

Samanburður á stafkaflanum D í Ritmálsskrá Orðabókar Háskólans (RMS), orðalista úr formálsorðabókinni í Kaupmannahöfn (ONP) og orðalista úr Orðabók Jóns Ólafssonar úr Grunnavík

Kristín Bjarnadóttir

8.5.2007

Erindi flutt á fundi Góðvina Grunnavíkur-Jóns.

0 Inngangur

Þetta eru drög að punktum um samkeyrslu á stafkaflanum d úr RMS, ONP og JÓIGrv. Skráin er óyfi rlesin þar sem hún er hvorki uppsett á viðunandi máta né hefur hún verið prentuð út til yfi rlestrar. Úr þessu verður bætt síðar og skjalið þá sent þeim sem vilja. Athugasemdir eru vel þagnar!

1 Gögnin

Upprunalegu gögnin í þessum samanburði voru listi um flettiorð úr d-i úr ONP sem fenginn var hjá Bent Chr. Jacobsen (1.545 orð) og listi úr Ritmálsskrá Orðabókar Háskólans (16.815 orð). Síðar var orðalistum úr seðlasafni Jakobs Benediktssonar úr Orðabók Jóns Ólafssonar úr Grunnavík (JÓIGrv.) og orðalista úr Talmálssafni (TMS) Orðabókarinnar bætt við. Dæmi fyrstu þremur heimildunum fylgja hér til að sýna formið á skránum; 10 línur valdar af handahófi. Síðasta uppflettiorðið er það sama í öllum skráum.

<i>ONP:</i>	<i>RMS:</i>	<i>JÓIGrv.:</i>
dáðafullr,adj.,	dáðrakkur adj 18s	dáðfyndnislæti npl
dáðalauss,adj.,	dáðreifður adj 19f	dáðgóður adj
dáðaskáld,sb.,n.	dáðrekki n n 19m	dáði m
dáðgjarn,adj.,	dáðreyndur adj 19m	dáðigur adj
dáðlauss,adj.,	dáðríki n n 19ms	dáðindi
dáðleysi,sb.,n.	dáðríkismaður n m 19s	dáðindismatur m
dáðleysingi,sb.,m.	dáðríkur adj 18s	dáðlaus adj
daðra,vb.,	dáðrómur n m 18s	dáðlegur adj
dáðrakk,adj.,	dáðsamlega adv 17s	dáðleysi n
dáðsamliqr,adj.,	dáðsamlegur adj 16s	dáðsamlegur adj

2 Normalísing og samræming á listunum

Núttímamálsstafsetningin í Ritmálsskrá Orðabókarinnar (RMS) var lögð til grundvallar í samkeyrslunni og vegna þessa þurfti að normalísera stafsetninguna í listanum úr ONP. Orðalistar úr JÓIGrv. og TMS eru að mestu normalíseraðir á sama hátt og RMS, enda eru hausar á seðlunum í þessum söfnum settir upp af Orðabókarmönnum. Að hluta var normalísing á ONP gerð með strengjaskiptingum og að hluta „í höndunum“, þ.e. með því að skoða hverja breytingu. Þessu efni var haldið til haga til þess að hægt væri að nota það síðar í skriftur til að vinna úr öðrum stafköflum, ef til þess kemur.

Þá var orðflokkamerkingum í ONP breytt til samræmis við merkingar í RMS, röð á sviðum breytt og heimildaskammstöfunum bætt við. Hér eru bútar úr samkeyrsluskránum sem eru sambærilegir við dæmin hér á undan en stafrófsröðin hér er önnur, sbr. það að *dadra* fellur út í ONP-listanum:

ONP:		RMS:	
dáð_n_f	ONP:dáð	dáðrakkur_adj_	RMS:18s
dáðafullur_adj_	ONP:dáðafullr	dáðreifður_adj_	RMS:19f
dáðalaus_adj_	ONP:dáðalauss	dáðrekkki_n_n	RMS:19m
dáðaskáld_n_n	ONP:dáðaskáld	dáðreyndur_adj_	RMS:19m
dáðgjarn_adj_	ONP:dáðgjarn	dáðríki_n_n	RMS:19ms
dáðlaus_adj_	ONP:dáðlauss	dáðríkismaður_n_m	RMS:19s
dáðleysi_n_n	ONP:dáðleysi	dáðríkur_adj_	RMS:18s
dáðleysingi_n_m	ONP:dáðleysingi	dáðrómur_n_m	RMS:18s
dáðrakkur_adj_	ONP:dáðrakkur	dáðsamlega_adv_	RMS:17s
dáðsamlegur_adj_	ONP:dáðsamligr	dáðsamlegur_adj_	RMS:16s

Ekkert viðbótarefni bætist við samkeyrsluskrána úr JÓIGrv. og TMS þar sem einu upplýsingarnar úr þeim í samkeyrslunni eru einfaldlega hvort orðin koma fyrir eða ekki. Þess vegna er þessum heimildum sleppt í þessari töflu.

