grundvelli hinnar vélrænu talningar.

Upp úr 1980 hóf ég að undirbúa tölvuskráningu flettiorða úr orðabók Blöndals og hugðist nota þann orðaforða, ásamt viðbótarefni, í stóra íslenska orðaskrá, sem ekki verður fjölyrt um hér. Ég hafði svolitla fjárveitingu árlega til þessarar skráningar og notaði hana til að greiða aðstoðarfólki laun og borga fyrir tölvunotkun. Þegar fjárveitingin var tekin af fyrir nokkrum árum stóðst á endum að skráningu flettiorða var lokið úr aðalsafni Blöndalsorðabókar og úr viðbættinum frá 1963, og náðist með herkjum að lesa saman tölvusafnið og hinar prentuðu bækur og færa inn lagfæringar. Svo á því að heita að ég hafi tölvutæk öll flettiorð Blöndalsbókar eins og hún leggur sig.

Ýmsan fyrirvara verður þó að hafa á þessu, og er þá rétt að gera um leið nokkra grein fyrir tölvuskráningunni, þótt ekki varði það allt stærð orðaforðans.

Fyrst er þess að geta að ýmislegt er enn ólagfært eftir tölvuskráninguna þrátt fyrir samlesturinn. Ýmiss konar villur slæddust inn vegna þess að verkið var unnið á löngum tíma og ritarar komu og fóru. Afstaða mín til sumra atriða breyttist á tímabilinu, og stundum fengu ritarar ekki nægilega skýr fyrirmæli um hvernig að skyldi fara. Því vantar nokkuð á að fullt samræmi sé í allri skráningu.

Í umræðu hér á eftir verða einstakir hlutar orðabókarinnar auðkenndir með skammstöfunum. Aðalsafnið sem út kom 1920–1924 er nefnt GBl., meginhluti þess Bl. og viðbæturinn Vb. Nýi viðbæturinn frá 1963 er nefndur Nvb. og allt safnið í heild sinni ABl.

Sumum flettiorðum var sleppt viljandi, t.d. beygingarmyndum sterkra sagna, sem allviða eru í bókinni. Nafnháttur var látinn nægja eða nefnifall fallorða þegar því var að skipta. Undantekning var þó gerð um lýsingarhætti sagnorða, sem oft eru sérstök flettiorð. Í Bl. töldust 299 slík dæmi, ekkert í Vb. og 5 í Nvb., eða samtals 304 flettiorð. Þetta eru oftast lýsingarhættir sterkra eða óreglulegra sagna, t.d. *brugðinn af bregða* og *róinn af róa*, en einnig af veikum sögnum, t.d. *barinn af berja* og *hulinn af hylja*. Þessar tölur hafa engin áhrif á heildarniðurstöður. Allar þessar orðmyndir teljast til sagnorða sem skráð eru í nafnhætti.

Allmörg af flettiorðum bókarinnar eru ekki heil orð, heldur forliðir eða viðliðir, og voru þeir skráðir. Þeir töldust vera 112 í Bl., 4 í Vb., þ.e. 116 alls í GBl., og 655 í Nvb., samtals 771 í allri orðabókinni. Athygli vekur hve margir þeir eru í nýja viðbætinum. Það er vegna þess að í sparnaðar skyni voru samsetningar með sama forlið ekki nærri alltaf skrifaðar fullum stöfum. Þegar efni var tölvuskrað úr viðbætinum hafa þessir forliðir verið skráðir sérstaklega en auk þess allar samsetningar sem þá hafa. Þannig er að vissu leyti um tvítekningu að ræða. Þessari tegund flettiorða er haldið utan við heildarniðurstöður talningar.

Stafsetningu var ætlunin að færa til nútímahorfs, en nokkur misbrestur varð á því.

Auk þess er talsvert um rangar færslur, t.d. þannig að vitneskja um orðflokk eða tölu hefir lent í röngu sviði. Flettiorð gæti þá t.d. heitið „adj.“ (adjektiv) eða vitneskja um tölu, „sg.“ (singular) eða „pl.“ (plural), lent í sviði orðflokks. Þessi dæmi eru að vísu tilbúin, en skýra hvað við er átt.

