

Jón Hilmar Jónsson

Í áttina að samfelldri orðabók

– nokkrir megindrættir í Íslensku orðaneti

1 Inngangur

Þess er ekki að vænta að orðabækur, fremur en önnur mannanna verk, haldi gildi sínu óskertu um aldur og ævi. Hitt er að líkindum sönnu nær og kunnuglegra, að þær sýni fljótt nokkur merki þess að hafa orðið til á liðnum tíma. Þau ummerki geta komið fram á margvíslegan hátt jafnt í innri efnisþáttum sem í ytri búningi orðbókarinnar. Ekki er þar með sagt að orðabækur þyki fljótt úreltar, flestar þeirra hafa þvert á móti til að bera eiginleika sem tryggja notagildi þeirra um langan aldur. Skilningur notenda og samfélagsins almennt á endingu og gildistíma orðabóka er óhjákvæmilega háður viðteknum skoðunum á hlutverki þeirra og þeim notkunarvenjum sem hver og einn hefur til einkað sér. Þegar ný orðabók er mótuð er því eðlilegt að tekið sé mið af ríkjandi hefð og að sama skapi farið varlega í að ætla notendum að víkja frá rótgróinni venju. Þessar aðstæður skýra vafalaust að nokkru leyti þá staðreynd að breytingar á flestum megingrundum almennra orðabóka hafa lengi verið fremur hægfara.

En hér er einnig að verki sú einfalda og alkunna staðreynd að viðfangsefni almennrar orðabókar, þ.e.a.s. orðaforðinn og notkun hans, tekur engum stökkbreytingum á skömmum tíma, og það orðafar sem hverfur úr notkun fær að miklu leyti að halda stöðu sinni sem orðabókarefni. Þess gætir jafnvel að orðabókarlýsingin sveigist af þessum

sökum meira en ella til hins sögulegra sjónarhorns, merkingar- og notkunareinkenni orðanna séu rakin í sögulegu samhengi og í sinni föstu mynd en minna sé hirt um vitnisburð lifandi málnotkunar. Hér ræður að nokkru sú rótgróna hugmynd um notkun orðabóka að þær megi kalla til vitnisburðar um gamla og rótgróna málvenju þegar málnotendur standa andspænis ókunnuglegri eða undarlegri mál- og orðanotkun.

Í þessu ljósi er vert að hugsa til þess að hinar stærstu og efnismestu orðabækur hljóta að sínu leyti að eiga erfiðast með að endurnýjast til samræmis við líðandi stund en efnisgnótt þeirra er hins vegar líkleg til að tryggja að þær haldi þrátt fyrir það betur gildi sínu en efnisminni orðabækur. Þegar haft er í huga að slík meginverk verða gjarna viðmiðun og fyrirmynd annarra minni orðabóka er skiljanlegt að framþróun í gerð orðabóka sé víða tiltölulega hæg.

Pótt orðabækur og gerð þeirra mótist enn af þessum aðstæðum hafa á undanfórnum árum átt sér stað breytingar sem kalla á ný viðhorf til orðabókarlýsingar og skapað hafa forsendur til að byggja upp heildstæð orðabókarverk á traustari og víðtækari grunni en áður. Hér er að baki sú tæknilega bylting sem gerir mögulegt að safna, flokka og greina orðabókarlegan efnivið á stórvirkan og markvissan hátt og skila orðabókarlýsingu í rafrænum búningi. Sú efnisskipan og framsetning sem sniðin var að hinum prentaða búningi ræður ekki lengur ferðinni og nýjar leiðir hafa opnast til að miðla orðabókarlegum upplýsingum á beinni og skilvirkari hátt en áður. Í því felast m.a. möguleikar á að bregðast betur en áður við þeim vanda sem leiðir af framvindu tímans.

Hér verður reynt að rýna ofurlítið í þá stöðu sem orðabókum og orðabókarlýsingu er búin að þessu leyti og þá bæði litið til ríkjandi hefðar og nýrra möguleika. Andspænis þeirri kyrrstæðu mynd sem prentaðar orðabækur birta verða reifaðar hugmyndir um samfellda og síkvika orðabókarlýsingu í rafrænu formi með margbreytilegri efnisþáttum og sveigjanlegri aðgangsskipan.

2 Skorður og takmarkanir prentaðrar orðabókarlýsingar

Prentaðar orðabækur hafa búið við margvíslegar skorður og takmarkanir sem orðið hafa enn skýrari en áður eftir að rafræn orðabókarlýsing kom til sögunnar. Hér verður dregið á þrjá veigamikla þætti sem hafa haft mótandi áhrif í þessu sambandi en horfa allt öðruvísi við þegar ekki þarf lengur að taka mið af prentuðum búningi.

2.1 Efnisleg takmörkun

Einn nærtækasti flokkunarþáttur orðabóka varðar stærð þeirra og efnismagn. Þótt hinar stóru orðabækur rúmi augljóslega meira efni en þær sem minni eru, er ekki þar með sagt að þeir sem vinna við gerð þeirra þurfi síður að takmarka það efni sem er í boði og unnið er úr. Krafan um mat á gildi flettiorða og takmörkun flettiorðaforðans er áþreifanleg hvort sem í hlut eiga stórar orðabækur eða litlar. Eftirtektarvert er að sá vandi sem þar er við að eiga er sjaldnast til umræðu þegar gerð er grein fyrir tilurð og gerð orðabóka.

