

Jón Hilmar Jónsson

Að byggja undir íslensk-erlenda orðabók

Forgreining og orðabókarefni

1 Inngangur

Í þessari grein er fengist við að lýsa einkennum og meðferð íslenskra orðabókargagna í tvímála og þá einkum íslensk-erlendum orðabókum.¹ Gerð er stuttlega grein fyrir stöðu íslenskrar orðabókargerðar þessarar tegundar, m.a. með hliðsjón af breyttum aðstæðum með tilkomu netsins og veforðabóka. Lýst er forystuhlutverki viðfangsmálsins og ólíkum aðstæðum eftir því hvort íslenska er í stöðu viðfangsmáls eða markmáls.

Meginefni greinarinnar snýr að íslensku sem viðfangsmáli, út frá efnisföngum, efnisvali og samsetningu flettulistans og því hlutverki hans að mynda umgjörð um orðabókarefnið. Bent er á þær takmarkanir sem eru á flokkun flettulistans og þeirra gagna sem liggja þar að baki og lögð áhersla á gildi þess að gerð hans og uppbygging styðjist við merkingargreiningu og merkingarlega flokkun. Í því sambandi er vísað til *Íslensks orðanets* sem hefur að geyma margbreytileg orðabókargögn með viðtækri greiningu á formgerðarlegum og merkingarlegum venslum þar sem fleiryrtar merkingarbærar einingar hafa sjálfstæða stöðu til jafns við stök orð.

Fjallað er um notagildi stafrænna textasafna og mismunandi grein-

1 Ég þakka ritstjóra og ritrýnendum tímaritsins góðar og gagnlegar athugasemdir og ábendingar við upphaflegan texta greinarinnar.

ingar á þeim með tilliti til lýsingar á notkunareinkennum og notkunarumhverfi orðanna og vali notkunardæma. Þá er víkið að sérkennum íslensks orðabókarefnis þar sem íslenska er markmálið, bent á hversu merkingarþátturinn er gildur við meðferð þess og greiningu og færð rök að því að rækileg forgreining eigi jafnt við hvort sem fengist er við íslensku sem viðfangsmál eða markmál. Loks er fjallað stuttlega um kosti þess að víkja frá hinum orðbundnu venslum málanna og viðhafa hugtakslega nálgun við orðabókarlýsinguna þar sem meira jafnvægi getur orðið með málunum.

2 Orðabók milli mála

Það er eðlilegt metnaðarmál hverju málsamfélagi að bera mál sitt og orðaforða að öðrum tungumálum í mynd orðabóka. Aðstæður til framtaks á því sviði eru vissulega ólíkar frá einu málsamfélagi til annars og mótast m.a. af stærð og krafti málsamfélagsins og hvernig menningar- og félagslegum tengslum þess við önnur málsamfélög er háttað. Landfræðilegt nágrenni, söguleg tengsl og málsögulegur skyldleiki eru virkir hvatar í þessu sambandi en áhrifavald stórra málsamfélaga segir einnig til sín eins og best sést á því hversu orðabækur gagnvart ensku eru almennt fyrirferðarmiklar.

Gerð nýrrar tvímála orðabókar er gjarna a.m.k. öðrum þræði menningarlegt framtak þar sem atbeini beggja málsamfélaga kemur til. Orðabókinni er þá ætlað að styrkja tengsl viðkomandi þjóða, efla viðskipti þeirra, styðja við ferðamennsku o.s.frv. Í öðrum tilvikum skírskotar verkið fremur inn á við, t.d. þegar orðabókin hefur það hlutverk að greiða götu innflytjenda og einstakra innflytjendahópa í nýju samfélagi. Enn önnur verk snúa beinlínis að námi og menntun í viðkomandi tungumáli á ólíkum stigum.

Margt hefur breyst í umhverfi orðabókagerðar og orðabókanotkunar á undanförunum árum. Með tilkomu netsins hefur aðgengi að hvers konar upplýsingum orðið stórum betra en áður og að sama skapi hefur orðið auðveldara að koma upplýsingum á framfæri. Upplýsingar um orð og orðanotkun þurfa hér minni umbúnað en í prentheimi, og þær geta átt gilt erindi í ýmsu formi, t.d. í mynd einfaldra orðalista sem auka má í eftir þörfum og aðstæðum, án þess að um sé að ræða frágengið eða fullbúið orðabókarverk í hefðbundnum skilningi. Notendur netsins njóta góðs af þessu en geta þó stundum átt erfitt með að meta gildi þeirra verka sem ratað er að, hversu vel

er til þeirra vandað og á hverju einstök efnisatriði og upplýsingar eru byggðar.