3 Vandamál í normalísingunni

Það er aðallega tvennt sem veldur vandkvæðum í normalísingunni á stafsetningunni í ONP til samræmis við RMS. Í fyrsta lagi er stafsetningin á Ritmálsskránni ekki fullkomlega samræmd og hending ræður því hvaða stafsetningarafbrigði er gert að uppflettingu. Þetta er áberandi í tökuorðum, t.d. þar sem broddstafi r eru ýmist notaðir eða ekki, eins og í orðunum *disputera* og *dispútant*. Þá eru afbrigði í samsettum orðum í ONP stundum gefi n í svigum, t.d. í *dreng(s)skapr*, en svigar eru ekki notaðir í RMS. Þá þarf sérstaklega að skoða upphafsstafi í RMS þar sem notaður er lágstafur í ONP, t.d. RMS: *Dýrfirðingur*/ONP: *dýrfirðingr*.

Úr þessum vanda var leyst í þetta sinn með því að fara sérstaklega yfi r öll orð sem bara komu fyrir í ONP og leita að hugsanlegum afbrigðum í RMS. Var þá stafsetningin í RMS látin ráða normalísingunni, jafnvel þótt hún orkaði tvímælis.

Rétt er að taka það fram að normalísingunni var unnin í flýti og án þess að dæmi væru gátuð, enda er hér aðeins um tilraun að ræða. Ekki er alltaf ljóst af rithættinum einum hvort um eitt og sama orðið er að ræða í báðum heimildum (sbr. t.d. ONP: *dani/danr* og RMS: *Dani*).

Sama vandamál kemur að nokkru leyti fram í efninu úr JÓIGrv., t.d. í afbrigðunum *dándikona* (RMS, ONP) og *dandikona* (JÓIGrv.)

4 Vandamál í orðflokkingum

Reglur um uppsetningu uppflettimynda í heimildunum tveimur eru ekki eins. Í ONP virðast nafnorð nánast alltaf vera gefi n í eintölu, jafnvel þótt um óumdeilanleg fleirtöluorð sé að ræða, sbr. t.d. *draumórr* í stað *draumórar*:

draumórr it sb. m. [÷; -ar]
(pl. draumórar) ...
(ONP 3:195)

Þessu er þó ekki fylgt út í hörgul, sbr. *dyrr*:

¹**dyrr** sb. f. pl. ...
(ONP 3:419)

Þá eru uppflettimyndir sagna hafðar í germynd í ONP, jafnvel þótt þær komi aðeins fyrir í miðmynd eða lýsingarhætti, eins og í miðmyndarsögninni og lýsingarhættinum *daufheyra* og *daufheyrdur* sem eru sameinuð undir uppflettiorðinu *daufheyra* í ONP (3:1238). (Hliðarmyndin *daufeyra* úr bókinni er ekki í listanum úr ONP sem notaður var í samkeyrslunni.)

Hér var Ritmálsskráin látin ráða eins og fyrr og fleirtölumyndin notuð sem uppflettiorð:

RMS:draumórar_n_m (pl)	RMS:16m	ONP:draumórr
RMS:daufheyra_v_mm	RMS:18m	ONP:daufheyra
RMS:dyr_n_f (pl)	RMS:16m	ONP:dyrr

(Af tæknilegum ástæðum vantar ‘pl’ inn í orðflokkingarnar í listanum úr RMS en úr því verður bætt síðar, sjá sjálfan samanburðarlistann.)

Loks er þess að geta að fara þarf sérstaklega yfir kyngreiningu í nafnorðum þar sem ONP gefur tvö möguleika í sama flettiorðinu. Leyst var úr þessum vandamálum í prufukeyrslunni með því að fara yfir útkomuna orð fyrir orð, leiðrétta normaliseringu og orðflokkinggreiningu og keyra svo aftur.

Í JÓIGrv. er dálítið um fleirtöluuppflettimyndir þar sem eintala er í RMS, t.d. *daudamörk* n.pl. (JÓIGrv.) og *daudamark* n. (RMS, ONP).