Gildi flettiorða má meta út frá ýmsum sjónarmiðum, og í flestum orðabókum er mikill hluti flettiorðanna studdur margþættu gildismati, sem m.a. getur byggst á notkunartíðni, merkingarlegri margbreytni, hugtakslegum skýrleika og notkunarfestu í orðasamböndum (sjá m.a. Svensén 2004:79–93). Eftir því sem slíkir einkennisþættir eru færri og veikari er að öðru jöfnu erfiðara að meta hvort tiltekið orð eigi rétt á sér sem flettiorð. Þar kemur m.a. til álitu hvernig líta ber á virka orðmyndun og hvort eðlilegt sé að gera henni nokkur skil með aðild mikilvægra for- og bakliða en takmarka að sama skapi aðild einstakra samsettra og afleiddra orða. Í þessu efni hlýtur afstaðan þó alltaf að mótast af því hvaða notendahópur er hafður í huga. Í einmála orðabók, þar sem einkum er gert ráð fyrir notendum sem hafa mál orðabókarinnar að móðurmáli, má hugsa sér mun strangari takmörkun á aðild merkingarlega gagnsærra samsetninga en í tvímála orðabók, þar sem meirihluti notenda kemur að viðfangsmálinu sem erlendu máli. Þá er hugsað til þess að notkunin beinist fyrst og fremst að skilningi á orðum og orðafari, þar sem merkingarskýringar eru í fyrirrúmi. En á móti kemur að sami notendahópur er líklegur til að hafa þörf fyrir aðgang að fjölbreyttu orðafari, þ. á m. margvíslegum

virikum samsetningum, þegar notkunartilefnið varðar orðaval í ræðu eða riti, eða áhuginn beinist að því að fá yfirsýn um orðafar á tilteknu merkingarsviði. Þessar aðstæður liggja að nokkru leyti til grundvallar hefðbundinni aðgreiningu orðabókartegunda, þar sem skilgreiningar og rækileg sundurgreining merkingarbrigða fer saman við aðhald í flettiorðavali (í skilgreiningaorðabók), en frjálsara flettiorðaval á við þegar gengið er út frá tiltekinni merkingu eða merkingarbrigði (t.d. í samheitaorðabók).

Í tvímála orðabókum togast einnig á tvenns konar notkunarþarfir að þessu leyti. Þegar hugsað er til þeirra sem koma að viðfangsmálinu sem erlendu máli er eðlilegt að takmarka orðaforðann, bæði gagnvart virkri orðmyndun og sjaldgæfu orðafari, og þá gjarna á breytilegan hátt eftir kunnáttustigi notenda. Gagnvart þeim sem hafa viðfangsmálið að móðurmáli og leita eftir viðeigandi jafnheiti flettiorðs í markmálinu er staðan önnur. Þar leita m.a. inngöngu samsett orð sem í sjálfu sér skýrast af merkingu orðliða sinna og sambandi þeirra en eiga sér skýrt merkingarmið sem álykta má að vísað sé til með skýru jafnheiti í markmálinu.

Efnislegar skorður prentaðra orðabóka eru enn augljósari þegar litið er til innri efnisþátta, svo sem orðasambanda af ýmsu tagi og notkunardæma. Þær upplýsingar sem fólgnar eru í þeim hlutum orðabókartextans nýtast notendum heldur ekki í þeim mæli sem þær gerðu ef hægt væri að draga slíka efnisþætti fram sem sjálfstæð athugunar-efni óháð því undir hvaða flettiorðum þeir standa. Hér koma líka við sögu margs konar samþjöppunaraðgerðir, sem m.a. birtast í misjafnlega skýrum táknunum, skammstöfunum og styttingum, sem notendur geta átt erfitt með að ráða í (sjá Wolski 1989).

2.2 Föst efnisskipan

Prentaðar orðabækur hafa ekki síður búið við fastar skorður að því er varðar efnisskipan og þar með aðgang notenda að þeim upplýsingum sem þær búa yfir. Þar mótar meginskipanin með flettiorðum sínum yfirleitt umgjörðina og opnar notendum hlið að margvíslegum efnisþáttum, sem sumpart eiga beint við viðkomandi flettiorð sem umsagnir um einkenni þess, en geta einnig haft sjálfstæða stöðu og krefjast sem slíkir sinna sérstöku umsagna eða skýringa. Flestar hin- ar almennari orðabækur eru bundnar stafrófsröð flettiorðanna, og þá

stendur hvert flettiorð um sig með sínum efnispáttum harla sjálfstætt og einangrað. Því hefur löngum reynst erfitt og fyrirhafnarsamt að tryggja viðunandi samræmi í lýsingu sambærilegra flettna. Val flettimyndanna styðst líka mjög við gamalgróna hefð, þar sem takmarkað tillit er tekið til breytilegs hlutverks í setningarlegu umhverfi. Í íslenskum orðabókum hafa atviksorð sem formlega eru hvorugkynsmyndir samstofna lýsingarorðs (myndir eins og *fljótt*, *skýrt*, *stíft*) ógjarna verið sjálfstæð flettiorð, og svipuð staða er uppi gagnvart lýsingarháttum sagna í lýsingarorðshlutverki (t.d. *meiddur*, sbr. *meiða(st)*, *klæddur*, sbr. *klæða(st)*).