Þessar aðstæður fela í sér nýjar áskoranir fyrir þá sem standa að gerð og útgáfu orðabóka og annars konar orðabókarefnis. Hinn prentaði búningur heldur að vísu enn velli auk þess sem margar prentaðar orðabækur er jafnframt að finna í rafrænu formi. En netið verður æ virkari upplýsingamiðill fyrir orðabókarefni, sem leitar þar síns forms og framsetningar, án þess að vera bundið af forskriftum prentaðs texta. Einkenni prentaðs orðabókartexta eru þó áfram greinileg enda má gera ráð fyrir að það komi mörgum notendum vel. En jafnframt leitar birtingin meiri sveigjanleika með fjölþættari aðgangi og tengingum innan orðabókartextans og út fyrir hann.

2.1 Íslenska í tvímála samhengi

Nýtt framtak og nýir áfangar í gerð tvímála íslenskra orðabóka mótast af þessum aðstæðum þar sem fjárhagslegar forsendur prentaðra orðabóka veikjast jafnt og þétt og netmiðillinn sækir á. Þörfin fyrir nýjar orðabækur markast þá öðrum þræði af því hversu brýnt þykir að búa lýsingunni umgjörð á netinu hvort sem prentuð orðabók gagnvart viðkomandi tungumáli er fyrir hendi eða ekki.

Tvímála orðabækur um íslensku eru að vonum sundurleitar enda mótaðar af ólíkum aðstæðum á ólíkum tímum. Sumar þeirra eru mjög komnar til ára sinna og hafa enn ekki verið leystar af hólmi en að drjúgum hluta er þó um að ræða nýlegar orðabækur. Að því er varðar tengslin við norræna málsvæðið markar veforðabókin ISLEX, sem tengir íslensku sem viðfangsmál við fimm norræn mál (dönsku, norskt bókmál, nýnorsku, sænsku og færeysku) vissulega tímamót, og gerð hennar opnar vonandi leiðir að sambærilegum samanburði við önnur markmál í mynd tví- eða margmála orðabóka á vefnum. Sem veforðabók getur hún líka betur en fyrri prentaðar orðabækur haft gildi við ólík notkunartilefni og sameinað þarfir íslenskra og erlendra notenda.

Þær orðabækur gagnvart norrænum málum þar sem íslenska er í hlutverki markmálsins eru harla ólíkar, og í þeim flokki er skýr þörf á umbótum og endurnýjun sem m.a. felur í sér notendaaðgang á vefnum. Yngst er *Dönsk-íslensk orðabók*, að stofni til frá 1992 og aukin í nýrri útgáfu 2004, en orðabækur gagnvart norsku og sænsku eru ívið eldri (*Svensk-isländsk ordbok* (1982), *Norsk-islandsk ordbok* (1987)).

Þegar litið er út fyrir norræna málsvæðið er staðan sem vænta

má í ýmsum tilvikum veikari, sérstaklega að því er varðar íslensku sem viðfangsmál. Gagnvart þýsku er augljós þörf á nýrri og vandaðri orðabók þótt þær íslensk-þýsku orðabækur sem tiltækar eru geti greitt götu margra notenda (*Íslensk-þýsk orðabók* (1993) og *Þýsk-íslensk/íslensk þýsk orðabók* (2004)). Gagnvart frönsku er staðan þó sýnu verri þar sem eina íslensk-franska orðabókin, *Íslenzk-frönsk orðabók* (1950), er orðin meira en hálfrar aldar gömul. Hér háttar hins vegar betur til þar sem íslenska er markmálið, því *Frönsk-íslensk orðabók* kom út árið 1995 og *Þýsk-íslensk orðabók* árið 2008. Gagnvart ensku er einnig þörf á endurnýjun þótt þar sé staðan betri. *Ensk-íslensk orðabók* (1984) er sannarlega efnismikil og var mikilsvert framtak á sínum tíma. Forlagið fylgdi þeirri útgáfu eftir með *Ensk-íslenskri skólaorðabók* (1986) sem síðan kom út aukin og endurbætt undir heitinu *Ensk-íslenska orðabókin* (2006). Íslenska er svo viðfangsmálið í *Íslensk-enskri orðabók* Sverris Hólmarssonar, Christophers Sanders og Johns Tucker, sem fyrst kom út árið 1989 en að nýju aukin og endurbætt í ritstjórn Christophers Sanders árið 2009. Ítölsku og spænsku hafa verið gerð skil með samstæðum orðabókum í báðar áttir (*Íslensk-ítölsk orðabók* (1994) og *Ítölsk-íslensk orðabók* (1999), *Spænsk-íslensk orðabók* (2007) og *Íslensk-spænsk orðabók* (2011)). Flestar þessara orðabóka eru jafnframt tiltækar sem veforðabækur í vefbókasafninu *Snöru*. Meðal tvímála orðabóka um íslensku ber svo sérstaklega að nefna hinar efnismiklu orðabækur gagnvart rússnesku, *Íslenzk-rússneska orðabók* (1962) og *Rússnesk-íslenska orðabók* (Helgi Haraldsson 1996) sem um margt hafa líkt yfirbragð þótt sú síðarnefnda sé enn efnismeiri og margbrotnari að gerð. Í báðum þessum bókum er orðalýsingin að ýmsu leyti rækilegri en notendur eiga að venjast í tvímála orðabókum.