5 Samkeyrslan

Tölur úr samkeyrslunni á d-inu, eftir að listarnir voru leiðréttil með samanburði á einstökum orðum, eru þessar:

Orðafjöldi í RMS:	16.815
Orðafjöldi í ONP:	1.524
Orðafjöldi í JÓ:	1.312
Orðafjöldi í TMS:	1.937
Samanlagður orðafjöldi:	18.596
Orð sem eru ekki í RMS:	1.781
Orð sem eru bara í RMS:	14.909
Orð sem eru bara í ONP:	646
Orð sem eru bara í JÓ:	222
Orð sem eru bara í TMS:	867
Orð sem eru bara í ONP og JÓ:	17

6 Niðurstöðurnar

6.1 Aldursgreining í RMS og samkeyrslan

Samkeyrslan gefur athyglisverða mynd af aldri elstu dæma í RMS. Orðin sem eru bæði í RMS og ONP eru alls 853. Það er áhugavert að athuga hvort Orðabók Jóns Ólafssonar úr Grunnavík (JÓIGrv.) gefur einhverjar viðbótarupplýsingar til að fylla upp í götin í RMS sem koma í ljós í samkeyrslu hinna heimildanna þar sem í fljótu bragði mætti ætla að Jón brúaði bilið á milli þeirra tveggja. Í töflunni hér á eftir eru tölur um aldursdreifingu í þessum orðum ásamt einu dæmi um hverja aldurskammstöfun í RMS.

Ljóst er að gott væri að athuga aldur elstu dæma í RMS, sérstaklega í seinni hluta þessa lista. Orðafjöldinn sem um er að ræða er allnokkur, en í D-inu eru orð úr ONP sem ekki koma fyrir fyrr en á 18. öld eða síðar alls 369. Ef þetta er framreiknað fyrir allt stafrófið þá eru orð af þessu tagi u.þ.b. 15.500, en D-ið er $\frac{1}{42}$ hluti af Ritmálsskránni allri (16.815 af 707.140). Á sama máta má fá út að orð sem koma ekki fyrir fyrr en á 20. öld séu u.þ.b. 3.300.

Vitað er að eitthvað af orðum hefur verið tekið upp aftur úr forna málinu en það á ekki sennilega ekki við nema um lítinn hluta þessara orða. Nú er í hæsta máta ólíklegt að ráðrúm gefi st til þess á OH að leita að þessum orðum með hefðbundnum aðferðum við orðtöku en lista af þessu tagi mætti nota sem samanburðarefni við vélræna orðtöku. Einnig væri gott ef notendum væri gefið n vísbending um það að orð komi fyrir í fornu máli þó að aldursmerking dæma í RMS gefi annað í skyn. Notendur heimasíðu Orðabókarinnar lesa nefnilega ekki leiðbeiningar frekar en aðrir orðabókanotendur en einhvers staðar á heimasíðunni eru auðvitað upplýsingar um það hvernig skilja ber aldursgreininguna.

Tafla um aldursgreiningu í RMS á orðum úr ONP (853 orð):

Án aldursmerkingar:	7			dammur_n_m	JólGrv.+
15. öld:	2	15	1	drykkjuhorn_n_n	JólGrv.-
		15s16f	1	dvelja_v_	JólGrv.+
16. öld:	280	16	11	dirfð_n_f	JólGrv.+
		16f	1	drykkjupör_n_n (pl)	JólGrv.-
		16m	167	dagfar_n_n	JólGrv.+
		16ms	6	daga_v_	JólGrv.+
		16s	89	dagatal_n_n	JólGrv.-
		16s17f	6	dauðasvefn_n_m	JólGrv.+
17. öld:	195	17	21	dafna_v_	JólGrv.+
		1718	2	digur_adj_	JólGrv.+
		1718f	4	darraður_n_m	JólGrv.-
		17f	35	danskur_adj_	JólGrv.+
		17fm	10	dauðamaður_n_m	JólGrv.+
		17m	56	dimma_v_	JólGrv.+
		17ms	19	drifhvítur_adj_	JólGrv.+
		17ms18	1	draga_n_f	JólGrv.-
		17ms18f	12	drettingur_n_m	JólGrv.+
		17s	27	drykkjarhorn_n_n	JólGrv.-
		17s18f	8	dysja_v_	JólGrv.+
18. öld:	89	18	2	dáraskapur_n_m	JólGrv.+
		18f	16	dagmálatíð_n_f	JólGrv.-
		18fm	2	danshringur_n_m	JólGrv.-
		18m	25	daufheyrast_v_mm	JólGrv.+
		18ms	4	dræmur_adj_	JólGrv.+
		18s	36	dauðdagi_n_m	JólGrv.-
		18s19f	4	dala_v_	JólGrv.+
19. öld:	202	19	3	dómstefna_n_f	JólGrv.-
		19f	25	Danaveldi_n_n	JólGrv.-
		19fm	3	dreyrugur_adj_	JólGrv.+
		19m	74	drápa_n_f	JólGrv.+
		19ms	21	dauðamyrkur_n_n	JólGrv.-
		19ms20f	1	drengilegur_adj_	JólGrv.+
		19s	56	dagafjöldi_n_m	JólGrv.-
		19s20f	19	dalsbotn_n_m	JólGrv.-
20. öld:	78	20	1	dúklaus_adj_	JólGrv.-
		20f	28	djúphyggja_n_f	JólGrv.-
		20fm	3	dauðasút_n_f	JólGrv.-
		20m	38	deilumál_n_n	JólGrv.-
		20ms	1	dýnulaus_adj_	JólGrv.-
		20s	7	dagdrykkja_n_f	JólGrv.-