Þar sem merkingarlýsing orðanna er í fyrirrúmi er sundurgrein-ing merkingarbrigða og innbyrðis skipan merkingarliða oft aðalviðfangsefni orðabókarlýsingarinnar. Í prentaðri orðabók er eðlilegt að röð merkingarliðanna sé sem mest látin ráðast af röklegu samhengi og reynt sé að endurspegla þá heildarmynd af orðinu sem notkunarheimildirnar skila. Þótt einstök notkunartilefni snúist aðeins um tiltekna merkingu blasir hún ekki við ein og sér heldur verður að svipast um í lýsingunni. Þetta fyrirkomulag er til þess fallið að styrkja hið málsögulega sjónarhorn við merkingarlýsinguna, sem aftur stuðlar að því að gömul merkingarbrigði haldi velli þótt þau eigi sér ekki lengur stoð í lifandi málnotkun. Og um leið verður til ákveðið viðnám gegn því að hróflað sé við orðlýsingunni eða flettiorðið sé fellt brott. Slík innri festa getur líka verið samofin öðrum efnisatriðum en merkingu orðsins. Þegar föst orðasambönd eru undirskipuð stökum flettiorðum, eins og sterk hefð er fyrir, geta einstök merkingarbrigði flettiorðsins átt við um þau sérstaklega. Þótt slíkt merkingarbrigði standi ekki undir sér að öðru leyti getur orðasambandið og notkun þess tryggt því líf í orðabókartextanum.

Þannig er áhersla á merkingarlýsingu hið innra háð formlegum eigindum orðanna á ytra borði lýsingarinnar. Merkingarlegt samhengi á milli orða og innan orðaforðans er hins vegar illa sýnilegt. Þá er öll yfirsýn um föst og sjálfstæð orðasambönd, jafnt formlega sem merkingarlega, háð aðgangi um stök flettiorð. Sú tilhögun er líkleg til að takmarka aðild slíkra orðasambanda verulega og torvela yfirsýn um samhengi þeirra og framsetningu.

2.3 Tegundarleg aðgreining

Orðabækur má flokka á ýmsa vegu eftir því hlutverki sem þeim er einkum ætlað að gegna. Að vísu er oft misbrestur á að hlutverkinu sé nægilega vel lýst af hálfu ritstjóra eða orðabókarhöfundar, og hafa verður í huga að í kynningu á orðabók sem söluvöru gætir tilhneigingar til að ætla henni meira og víðtækara hlutverk en hún stendur undir í raun. Á hinn bóginn getur þótt ástæða til að leggja áherslu á sérstöðu nýrrar orðabókar, vekja athygli á nýju hlutverki og koma til móts við nýjar notkunarþarfir. Við fræðilegt mat á orðabókum er mjög lítið til þess hversu vel hlutverkið er skilgreint og hvort orðabókin nær að skila því á fullnægjandi hátt. Með hliðsjón af því er eðlilegt að varlega sé farið við að hreyfa fastmótuð tegundamörk og traustara þyki að leita viðurkenndra fyrirmýnda.

Tegundarleg aðgreining orðabóka hefur eðlilega tekið mið af orðabókum í prentuðum búningi. Þær skorður sem þeim eru settar um efnisskipan og sjónarhorn við orðabókarlýsinguna hafa verið ráðandi þættir í þessu sambandi. Það gildir ekki síst um þann grundvallarmun sem gerður er á **segðarskipuðum** (e. semasiological) og **inntaksskipuðum** (e. onomasiological) orðabókum (sjá Hausmann 1989, sbr. einnig Sterkenburg 2003). Í segðarskipuðum orðabókum er lítið á flettiorðin og skipan þeirra frá hinni formlegu hlið orðsins, burtséð frá því hversu margbrotin merking þar er á ferðinni. *Íslensk orðabók* er skýr fulltrúi þessarar tegundar meðal íslenskra orðabóka. Í inntaksskipuðum orðabókum fela flettiorðin í sér vísun til (tiltekinnar) merkingar, þar sem þau sameina gjarna merkingarlega samstætt orðafar, eins og m.a. kemur fram í samheitaorðabókum.

Með tilliti til þessarar aðgreiningar hefur greinargerð orðabóka um merkingarlegan breytileika einstakra orða ekki átt greiða samleið með lýsingu á merkingarskyldu orðafari. Segðarskipaðar orðabækur eru mun fyrirferðarmeiri, hvort sem lítið er til fjölda þeirra, stærðar og efnismagns eða hlutdeildar á orðabókamarkaði. Hin stóru orðabókarverk um einstakar þjóðtungur eru skýrustu og áhrifamestu merkisberar þessarar orðabókartegundar. Staða inntaksskipaðra orðabóka hefur löngum verið mun veikari og óljósari og þær hafa ekki gert tilkall þess að gegna undirstöðuhlutverki við lýsingu á orðaforða einstakra tungumála. Efniviður þeirra hefur heldur ekki birst notendum á jafn skýran og áhrifamikinn hátt og efniviður margra orðabóka af hinni gerðinni, þar sem umfangsmik-

ið dæmasafn vitnar um margvísleg merkingar- og notkunareinkenni orðanna.

Inntaksskipaðar orðabækur búa líka við ákveðinn vanda að því er varðar flettimyndir og val þeirra. Í mörgum algengum orðabókum af því tagi, einkum samheitaorðabókum, hefur flettulistinn sams konar yfirbragð og þegar um segðarskipaða orðabók er að ræða, flettunum er raðað í stafrófsröð og hvert flettiorð getur sameinað ólík merkingarbrigði. Flettiorðið er því ekki alltaf einræður lykll tiltekinni merkingu orðsins, og notandanum er jafnvel ætlað að álykta af kunnáttu sinni við hvaða merkingarbrigði tiltekið samheiti á. Prentaðar samheitaorðabækur eru einnig í vanda gagnvart orðasamböndum sem samheiti eru tilgreind við. Notandinn verður að gera ráð fyrir að þeim sé skipað undir stök flettiorð á líkan hátt og í segðarskipaðri orðabók. Umgjörðin sem orðasamböndum er búin hefur vafalaust takmarkandi áhrif á aðild þeirra langt umfram það sem æskilegt væri.