2.2 Hlutverk í tvímála orðabók: viðfangsmál eða markmál

Hér verða einstakar tvímála orðabækur um íslensku ekki tíundaðar frekar þótt nefna megji ýmis markverð verk á því sviði. Þess í stað verður athyglinni beint að þeim aðstæðum og forsendum sem gerð slíkra orðabóka býr við, sérstaklega að því er varðar grundvallargögn, val, afmörkun og framsetningu orðabókarefnisins og aðgang notenda að orðalýsingunni og einstökum þáttum hennar. Þar er skýr munur á eftir því hvoru hlutverkinu íslenska gegnir, hvort hún er viðfangsmálið og þar með í forgrunni orðabókartextans (íslensk-erlendar orðabækur),

eða markmál sem bregður upp samsvörunum við erlent viðfangsmál orðabókarinnar hverju sinni (erlend-íslenskar orðabækur).

2.2.1 Viðfangsmálið í forgrunni

Viðfangsmálið og efnisþættir þess móta að jafnaði umgjörð tvímála orðabókar og afmarka umfang hennar og orðafjölda. Þar myndar flettulistinn uppistöðuna sem öðrum efnisþáttum er skipað undir. Þegar íslenska er í þessu hlutverki kemur það í hennar hlut að tefla fram þessum uppistöðueiningum, sem innan viðeigandi marka eiga að birta samfellda mynd af íslenskum orðaforða og orðanotkun. Þar með verður val flettiorða eðlilega í brennidepli á fyrstu stigum orðabókarverksins.

Samsetning flettulistans getur átt sér stað með ýmsum hætti og margar einstakar ákvarðanir ráðast fyrst og fremst af mati þess, höfundar eða ritstjóra, sem um vélar og ábyrgðina ber hverju sinni. Hinn óþreifanlegi vandi felst oftast en ekki í því hversu erfitt er að beita samræmdum mælikvarða á flettugildi orða, meta hvort tiltekið orð eigi að fá inni í listanum (og fái þá jafnframt jafnheiti eða annars konar skýringu á markmálinu) eða hafi einhver þau einkenni sem réttlæta að litið sé framhjá því. Þar koma til ólík greinimörk og tillit til ólíkra notendahópa. Það snertir m.a. mat á hlut samsetninga og virkri orðmyndun, viðhorf til málstöðlunar og málræktar og hvort orðabókarlýsingin á sem mest að einskorðast við samtímamál.

Annar og raunar stærri vandi, sem í vissum skilningi er óáþreifanlegur, lýtur að því hversu takmarkaðar og brotakerndar birtingarmyndir standa til boða af íslenskum orðaforða, samsetningu hans og notkunareinkennum. Við þær aðstæður getur þótt nærtækt að njóta eftir föngum góðs af þeim ákvörðunum sem fyrir liggja í öðrum hliðstæðum orðabókum og spara að sama skapi fyrirhöfn við sjálfstæða efnisöflun. Um þennan vanda og viðbrögð við honum verður nánar fjallað hér á eftir, einkum í 3. og 4. kafla.

2.2.2 Markmálið: samsvaranir við annað mál

Þegar íslenska er í hlutverki markmáls eru aðstæður í grundvallaratriðum aðrar. Þar þarf að finna íslenskar samsvaranir við einingar annars tungumáls sem mótar flettiorðavalið á sínum forsendum. Sá íslenski orðaforði sem ratar inn í það umhverfi ber svip af því. Ýmis gild orð með skýru flettugildi koma jafnvel ekki við sögu á meðan

önnur og óvirkari orð eru dregin fram eða mynduð eru ný orð til að ná fram viðeigandi samsvörun í mynd íslensks jafnheitis. Annars er eðli málsins samkvæmt erfitt að fá heillega yfirsýn um orðaforðann og orðafarið markmálmegin, eins dreift og það er um orðabókartextann. Hér er hlutverk þess sem á heldur að hnitmiða jafnheitavalið eins og kostur er. Þar getur stundum virst fátt um vænlegar lausnir, í öðrum tilvikum virðast ólíkar leiðir í boði, a.m.k. við fyrstu sýn.