6.2 Orð sem finnast bara í ONP

Út frá þeim samanburði sem hér er gerður má skipta orðunum sem koma bara fyrir í ONP í fernt:

1. Orð sem enn gætu leynst í RMS
2. Orð sem vantar í RMS
3. Eiginleg „formálsorð“
4. Umframt efni (t.d. örnefni, nöfn; villur . . .)

6.2.1 „Leitið betur“

Gaman væri að athuga skrá um orðin sem fi nnast bara í ONP betur, þau eru nú 675 orð. Sum þessara orða leynast e.t.v. í RMS þegar betur er að gáð. Skoða þarf afbrigði í stafsetningu, samsetningarhætti (ef.et./ef.ft./stofn í fyrri hluta), hástafi /lágstafi , breytingu á kyni o.s.frv. Þetta felur í sér allnokkra handavinnu.

6.2.2 Orð sem vantar í RMS

Orðtakan í Ritmálssafninu er gloppótt, af eðlilegum ástæðum sem oft hefur verið fjallað um. Það er t.d. mjög undarlegt að eftirfarandi orð eru ekki flettiorð í RMS (ekki heldur með stóru D-i):

```
danadrottning_n_f|ONP:|danadróttning  
danakonungur_n_m|ONP:|danakonungr
```

Þessi orð myndu sennilega mörg hver skila sér í vélrænni orðtöku þar sem hafa mætti listann til hliðsjónar.

6.2.3 Eiginleg „formálsorð“

Hér er efni sem við á Orðabókinni þurfum ekki að hafa áhyggjur af að vanti í RMS en þetta er áhugavert til málfræðirannsókna.

6.2.4 Umframt efni

Í báðum orðalistunum er sennilega heilmikið umframt efni, sérstaklega ef þeir væru skoðaðir í heild. Ritmálsskráin er gríðarlega stór og hún er full af ósamræmi hvers konar, bæði í stafsetningu og vali flettimynda. Stafkaflinn D úr ONP er hins vegar mikið unninn, enda er hann kominn út í prentaðri orðabók og þar hafa flettiorð þegar verið valin. Síðari hluti stafrófsins í ONP er á vinnslustigi en hann er áreiðanlega ekki í verra standi en Ritmálsskráin, a.m.k. ekki ef marka má listann á heimasíðu ONP.

Ég fæ ekki séð að umframa efnið skipti miklu máli í samkeyrslunni, einfaldlega vegna þess að það mun ekki fi nnast í báðum heimildunum. Því er auðvelt að láta það einfaldlega detta „út af borðinu“ þegar sjálfri samanburðarvinnunni er lokið en gott að geta gjóað á það augunum til þess að fi nna það efni sem nefnt er í 6.2.1-2. Það yrði þá gert vitandi vits um þá galla sem þarna kunna að leynast. Gagnlegasta efnið fyrir okkur á OH er það sem fram kemur í báðum heimildunum (sbr. kafla 6.1).

7 Og svo ...

Ef litið er á samkeyrslu af þessu tagi sem vinnutól þá held ég að bæði þeir sem vinna við RMS og ONP geti haft af henni gagn. Hér að ofan eru hugmyndir um það efni sem áhugavert er fyrir OH en mér finnst sennilegt að hægt væri að hafa af henni gagn við vinnuna við flettulistann í ONP. Við á OH getum svo stuðst við JÓIGrv. til að fylla í göt í RMS og athuga normaliseringuna betur.

KB 8.5.2007

Sjá orðalista sem þessi athugun er byggð á, <http://www.lexis.hi.is/kristinb/D-rms-onp.pdf>.