Ráðið til að rífa sig út úr þessu skjóli segðarskipaðra orðabóka hefur verið það að grípa til flokkunarkerfis eða flokkunarheita sem byggjast á merkingarlegum einkennum. Hugtakaorðabók Rogets, *Roget's Thesaurus of English Words and Phrases*, er klassískur fulltrúi fyrrnefndu aðferðarinnar, síðarnefndu aðferðina þekkja íslenskir orðabókanotendur af orðabókinni *Orðaheimi* (Jón Hilmar Jónsson 2002). Hvor leiðin sem farin er verður að búa notendum í hendur stafrófsraðaða orðaskrá til fyrstu leitar, eins og orða- og orðasambandaskrár þessara tveggja orðabóka sýna.

Eins og áður var tekið fram eiga inntaksskipaðar orðabækur sér óskýrari undirstöðu en hinar segðarskipuðu og það sama gildir um efnisleg mörk þeirra. Í almennt segðarskipaðri skilgreiningaorðabók má takmarka flettiorðavalið með hliðsjón af ýmsum áþreifanlegum þáttum, skýringarþörf, tíðni í texta- eða heimildasafni, landfræðilegri og félagslegri útbreiðslu, aldri o.s.frv. Í inntaksskipaðri orðabók, ekki síst samheitaorðabókum, er skýringarþörfin ekki mælikvarði heldur er dreginn að sá orðaforði sem skipa má í merkingarsamhengi við einstök orð eða stærri samstæður orða. Og gera má ráð fyrir að flettiorðafjöldinn segi ekki mikið um umfangið því aðeins hluti orðaforðans kemur fram í því hlutverki.

Þótt þessar tvær megintegundir gangi hvor sína leið, ef svo má segja, fást þær vitaskuld að miklu leyti við sama orðaforðann og sjá má ýmis merki þess að önnur njóti góðs af því sem hin hefur komið til

leiðar. Hér eru segðarskipaðar orðabækur mun oftast í hlutverki veitandans eins og eðlilegt má telja. Það kemur m.a. fram þegar athugað er hvaðan *Íslensk samheitaorðabók* sækir efni sitt. Þar hafa, samkvæmt formála ritstjóra, tvær segðarskipaðar orðabækur lagt til stofninn, *Íslensk orðabók* Menningarsjóðs frá 1963 *Dönsk orðabók* Freysteins Gunnarssonar frá 1926, en sú síðarnefnda er tvímála orðabók með íslenskum skýringum.

3 Úr prentaðri í rafræna mynd

Með tilkomu rafrænnar gagna- og upplýsingamiðlunar og tölvuvinnslu orðabókargagna hafa ýmsar grundvallarforsendur orðabókargerðar breyst verulega. Það á m.a. við um allt það sem rakið hefur verið hér að framan um þær skorður sem prentaðri orðabókarlýsingu eru settar. Í fyrsta lagi horfir efnisleg afmörkun orðabókarinnar við á allt annan hátt því að engin bein þörf er á að takmarka fjölda þeirra orða og annarra viðfangseininga sem lýsingin beinist að. Í öðru lagi má viðhafa breytilega efnisskipan, bjóða margs konar leitarmöguleika, raða orðum og efnisþáttum á ýmsa vegu og tengja saman samstæð atriði eftir því sem við á. Síðast en ekki síst skapast forsendur til að móta orðabókarlýsingu á nýjan hátt, óháð þeirri tegundaraðgreiningu sem sprottin er úr umhverfi prentaðra orðabóka.

Sú þróun sem hér er á ferðinni getur haft veruleg áhrif á margan hátt, m.a. á það hvernig orðabækur geta endurnýjast og haldið velli þótt tímar líði. Það orðafar og sú orðanotkun sem heyrir sögunni til og ástæða getur þótt til að ryðja burt í prentaðri orðabók til að rýma fyrir lýsingu á lifandi orðanotkun samtímans getur fengið að standa í rafrænni orðabók til vitnisburðar um sögu máls og orðaforða, og slíkt efni heldur áfram gildi sínu í ýmiss konar samhengi. Raunar má gera orðafari sem lýtur að samfélags- og menningarháttum liðins tíma betri skil en áður voru tök á og þá m.a. birta það sem baksvið þeirrar lýsingar sem beinist að orðaforða samtímans.

3.1 Stöðug endurnýjun, nýir miðlunarkostir

Einn augljósasti kostur rafrænnar orðabókarlýsingar felst í því hversu auðvelt er að endurbæta lýsinguna, lagfæra villur, auka við og fella brott eftir þörfum, og láta þær aðgerðir skila sér strax til notenda. Slík-

ar aðgerðir eru ekki bundnar við stök atriði, þær geta einnig beinst að afmörkuðum efnispáttum og einkennum og þá gjarna byggst á úrvinnslu og greiningu nýrra gagna. Jöfn og stöðug endurnýjun, sem gerð er sýnileg gagnvart notendum, er til þess fallin að vekja áhuga þeirra og traust á orðabókarlýsingunni og ljá henni æskilegan ferskleika.

Nýir valkostir við miðlun orðabókarlegra upplýsinga hafa einnig áhrif í þá átt að skerpa athygli notenda. Myndefni getur gegnt veigamiklu hlutverki og komið fram á margvíslegan hátt, m.a. stutt við skýringar í rituðu máli, hreyfimyndir geta lýst samsettu atferli (svo sem ýmiss konar verklagi og vinnubrögðum) betur en nokkur orð, og myndræn auðkenning efnisatriða lífgar upp á orðabókartextann. Hljóðefni hefur stóraukið gildi þeirra upplýsinga sem veita má um framburð orða, og það má einnig nýta til stuðnings rituðum skýringum á orðafari sem lýtur að hljóðum í ýmsum myndum. Galdurinn er í því fólgin að flétta þessa miðlunarkosti saman á sem haganlegastan hátt. Með markvissri notkun mynd- og hljóðefnis má einnig sækja lengra en áður inn á svið alfræðilegra upplýsinga og koma fyrir ýmiss konar fróðleik í tengslum við orðabókarefnið.