Einnig við þessar aðstæður er hætt við að erfitt reynist að framkalla heillegar myndir af íslenskum orðaforða og einstökum hlutum hans til stuðnings jafnheitavalinu. Að því leyti er enn erfiðara um vik að það orð eða orðafar sem rata skal að á sér engan formlegan íslenskan búning fyrir fram heldur tengist það og kviknar af merkingarlegum eigindum orða sem fengist er við í erlenda viðfangsmálinu. Íslenska samsvörun sem fyrst kemur upp í hugann þarf að máta við önnur hugsanleg jafnheiti á sama merkingarsviði en þar er yfirsýn oftast en ekki takmörkuð.

3 Efnisföng, efnisval og flettulisti

Í tvímála orðabók er teflt saman orðaforða tveggja tungumála með skírskotun til markhópa báðum megin þótt með ólíkum hætti sé. Markhópurinn markmálmegin nálgast lýsinguna fyrst og fremst til skilnings á einstökum orðum og orðasamböndum viðfangsmálsins og til glöggvunar á orðanotkun í því. Markhópurinn sem hefur viðfangsmálið að móðurmáli eða er handgenginn því leitar á hinn bóginn samsvarana við einingar þess (flettiorð og orðasambönd) í markmálinu.

Í íslensk-erlendri orðabók er þannig jöfnum höndum hugsað til þarfa íslenskra og erlendra notenda. Í fræðilegri umræðu um tvímála orðabækur hefur mjög verið fjallað um mismunandi notkunarhlutverk og vandkvæði á að þau fari saman í einu orðabókarverki (sbr. m.a. Svensén 2004:17–23). Þannig geti orðabók sem leggur áherslu á að greiða fyrir skilningi á merkingu orða í viðfangsmálinu ekki að sama skapi þjónað þeim sem leita upplýsinga um orðanotkun (notkunarþætti jafnheita) í markmálinu. Samkvæmt því hafa margar tvímála orðabækur verið samdar með býsna sérhæft hlutverk í huga. Slík aðgreining tekur reyndar mið af prentuðum orðabókum en í orðabókum á netinu er mun hægara að koma til móts við mismunandi þarfir ólíkra notendahópa, með breytileika í efnisröðun, beinni leit

að efnispáttum innan orðsgreina og innbyrðis tengingum. Hvort sem hægt er að líta svo á að gagnvart íslensku sé aðeins rúm fyrir eina tvímála orðabók í hvora átt um hvert málapar (t.d. eina íslensk-enska og eina ensk-íslenska, sbr. Sanders 2005:53) styrkir þetta mjög netið sem vænlegan miðil þegar hugsað er til nýrra tvímála íslenskra orðabóka.

3.1 Flettulistinn sem umgjörð um orðabókarefnið

Gagnvart íslensku í íslensk-erlendri orðabók er viðfangsefnið í upphafi að afla efniviðar til að byggja upp íslenskan flettulista og greina notkunarþætti flettiorðanna, aðgreina merkingarbrigði, tilgreina orðasambönd og lýsa einkennandi orðanotkun. Í þessu ferli er afmörkun flettulistans gjarna í fyrirrúmi, þar sem fyrst er úrskurðað um flettugildi orða í því safni orða sem efnisöflunin leggur til, með tilliti til áætlaðrar stærðar verksins og hlutverksins sem því er ætlað. Síðan er lýsing hverrar flettu tiltölulega sjálfstætt og afmarkað verkefni, þar sem gera þarf sérstaka atrennu að efnisöflun, velja úr tiltæku efni og setja lýsingunni hæfilegar skorður. Önnur aðkoma er að gefa orðanotkuninni, og þá sérstaklega merkingarbærum orðasamböndum, meiri gaum þegar við fyrstu afmörkun samhliða flettuvalinu.

Hvernig sem afmörkuninni er háttað kallar hún á viðeigandi verklag við lýsingu orðanna sem mynda flettulistann. Hætt er við að erfitt reynist að fá æskilega yfirsýn um samsetningu flettulistans svo að hægt sé að nálgast verkið í skipulegum áföngum. Þar togast m.a. á hagræðið að því að láta stafrófsröðina vísa veginn og meta út frá því hvernig verkinu miðar, fást við hvern og einn orðflokk sérstaklega og treysta með því innbyrðis samræmi eða fylgja eftir lýsingu tiltekinna merkingarflokka.

3.2 Þörf á merkingarflokkun

Almenn flokkun og greining flettulista og annarra orðalista takmarkast að mestu við orðflokkaskiptingu. Að auki dregur stafrófsröðin fram orðmyndunarleg vensl og þau vensl geta komið frekar í ljós ef leitað er tiltekinna bakliða samsettra orða, t.d. fyrirferðarmikilla nafnorðaliða eins og *-maður* eða lýsingarorða með liðum eins og *-látur* og *-laus*. Merkingarlegir flokkunarþættir eru hins vegar yfirleitt ekki fyrir hendi og þá ekki heldur gagnvart fleiryrtum merkingarbærum einingum sem gerð eru skil innan flettugreina.