Ýmiss konar álitamál voru látin í eins konar ruslakistu sem kölluð var „ATHS“, og þar hefir fleira lent en þangað átti að fara. Enn er allmikið verk eftir við að greiða úr þessu öllu.

Eins og ráða má af þessu var ýmiss konar vitneskja skráð um hvert flettiorð. Helst er að nefna orðflokk og beygingarendingar þegar svo bar undir. Orðflokktákn vantar stundum og raunar oft í nýja viðbætinum frá 1963, og var þá reynt að bæta úr því.

Í gamla safninu frá 1920–1924 eru sum flettiorð auðkennd með sérstöku merki fyrir framan orðið og segja til um málsnið. Slík merki eru alls fjögur og tákna: „úrelt orð“, „skáldskaparorð“, „nýyrði“ eða „aðskotaorð“ (sem er ekki viðurkennt í ritmáli). Öll þessi atriði voru tölvuskrað hvert í sitt svið.

Orð sem töldust staðbundin voru auðkennd með skammstöfunum, svo sem „Af.“ = Austfirðir, „Árn.“ = Árnessýsla o.s.frv. Algengt er að slík auðkenning eigi ekki við

flettiorðið sjálft heldur einhverja merkingu þess, þegar merkingar eru fleiri en ein. En úr því að merkingar orða voru ekki tölvuskráðar var auðkenning af þessu tagi því aðeins skrásett að hún ætti við flettiorðið sjálft.

Þessi greinargerð verður að nægja og þeir fyrirvarar sem minnst hefir verið á. Fleira mætti nefna, en aðalatriðið er að vélræna talningin nær ekki beinlínis til orðabókar Blöndals heldur til hins tölvuskráða orðasafns sem enn þarf ýmissa lagfæringa við.

2

	Bl.	Vb.	GBL.	Nvb.	ABL.
nafnorð					
kk.	27.581	979	28.560	10.637	39.197
kv.	26.809	891	27.700	13.577	41.277
hk.	22.175	819	22.994	9.811	32.805
alls	76.565	2.689	79.254	34.025	113.279
sérnöfn	1.196	19	1.215	1	1.216
lýsingarorð	22.133	893	23.026	4.557	27.583
sagnorð	7.487	308	7.795	822	8.617
atviksorð	2.456	57	2.513	136	2.649
upphrópanir	136	3	139	13	152
forsetningar	102		102	3	105
fornöfn	66		66		66
töluorð	90	2	92		92
samtengingar	30		30		30
greinir	2		2		2
nafnháttarmerki	1		1		1
óflokkuð	138	219	357	5	362
Alls	110.402	4.190	114.592	39.562	154.154

Tafla 1: Greining orðafordans í orðabók Blöndals

Á töflunni hér fyrir ofan sjást niðurstöður talningarinnar. Magnús Gíslason reikni-fræðingur var mér hjálplegur bæði við talninguna og gerð töflunnar. Í fremsta dálki eru orðflokkar. Í næsta dálki, sem merktur er „Bl.“, eru tölur úr aðalsafninu 1920–1924. Í dálki merktum „Vb.“ eru orð sem tekin voru úr viðbætinum í þeirri útgáfu, og samtölur þessara tveggja dálka eru birtar undir „GBL.“. Því næst koma tölur úr nýja viðbætinum, „Nvb.“, sem út kom 1963, og loks eru heildartölur birtar í síðasta dálki, „ABL.“. Neðst

eru samtölur úr hverjum dálki, og má þar sjá stærð hvers bókarhluta og bókarinnar í heild.

Hafa ber í huga að í Vb. er slæðingur af orðum sem einnig eru í Bl., en endurtekin vegna viðbótarmarkeringar eða leiðréttingar. Slík orð eru einungis talin til orðaforða Bl. Í Nvb. eru enn fleiri orð endurtekin úr GBl. af sömu ástæðum, en um þau gildir hið sama; þau eru talin til orðaforða Bl. eða Vb., en ekki Nvb.