3.2 Sveigjanleg umgjörð orðabókartextans

Gagnerasta breytingin sem leiðir af tilkomu rafrænnar miðlunar orðabókarlegra upplýsinga er þó fólgin í því að þar með er rofin sú fasta umgjörð sem orðabókartexinn óhjákvæmilega hefur í prentaðri orðabók. Segja má að með því opnast dyr í allar áttir. Meðferð flettulistans sýnir þetta í einfaldri mynd. Hann þarf ekki lengur að koma fram í fastri röð, og reyndar þarf hann ekki að birtast sem slíkur því notandinn getur einfaldlega ritað þá flettimynd sem hann hefur valið, séð þá lýsingu sem þar er á bak við og eftir atvikum fetað sig áfram í leit að frekari upplýsingum. En það getur verið gagnlegt að hafa flettulistann fyrir augunum, eiga þess kost að velja viðeigandi flettu án þess að rita hana í leitarreit, sjá þann orðaforða sem stendur næst leitarorðinu í stafrófsröð eða orðmyndunarlega o.s.frv. Flettulistann má auk þess flokka á ýmsa vegu án tillits til flettustrengsins sjálfs, t.d. málfræðilega eftir orðflokki og jafnvel (ef viðeigandi greining hefur farið fram) merkingarlega eftir hugtaki eða merkingarsviði. Slík greinimörk má svo eftir atvikum tengja saman til að hnitmiða leitina enn frekar.

3.3 Sjálfstæði orðasambanda

Í rafrænni orðabók er engin ástæða til að takmarka flettulistann við einyrtar flettur. Hann getur einnig rúmað fleiryrtar einingar, sérstaklega þær sem standa sjálfstæðast gagnvart stökum flettiorðum og kalla á sína eigin lýsingu. Föst orðasambönd af því tagi geta raðast með einyrtum flettum og verið sýnileg innan allsherjarflettulista en þau þurfa einnig að vera aðgengileg með frjálsri orða- og strengjaleit. Málfræðileg og merkingarleg flokkunaratriði koma hér ekki síður til greina en meðal stakra orða, svo að velja megí orðasambönd af tiltekinni formgerð eða sækja sér orðafar sem tengist sérstöku merkingarsviði.

Slík meðferð fastra orðasambanda hefur afgerandi áhrif á alla efnisskipan orðabókarlýsingarinnar, brýtur hana upp í fleiri viðfangseiningar en tengir hana jafnframt saman í mun samfelldari heild. Og innri þættir lýsingarinnar geta líka verið hreyfanlegir og virkir í heildarmyndinni. Notkunarsambönd og bein notkunardæmi eru fyrirferðarmikil í flestum meiri háttar orðabókum, þar sem þau vitna um notkun og notkunareinkenni einstakra orða, hvort sem slík lýsing er aðalviðfangsefnið eða hún er samofin greinargerð um megineinkenni eins og merkingu orðanna. Í almennum prentuðum orðabókum hafa laustengdari orðasambönd eins og orðastæður (e. collocations, d. kollokationer) yfirleitt fengið takmarkað rými, og orðabókarhöfundum hefur oft reynst erfitt að velja og hafna andspænis ríkulegu úrvali slíkra notkunarsambanda í heimildum. Það upplýsingagildi sem fólgið er í orðastæðum sem tiltækar eru í orðabókargrunninum skilar sér oft aðeins að litlu leyti í orðabókartextanum.

3.4 Orðasambönd vitna um merkingarvensl

Með því að gera orðastæðum skil sem sjálfstæðu viðfangsefni í orðabókarferlinu og láta þær mynda heildstætt safn með tengingum við þau orð sem þær eru myndaðar úr má fá þeim mun víðtækara hlutverk en prentaðar orðabækur geta boðið. Það felst í því að láta orðastæður leiða saman merkingarlega samstætt orðafar. Í orðastæðusafni einstakra nafnorða má víðast greina svo og svo mikinn fjölda samheita meðal lýsingarorða sem með þeim standa, og það sama gildir um einstök lýsingarorð sem eiga sér nafnorð að fylgdarorðum. Undir nafnorðinu *andlit* má t.d. greina samheitaklasa eins og *útgrátinn*, *társtokkinn*, *tárugur* og *festulegur*, *meitlaður*, *skarplegur*, og lýsingarorðið *ljúfur*

sameinar nafnorðin *geð*, *lund* og *skap*. Sambærileg tengsl eru að þessu leyti á milli nafnorða og sagna, sömuleiðis á milli sagna og atviksorða (atviksliða). Með þeim stórvirku vinnslu- og greiningaraðgerðum sem nú má viðhafa á orðabókarlegum efniviði í málheildum og textasöfnum er hægt að fá orðastæðum virkara hlutverk en áður og láta þær jöfnum höndum endurspeglar setningarleg og merkingarleg einkenni.

Notkunardæmi í lengri mynd sem sótt eru í textasöfn eða aðrar frumheimildir geta vitaskuld fengið meira rými en áður og birt margbreytilegar upplýsingar, þótt erfiðara sé að koma við flokkun og greiningu sem nýta má við aðra efnisþætti orðabókarlýsingarinnar.