Þörfin fyrir merkingarlega flokkun er þó næsta augljós enda er sam-
anburður viðfangsmáls og markmáls merkingarlegs eðlis. Vandinn er
hins vegar sá að ekki dugir að staðnæmast við flettimyndirnar ein-
ar heldur verður jafnframt að greina á milli merkingarbrigða tví- og
margræðra orða. Miðað við hefðbundið lýsingarferli eru því ekki for-
sendur fyrir merkingarlegri flokkun orðabókarefnisins fyrr en allri
margræðni hafa verið gerð skil.

Af þessu leiðir að merkingarleg flokkun verður ekki það viðfangs-
efni í orðabókarferlinu sem hún í raun þyrfti að vera, og reynslan úr
einu orðabókarverki nýtist að takmörkuðu leyti í öðru sambærilegu
verki. Gildi merkingarflokkunar snýr bæði að viðfangsmálinu sem
slíku og að samhengi þess við markmálið. Ef slík flokkun tekur til
tiltæks orða- og orðasambandasafns í heild sinni án tillits til síðari
takmörkunar og endanlegs efnisvals getur hún gegnt virku hlutverki
við sjálft flettuvalið, og sá orðaforði sem út af stendur er þá áfram í
greiningartengslum við flettulistann. Þegar kemur svo að lýsingu og
greiningu flettnanna skiptir miklu að hafa yfirsýn um merkingarskyld
orð og sambönd, svo að ná megi fram samfellu í greiningunni með
hliðsjón af skyldu orðafari. En gagnsemin er jafnvel enn áþreifanlegri
fyrir markmálið, því flokkunin dregur fram það orðafar í markmál-
inu sem til álita kemur að tilgreina sem jafnheiti og stuðlar þannig að
hnitmiðuðu jafnheitavali og markvissum skýringum.

3.3 Merkingarflokkun í *Íslensku orðaneti*

Merkingarflokkun af þessu tagi hefur að nokkru leyti sannað gildi sitt
við greiningu orðabókarefnisins í ISLEX (sjá Þórdís Úlfarsdóttir (2013
í þessu hefti)) en þar er byggt á viðtækari og margþættari flokkun á
mun stærra safni orða og orðasambanda í *Íslensku orðaneti* (sjá Jón
Hilmar Jónsson 2012). Gildi orðanetsins í þessu sambandi er að miklu
leyti fólgið í því hvernig flettimyndum þess og flettuskipan er háttað.
Annars vegar eru flettturnar merkingarlega einræðar þannig að hvert
merkingarbrigði er sjálfstæð eining. Hins vegar rúmar flettulistinn
auk stakra orða (einyrtra flettna) mikinn fjölda fleiryrrtra flettna þegar
um er að ræða merkingarbær orðasambönd. Þannig má greiðlega
tengja greiningu orðanetsins við flettiorð, einstök merkingarbrigði og
merkingarbær orðasambönd innan tiltekins orðabókartexta.

Auk merkingarflokka í mynd lýsandi hugtakaheita (nafnorða
eins og ÁHUGI, ÁHUGALEYSI, ÁHUGAMÁL, ÁHYGGJUR, ÁHYGGJULEYSI,
ÁHÆTTA) kemur merkingargreining orðanetsins fram í samheitum og

andheitum. Innan merkingarflokksins ÁHYGGJUR eru einyrtar flettur m.a. áhyggjur, búksorgir, hugarvöl og sút og meðal tvíyrtra flettna eru hafa áhyggjur, bera ugg í brjósti, vera uggandi um sinn hag, vera í öngum sínum, <honum, henni> vex <þetta> í augum og <þetta> stendur <mér> fyrir svefni. Flettan <þetta> stendur <mér> fyrir svefni er auk þess í samheitavenslum við flettuna <þetta> heldur vöku fyrir <mér> og sams konar vensl eru á milli flettnanna láta sér <þetta> í augum vaxa og mikla <vandann> fyrir sér.

Þannig getur Íslenskt orðanet stutt merkingarflokkun í greiningar- og vinnsluferli íslensk-erlendra orðabóka þótt gera verði ráð fyrir að slík flokkun geti seint orðið tæmandi. Í máltaekniverkefninu Íslenskur merkingarbrunnur (sjá Anna B. Nikulásdóttir 2012) er einnig fyrir hendi víðtæk merkingarflokkun á íslenskum orðaforða sem hagnýta má í sama skyni en þar er beitt öðrum og vélrænni aðferðum við greininguna. Gögnin sem þessi verkefni leggja til geta gegnt virku hlutverki þegar byggt er upp orða- og flettusafn íslensk-erlendar orðabókar og auðveldað efnisafmörkun sem hæfir því hlutverki sem orðabókinni er ætlað að gegna.