Í gömlu útgáfunni eru samtals hátt í 115 þús. orð, tæp 40 þús. orð í nýja viðbætinum og ríflega 154 þús. orð í Blöndalsorðabók allri. Þetta er langstærsta orðabók um íslenskt mál sem prentuð hefir verið. Hafa má til hliðsjónar að Árni Böðvarsson taldi vera um 65 þús. flettiorð í fyrri útgáfu orðabókar sinnar (*Íslenskrar orðabókar handa skólum og almennungi*) 1963 og um 85 þús. í síðari útgáfunni 1983 (sjá formála þeirrar útgáfu, bls. V og IX).

Enn fremur vekur athygli að dálítið efni er enn óflokkað, 362 orð alls. Þetta er það sem fór í ruslakistuna „ATHS“ sem fyrr var nefnd. Greining á þessum orðum myndi auðvitað raska eitthvað tölum á öðrum stöðum í töflunni, en væntanlega kemur sú röskun víða niður og breytir varla hlutföllum svo að máli skipti né heildartölum.

Að öðru leyti má láta lesendum eftir að lesa úr töflunni, hverjum eftir sínu höfði. Hér skal þó drepið á tvö atriði sem vöktu athygli mína.

Á töflunni kemur býsna glögglega fram skiptingin í lokaða orðflokka og opna. Ef sérnöfnum er steyppt saman við nafnorð teljast orðflokkanir ellefu. Af þeim mega sex heita lokaðir: fornöfn, forsetningar, greinir, nafnháttarmerki, samtengingar og töluorð. Í Nvb. eru þó þrjár forsetningar sem ekki voru í GBl. Opnu eða frjóu flokkarnir eru fimm: atviksorð, lýsingarorð, nafnorð, sagnorð og upphrópanir.

Úr lokuðu orðflokkanum sex koma aðeins 296 flettiorð eða 0,19% af orðaforða bókarinnar, en 153.858 orð eða 99,81% úr lokuðu flokkunum fimm. Óflokkuð eru 362 orð eða 0,23%. Stærstu flokkarnir eru:

nafnorð (með sérnöfnum)	114.495	74,27%
lýsingarorð	27.593	17,90%
sagnorð	8.617	5,59%
atviksorð	2.649	1,72%
Alls	153.354	99,48%

Tafla 2: Orðflokkskipting í orðabók Blöndals

Þessir fjórir orðflokkar ná m.ö.o. yfir nærri 99,5% af öllum orðaforða Blöndalsorðabókar, og um ¾ af öllum flettiorðum í bókinni eru nafnorð.

Hitt atriðið sem vakti sérstaka athygli mína er að hlutföllin milli kynflokka nafnorða hafa breyst frá GBl. til Nvb. Í GBl. eru karlkynsorðin flest, en í Nvb. eru kvenkynsorðin svo miklu fleiri (41.277) að þau hafa vinninginn í heildartölunum. Hlutföllin eru nánara tiltekið þannig:

	GBI.	Nvb.	ABI.
karlkynsorð	36%	31%	34,6%
kvenkynsorð	35%	40%	36,4%
hvorugkynsorð	29%	29%	29,0%

Tafla 3: Kyn nafnorða í orðabók Blöndals

Hvorugkynsorðin halda sinni hlutdeild óbreyttri, en kvenkynsorðum hefir fjölgað á kostnað karlkynsorða. Þetta bendir óneitanlega til þess að kvenkynsorðin hafi verið í sókn um miðbik aldarinnar. Hér skal þó ekkert fullyrt um slíkar hræringar í málinu, en ef til vill er nærtækast að ætla að kvenkynsorð hafi oftast orðið viðliðir í nýjum samsetningum en karlkynsorð. Í Nvb. er mjög mikið um samsetningar.

Þrátt fyrir allt er ég trúaður á að tölurnar séu nægilega traustar til að gefa sæmilega sanna mynd af stærð orðabókarinnar og stærðarhlutföllum. Heildartölurnar fara býsna nærri þeim niðurstöðum sem fengust við að telja í höndunum forðum.