Í ljósi þessa getur merkingarleg greining og flokkun orðaforðans að töluverðu leyti farið fram án þess að þar sé byggt á hefðbundnum skilgreiningum og merkingarskýringum. Með því að styðjast við notkunarsambönd orðanna og formleg einkenni er stigið mikilvægt skref í þá átt að fella segðarskipaða og inntaksskipaða orðabókarlýsingu saman í samstæða heild, með virkum tengslum á milli ólíkra hluta.

Hvatinn að því að byggja upp samfellda orðabókarlýsingu af þessu tagi er ekki aðeins fólgin í því að leita álitlegra lausna í samræmi við þá miðlunarkosti sem orðabókarefnið og greining þess hefur að bjóða. Þar er einnig að baki sú ályktun að notendur rafrænnar orðabókar vænti þess almennt að hún geti gegnt sem víðtækustu hlutverki og þeir eigi auðvelt með að tileinka sér nýja og nýstárlega notkunarmöguleika.

4 Íslenskt orðanet – nokkrir megindrættir

Þær hugmyndir sem hér hafa komið fram hafa mótað orðabókarverkefni sem greinarhöfundur vinnur nú að ásamt Þórdísi Úlfarsdóttir og hefur verið valið heitið *Íslenskt orðanet*. Hér á ekki við að lýsa orðanetinu í einstökum atriðum en ástæða er til að draga fram nokkur megin-einkenni sem ef til vill geta skýrt ofurlítið myndina af því sem fjallað hefur verið um hér á undan. Í því sambandi skal tekið fram að orðanetið á sér ekki erlenda fyrirmynd, og það hefur að ýmsu leyti sérstöðu meðal sambærilegra merkingarneta. Þar má í fyrsta lagi nefna samhengið við víðtæka og heildstæða orðasambandalýsingu, sem endurspeglast í því að orðasambönd eiga sjálfstæðari aðild að netinu en venja er. Í öðru lagi gegna orðmyndunarleg vensl mikilvægu hlut-

verki sem tengiliður við merkingargreininguna. Í þriðja lagi spannar orðanetið feikilega víðtækan orðaforða, án tíðni- eða aldursbundinnar takmörkunar. Jafnframt því að skila fræðilegri greiningu og flokkun á íslenskum orðaforða er orðanetinu ætlað að vera undirstaða nýrra orðabókarverka, einkum samheita- og hugtakaorðabóka.

4.1 Grunnefni orðanetsins

Grunnefni orðanetsins er að meginhluta til af tvennu tagi: a) orðasambönd (föst orðasambönd og orðastæður) úr *Stóru orðabókinni um íslenska málnotkun* (STOB, Jón Hilmar Jónsson 2005) og orðasambandskrá Orðabókar Háskólans (samtals rösklega 100.000 sambönd), b) merkingarlega flokkaðar samsetningar einstakra flettiorða í STOB (áður *Orðastað* (Jón Hilmar Jónsson 2001b); samtals um 100.000 orð). Meginhlutar einstakra orðasambanda og samsetninga eru tilgreindir sem lykilorð og hvert og eitt lykilorð tengir saman öll þau sambönd þar sem það kemur fram.

Við þennan grunn hefur bæst efniviður úr ólíkum áttum, einkum þó nýjar orðastæður sem sóttar eru í tölvutæk textasöfn. En hér er ástæða til að vekja athygli á tvenns konar efniviði sem hefur nokkra sérstöðu í þessu samhengi. Val jafnheita í erlend-íslenskum orðabókum er víða til marks um mat á merkingarskyldleika íslenskra orða, sérstaklega þegar um er að ræða hliðskipuð jafnheiti sem eiga við tiltekið flettiorð. Í orðanetinu er að finna slík hliðskipuð sambönd úr tveimur íslensk-erlendum orðabókum, *Dansk-íslenskri orðabók* (1992) og *Ensk-íslensku orðabókinni* (2006), sem góðfúslega hefur verið veitt leyfi til að nýta í þessu samhengi. Úr talmálssafni Orðabókar Háskólans má lesa upplýsingar um merkingarsamband fjölda íslenskra orða. Með því að styðjast við frumgreiningu á orðaforða safnsins í talmálsskrá hafa merkingarskyld orð (einkum samheiti) verið felld inn í orðanetið.

4.2 Uppbygging flettulista

Auk þess að tengja saman sambönd með sameiginlegum lið mynda lykilorðin samfelldan flettulista. Frekari uppbygging flettulistans á sér síðan stað sem liður í úrvinnslu og greiningu grunnefnisins. Þar liggur það sjónarmið til grundvallar að hver og ein fletta eigi að vera sem

mest merkingarlega einræð. Meðal nafnorða má koma í veg fyrir samslátt merkingarbrigða með raðnúmerum sem greina að flettur með sömu ritmynd. Erfiðara getur verið að beisla lýsingarorðin á þann hátt en þar kemur sú lausn til greina að koma fyrir aðgreinandi þætti innan flettimyndarinnar, t.d. með því að tilgreina lýsandi nafnorð sem lýsingarorðið á við í tiltekinni merkingu. Þannig gæti lýsingarorðið *hvass* t.d. komið fram í aðgreindum flettimyndum: **hvass <augnaráð>**, **hvass <brún>**, **hvass <gagnrýni>**, **hvass <hnífur>**, **hvass <vindur>**. Mesta sérstöðu hafa þó sagnir með sínum mikla breytileika jafnt í merkingu sem formgerð. Til að merkingarleg einræðing meðal sagna nái fram að ganga eru flettimyndir þeirra að langmestu leyti settar fram sem tví- eða fleiryrt sagnarsambönd með rökliðum sínum. Í orðanetinu er tekið mið af þeirri framsetningu sem gildir um sagnarsambönd í STOB, þar sem m.a. er greint á milli sambanda sem eiga við persónuvísandi frumlag í nefnifalli og sambanda þar sem frumlagið vísar til hluta og fyrirbæra (sjá Jón Hilmar Jónsson 2005: XXI–XXIII, sbr. einnig Jón Hilmar Jónsson 2001a). Í fyrrnefnda tilvikinu hefst sambandið (flettimyndin) á nafnhætti sagnarinnar: **ganga að <skilmálunum>**, **ganga af vitinu**, **ganga ekki heill til skógar**. Í hinu síðarnefnda fer nafnliðarstrengur á undan sögninni, og sögnin kemur fram í nútíðarmynd 3. persónu: **<ferðin> gengur að óskum**, **<veðrið> gengur niður**, **<varan> gengur út**.