4 Formgerð og framsetning merkingarbærra orðasambanda

Hér að framan var minnst á orðflokkgreiningu sem helsta greiningarþátt orðabókarefnis samkvæmt rótgróinni hefð. Þótt orðflokkgreining eigi sér langa sögu hefur hún sjaldnast gegnt beinu flokkunarhlutverki og segja má að takmörkuð áhersla hafi verið lögð á sérstöðu einstakra orðflokka í almennum orðabókum. Í stafrænu umhverfi eru breyttar forsendur að þessu leyti, hægt er að skilyrða efnisflokkun og orðaleit við einstaka orðflokka og hægara er að samræma meðferð orðflokka í orðabókartextanum. En hefðbundin orðflokkgreining nær aðeins til stakra orða og með því verða fleiryrtar merkingarbærar einingar út undan og eiga sér ekki sambærilegan greiningarþátt.

Í Íslensku orðaneti er viðhöfð málfræðileg mörkun á fleiryrtum flettum sem gefur færi á formgerðarlegri flokkun þeirra (sjá Jón Hilmar Jónsson 2012:46-49). Mörkunin tekur að nokkru leyti mið af hinu orðabókarlega umhverfi og er í vissum atriðum einfaldari en venjubundin málfræðileg textamörkun. Auk orðflokkgreiningar er megináherslan á fallmörkun, lexíkalskan breytileika í orðasamböndum og innbyrðis

vensl setningarliða. Markið so <no-dg> fs no-a (sögn með eftirfarandi nafnorði í þágufalli með greini, að viðbætti forsetningu með nafnorði í þolfalli) sameinar t.d. flettur *koma* <húsinu> í *stand*, *stofna* <samkomulaginu> í *hættu*, *kasta* <hlutleysinu> fyrir *borð* og *skjóta* <ákvörðuninni> á *frest*, þar sem oddklofarnir afmarka lexíkalskan breytileika. Á sama hátt tengir markið <no-ag> so saman flettur *<gluggann>* *hrímar*, *<bátinn>* *fyllir*, *<sjóinn>* *gárar* og *<vindinn>* *hægir*. Formgerðarflokkunin kallast svo eftir atvikum á við merkingarlega greiningu og flokkun á fleiryrtum flettum og styður hana reyndar gjarna þar sem samfella í formgerð á sér oft samsvörun í merkingarlegum skyldleika.

Mörkunin byggist á því að fleiryrtar flettur í orðanetinu eru settar fram á staðlaðan hátt, m.a. með því að afmarka og auðkenna orðbundinn breytileika. Þessa meðferð fleiryrttra merkingarbærra eininga má nýta utan orðanetsins til greiningar og flokkunar á orðabókarefni svo að betri yfirsýn gefist um orðasambönd, fyrirferð þeirra og stöðu innan orðabókartextans. Slíkt notagildi er í sjálfu sér óháð því hvernig merkingarbærum orðasamböndum er fyrir komið í textanum og hvaða snið er haft á þeim. Hliðsjón af orðanetinu og samanburður við almennar orðabækur getur hins vegar sýnt hversu þröngt er um orðasambönd sem undirskipaðar einingar innan orðabókartexta og hversu erfitt er að gera þau sýnileg á sjálfstæðan hátt.

Atviksliðasambönd af ýmsu tagi eru skýrt dæmi um fleiryrtar merkingarbærar einingar sem full þörf er á að séu í sjónmáli innan viðfangsmálsins í tvímála orðabókartexta. Upphafseiningar slíkra sambanda eru gjarna forsetningar, atviksorð eða samtengingar, en þeim orðflokkum er jafnan gefið takmarkað rúm þar sem hið málfræðilega hlutverk viðkomandi orðs er í fyrirrúmi umfram einstök sambönd. Slík sambönd skila sér hins vegar ekki sem skyldi ef þeim er ætlað sæti undir gildari orðum, t.d. nafnorðum eða fornöfnum, og þar getur reynst erfitt að rata að einstökum samböndum í efnismiklum flettugreinum.

Á vinnslustigi getur einnig reynst torvelt að fá æskilega yfirsýn um margs kyns sagnasambönd með smáorði og sama má segja um orðtök í mynd sagnliða þar sem oft ber á breytileika í orðum og orðmyndum. Þar getur mörkun fleiryrttra eininga í stærra gagnasafni komið að góðum notum.

5 Notkunareinkenni og notkunardæmi

Forgreining orðabókargagna af því tagi sem lýst hefur verið hér að framan á fyrst og fremst að hafa almennt gildi og vera óháð því hvernig efnisskipan er háttað í einstökum orðabókum sem notfæra sér gögnin. Almennir flokkunarþættir, málfræðilegir sem merkingarlegir, veita aðhald og yfirsýn, hvort sem efnisskipan í endanlegum texta er með hefðbundnu sniði eða hún er látin mótast af greiningunni í ríkari mæli. En greiningin stuðlar að frjálsari efnisskipan þar sem merkingarleg einkenni geta fengið aukið vægi og fleiryrt sambönd haft meira sjálfstæði en hefð er fyrir eins og greinarhöfundur hefur áður lýst (sjá Jón Hilmar Jónsson 2005).

Hvernig sem efnisskipaninni er háttað er lýsing viðfangsmálsins að drjúgum hluta fólgin í því að draga fram einkennandi notkun orða og orðasambanda, m.a. með beinum dæmum. Fyrirferð notkunardæma getur verið afar breytileg og í raun er það á valdi ritstjóra hverju sinni hve miklu hlutverki þau eru látin gegna og hvernig þau eru sett fram. Í texta tvímála orðabókar er staða notkunardæma að því leyti önnur og sjálfstæðari en í einmála orðabók að dæmin styðjast ekki við undanfarandi merkingarskýringu á sama máli og skerpa þá mynd sem þar kemur fram, heldur staðfesta þau merkingarlega samsvörun í mynd erlends jafnheitis eða jafnheita. Notkunardæmin eiga því fyrst og fremst erindi við þá notendur sem nálgast viðfangsmálið sem erlent mál og hafa um leið vald á markmálinu.

Þessar aðstæður fela í sér ákveðinn vanda að því leyti að annars vegar þarf að velja og mynda notkunardæmin með skýrri yfirsýn um viðfangsmálið, íslenska orðanotkun, breytileika hennar og blæbrigði. Hins vegar kemur til þörf erlendra notenda fyrir að fá fram tiltekin megineinkenni í orðanotkun, sérstaklega með tilliti til þess að geta tileinkað sér málið og beitt því við raunverulegar aðstæður.

5.1 Form og framsetning notkunardæma

Form og framsetning notkunardæma getur verið með ólíkum hætti eftir því hvaða hlutverki þeim er ætlað að gegna. Þegar áherslan er á setningargerð er í norrænni orðabókarhefð gerður greinarmunur á „lífandi“ og „dauðum“ dæmum („levande og döda språkprov“, sjá Svensén 2004:193–195). Hin fyrrnefndu standa nær beinu notkunarsamhengi, með nafnliðum í mynd nafnorða eða persónufornafna

og sögnum í persónuhætti en hin síðarnefndu birta eins konar notkunarmunstur þar sem notkunarsamhengið er tilgreint á knappan og formfastan hátt, sagnir í nafnhætti og nafnliðir í mynd óákveðins fornafns. Í íslenskum orðabókum eru slík munsturdæmi einkum bundin skammstöfunum fornafnsins *einhver*, t.d. *hitta e-n*, *mæta e-m/e-u*, *gefa e-m e-ð*. Munsturdæmi draga fram tiltekin megineinkenni eins og fallstjórn en þau eru ekki til þess fallin að sýna merkingareinkenni og merkingarlega afmörkun nálægra orða eins og frumlags eða andlags sagna og því getur þótt ástæða til að fylgja þeim eftir með frjálsari dæmum. Þriðji kosturinn er að sameina þessi dæmahlutverk með því að tilgreina dæmigerða fulltrúa fylgdarorða með skýrum málfræðieinkennum í formföstum dæmum eins og m.a. er gert í *Stóru orðabókinni um íslenska málnotkun* (2005) og síðan í *Íslensku orðaneti*. Notkun persónufornafnsins í stað fornafnsins *einhver* gegnir þá því hlutverki að ljá dæmunum skýrari svip beinnar málnotkunar: *hitta* <*hann*, *hana*>, *mæta* <*honum*, *henni*>, *mæta* <*andstöðu*, *mótspyrnu*; *andstreymi*>, *gefa* <*honum*, *henni*> <*gjöf*; *mat*; *leyfi*>. Frá öðru sjónarhorni getur dæmigerð orðanotkun birst í virkum og áberandi orðastæðum, þar sem flettiorðið kemur fram með einkennandi fylgdarorði: *rennandi blautur* (undir *blautur*), *brimið svarrar* (undir *brim*), *detta kylliflatur* (undir *detta*).