Þessi aðgerð gjörbreytir ásýnd og umfangi flettulistans og færir sagnarsamböndin upp að hlið stakra orða sem fullgildar flettur. Önnur aðgerð í sömu átt er í því fólgin að skilgreina föst orðasambönd meðal atviksliða sem tví- og fleiryrtar flettur.

Með þeim umbreytingum sem hér er lýst verður til flettulisti sem rúmar flettur af ólíku tagi, allt frá stökum ósamsettum orðum til margliða orðasambanda. Og flettulistann má samræma og endurbæta eftir því sem úrvinnslu og greiningu efnisins miðar áfram, og nýjar flettur geta bæst við hvenær sem er í því ferli.

4.3 Merkingarsambönd og merkingarheildir

Meginhlutverk orðanetsins lýsir sér að nokkru í því sem hér hefur verið rakið um efniviðinn og uppbyggingu flettulistans. Markmiðið er að rekja merkingarleg vensl á milli orða og orðasambanda og skipa saman samstæðu orðafari. Merkingarvenslin eru byggð upp út

frá nánustu og skýrustu venslum (samheitum og andheitum) og það orðafar sem þannig er markað síðan tengt saman í stærri merkingarheildir undir hugtakaheitum. Þar gegna sameinandi lykilorð m.a. virku hlutverki, eins og áður var nefnt. Í því sambandi hafa orðasamböndin nokkuð ólíka stöðu. Hin fastari þeirra, svo sem orðtök, innihalda lítt eða ekki sameinandi liði og hafa því ekki virku hlutverki að gegna á sama hátt og lausari sambönd eins og orðastæður. Orðapör, þ.e. sambönd með hliðskipuðum orðum af sama orðflokki (*amstur og erfiði, tryggur og fylgispakur, dafna og þroskast*) hafa líka nokkra sérstöðu gagnvart öðrum orðastæðum. Þau geta endurspeglað samheiti en merkingarvensl þeirra eru þó lausari í sér, svo að samtenging orða á þeim grunni spannar gjarna víðara og grófara merkingarsvið.

Merkingarvenslin sem efni orðanetsins birtir má rekja og túlka á mismunandi vegu, og hinn merkingarlega skyldleika má að nokkru leyti marka af því hversu margar tengingar vitna um samband flettnanna og hversu nærstæð þau lykilorð eru sem tengja þær saman. Slíkar tölulegar upplýsingar opna möguleika á að lýsa merkingarvenslum innan orðaforðans á mun fíngreindari hátt en venja er til í orðabókarlýsingu.

4.4 Formlegar og málfræðilegar eigindir flettnanna

Það leiðir af samhengi orðanetsins við lýsingu orðasambanda og samræmdri framsetningu þeirra að hinar formlegu eigindir eru frá upphafi grunnþættir orðbókarefnisins. Með því að fara með sagnasambönd og önnur föst orðasambönd sem fullgildar flettur er hægt að viðhafa málfræðilega mörkun á flettimyndunum og ná þannig fram víðtækri og nákvæmri greiningu og flokkun flettulistans (sjá Þórdís Úlfarsdóttir 2006). Á grundvelli slíkrar greiningar má auðveldlega fá yfirsýn um tiltekna formskipan, t.d. sagnarsambönd með andlagi í tilteknu falli, hvort heldur sem er gagnvart einstökum sögnum eða innan flettulistans í heild. Eftir því sem greiningu orðanetsins vindur fram gefst svo kostur á að tengja hin formbundnu einkenni við merkingarlega greiningarþætti og þá hvort heldur sem er skoða einstakrar formgerðir í merkingarlegu ljósi eða virða fyrir sér merkingarlega samstætt orðafar með hliðsjón af formgerð.

Aðgreining orðflokka er að sjálfsögðu einnig tiltæk til flokkunar flettulistans og sem viðbótargreinimark gagnvart einstökum merk-

ingarsviðum. Þannig má búa notendum í hendur ýmsa samsetta leitarkosti sem ekki hafa verið í boði í fullunninni orðabókarlýsingu.

Hér er gert ráð fyrir að hin málfræðilega greining takmarkist við flettulistann en hægt er að hugsa sér að hún sé enn víðtækari og taki þá til orðasambandanna í heild sinni.

4.5 Afstaðan til skilgreiningaorðabókar

Eins og áður var nefnt hafa segðarskipaðar orðabækur um langan aldur verið í eins konar forystuhlutverki meðal orðabóka og þá sérstaklega þær orðabækur sem birta rækilegasta sundurgreiningu merkingarbrigða einstakra orða. Við samningu slíkrar orðabókarlýsingar er hver einstök fletta viðfangsefni út af fyrir sig með takmarkaðri útsýn til annarra flettna og þeirra sameiginlegu einkenna sem fram kunna að koma séu flettunar bornar saman. Merkingarlýsingin með skilgreiningum sínum og skýringum er í öndvegi og það markmið að semja sem nákvæmastar skýringar samræmist illa hugsanlegri viðleitni til að ná fram æskilegu samræmi í skilgreiningum á merkingarlega samstæðu orðafari. Eðlilegt er að þessi áhersla og það verklag sem henni fylgir gildi svo lengi sem efniviðurinn býr ekki yfir neinum upplýsingum um innra samhengi orðaforðans.

Í orðabókarlýsingu sem byggð er upp á þann hátt sem lýst hefur verið hér að framan er staða skilgreininga í orðabókarferlinu allt önnur. Þar geta þær notið góðs af þeim margvíslegu greiningarþáttum sem orðanetið leggur til, og eðlilegast er að fullbúnar merkingarskýringar komi ekki til sögunnar fyrr en hafa má fullan stuðning af þeirri greiningu. Svo er á það að líta hvort þær upplýsingar sem greiningin skilar draga úr þörfinni fyrir hefðbundnar merkingarskilgreiningar eða breyta forsendunum fyrir því hvernig þær eru settar fram.

5 Lokaorð

Í yfirskrift greinarinnar er umfjöllunarefni hennar kynnt undir heitinu **samfelld orðabók**. Með því heiti er reynt að vísa til þess sem hér hefur verið vakin athygli á og er í því fólgið að byggja upp heildstæða orðabókarlýsingu í skipulegum áföngum, þar sem formbundin og merkingarleg lýsing kallast á og þessir tveir meginhlutar geta

tengst á margvíslegan hátt og birt fjölþættar upplýsingar á grundvelli áþreifanlegra gagna.

Orðabókarlýsingin er einnig samfelld í þeim skilningi að hún tekur auðveldlega við viðbótarefni sem fella má að þeirri lýsingu sem fyrir er, m.a. með tilliti til breytinga á mál- og orðanotkun og því samfélagi sem orðin vísa til. En orðabókin getur jafnframt haldið til haga og varðveitt orðanotkun og orðafar liðins tíma og eftir atvikum komið því á framfæri í virku samhengi við lifandi orðafar.

Með þessu er stigið skref í þá átt að koma hreyfingu á þá kyrrstæðu mynd af máli og orðaforða sem orðabækur hafa löngum birt og láta orðabókarlýsinguna endurspegla síkvikt og breytilegt eðli þessara fyrirbæra.

Ritaskrá

- Dönsk-íslensk orðabók*. 1992. Ritstj. Hrefna Arnalds, Ingibjörg Johannesen. Reykjavík: Ísafoldarprentsmiðja.
- Ensk-íslenska orðabókin*. 2006. Reykjavík: JPV útgáfa.
- Freysteinn Gunnarsson. 1926. *Dönsk orðabók með íslenskum þýðingum*. Reykjavík: Ísafoldarprentsmiðja.
- Hausmann, F.J. 1989. Wörterbuchtypologie. Í: *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie*. I, bls. 968–981. Ritstj. Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand og Ladislav Zgusta. Berlin/New York: De Gruyter.
- Íslensk orðabók handa skólum og almenningi*. 1963. Ritstj. Árni Böðvarsson. Reykjavík: Bókaútgáfa Menningarsjóðs.
- Íslensk samheitaorðabók*. 1985. Ritstjóri: Svavar Sigmundsson. Reykjavík: Styrktarsjóður Þórbergs Þórðarsonar og Margrétar Jónsdóttur Háskóla Íslands.
- Jón Hilmar Jónsson. 2001a. Staða orðasambanda í orðabókarlýsingu. *Orð og tunga* 5:61–86.
- Jón Hilmar Jónsson. 2001b. *Orðastaður. Orðabók um íslenska málnotkun*. Önnur útgáfa, aukin og endurskoðuð. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2002. *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2005. *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.
- Orðasambandaskrá Orðabókar Háskólans: www.lexis.hi.is/osamb/osamb.pl*

Roget's Thesaurus of English Words and Phrases. 1988. London: Penguin Books.

Sterkenburg, P. van. 2003. Onomasiological specifications and a concise history of onomasiological dictionaries. Í: *A Practical Guide to Lexicography*, bls. 127–143. Ritstj. Piet van Sterkenburg. Amsterdam/Philadelphia: John Benjamins Publishing Company.

STOB = Jón Hilmar Jónsson 2005.

Svensén, B. 2004. *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Wolski, W. 1989. Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. Í: *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie*. I, bls. 956–967. Ritstj. Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand og Ladislav Zgusta. Berlin/New York: De Gruyter.

Pórdís Úlfarsdóttir. 2006. Málfræðileg mörkun orðasambanda. *Orð og tunga* 8:117–144.

Lykilorð

rafrænar orðabækur, hugtakaorðabækur, orðanet, samheiti, flettumyndir, orðastæður, orðasambönd

Keywords

electronic dictionaries, thesaurus, word net, synonyms, lemmatisation, collocations, phraseology

Abstract

In this article it is argued that a comprehensive electronic dictionary description must rest on a new and more wide ranging basis than previously possible for those printed works that for a long time have shaped lexicography the most. Instead of focusing on particular lemmas and their different meanings it is necessary first to identify semantic relations within the vocabulary, both regarding words and phrases, based on a systematic analysis of collocations and other phrases. It is described how such an analysis can connect semantic and formal features of phrases and, by offering combined searches, how users can employ various search facilities. The presentation of such a dictionary description is further described with reference to the ongoing project *Íslenskt orðanet* (Icelandic Word Net).

Jón Hilmar Jónsson

Stofnun Árna Magnússonar í íslenskum fræðum

Neshaga 16

IS-107 Reykjavík

jhj@hi.is