5.2 Samhengið við merkingarbær orðasambönd

Hvaða framsetning sem valin er á notkunardæmum í íslensk-erlendri orðabók hvílir lýsing íslensku flettiorðanna að verulegu leyti á því dæmasafni sem þau eru látin styðjast við. En dæmaþörfin snertir ekki aðeins flettiorðin sem slík og breytilega notkun þeirra heldur getur hún einnig átt við um margs konar sjálfstæð og föst orðasambönd sem skipað er undir flettiorðin. Þá getur vandinn verið sá hvernig megi takmarka og einfalda lýsingu einstakra flettugreina og þar með dæmasafnið svo að notendur missi ekki yfirsýn um efnisatriðin í ofhlöðnum orðabókartexta.

Sú áskorun reynist að vonum erfið í prentuðum orðabókartexta, eins og margar íslensk-erlendar orðabækur vitna um. Raunin er oft sú að föstum merkingarbærum samböndum og almennari notkunardæmum orðanna er skipað saman, utan við aðgreiningu merkingarbrigða sem gerð er grein fyrir með ólíkum jafnheitum, og innbyrðis röð gjarna látin ráðast af forsetningum eða atviksorðum sem fram koma í samböndunum (sjá Jón Hilmar Jónsson 1995). Þar með rofn-

ar samhengi merkingareinkenna orðsins við setningargerðina og merkingarbær orðasambönd skera sig ekki úr hópi annarra sambanda orðsins sem sjálfstæðar einingar. Slík tilhögun er í alla staði óheppileg en þó sérstaklega fyrir þá notendur sem leita leiðsagnar um málbeitingu á íslensku.

5.3 Greiðari leið að dæmasöfnum

Í veforðabók má gefa notkunardæmum mun meira svigrúm og um leið skýrara og margþættara hlutverk en tók eru á í prentaðri orðabók. Þau þurfa t.d. ekki í öllum tilvikum að birtast á yfirborði orðabókartextans heldur verið framkallanleg ef notandinn kys að skerpa tiltekna umsögn og notkunareinkenni. Aðstæður leyfa einnig að merkingarbær orðasambönd standi sjálfstætt og komi eftir atvikum fram með sínum notkunardæmum. Þá verður sá kostur æ nærtækari að leita dæma utan við orðabókartextann með því að svipast um í flokkuðu safni textadæma viðkomandi orðs í mynd orðstöðulykils eða markaðrar málheildar (e. *tagged corpus*). Notendum dönsku vefsíðunnar dsl.dk/sprog (sem *Det danske sprog- og litteraturselskab* stendur að) gefst t.d. aðgangur að stórrí danskri málheild, *KorpusDK*, samhliða aðgangi að stórum dönskum orðabókarverkum. Þangað geta notendur orðabókanna sótt viðbótardæmi um einstök orð og virt nánar fyrir sér stöðu orðanna í setningarlegu samhengi.

Að því er orðanotkun og notkunardæmi varðar má nú auðveldlega leita fanga með beinni leit í tölvutækum textum, hvort sem það er gert í tengslum við orðabókarnotkun eða óháð þeim aðstæðum. Stundum á best við að leita fanga í ógreindu textasafni, t.d. hinu geysistöra safni *Tímarit.is* sem tekur til íslenskra blaða og tímarita frá upphafi til dagsins í dag. Í öðrum tilvikum hentar betur að nýta sér greiningu sem fram hefur farið á viðkomandi textum. *Íslenskur orðasjóður* (2012) er textasafn af því tagi, sótt í íslenskar vefsíður, þar sem setningarlegt umhverfi orðanna hefur verið greint og notendur geta nýtt sér þá greiningu, t.d. svipast um í hópi áberandi vinstri eða hægri nágranna orðsins og ratað þaðan að beinum textadæmum. *Íslenskt textasafn* Stofnunar Árna Magnússonar í íslenskum fræðum hefur að geyma fjölbreytta íslenska texta þar sem raða má textadæmum eftir orðmynd leitarorðsins eða í stafrófsröð eftir undanfarandi eða eftirfarandi orði. *Mörkuð íslensk málheild* sameinar svo safn íslenskra texta frá árabílinu 2000–2009 þar sem orðin hafa fengið rækilega málfræðilega greiningu og leita má dæma um einstök orð og orðmyndir í tengslum við

málfræðileg og setningarleg einkenni eða draga fram slík einkenni óháð því hvaða orð eiga í hlut.²

Í fyrrgreindum textasöfnum má ganga að forgreiningu á ólíku stigi og hafa hana til stuðnings og hliðsjónar við val og framsetningu notkunardæma í orðabókartexta. Slík greining stuðlar að markvissara dæmavali en hún hefur vitaskuld einnig verulegt gildi gagnvart öðrum þáttum orðabókartextans, svo sem skilum og breytileika í merkingu, áberandi orðastæðum og ekki síst einkennandi setningargerð.

5.4 Orðastæður og forgreining notkunarsambanda

Gagnvart orðabókarnotum er þörf á greiningu sem færir notkunar-sambönd nær búningi eiginlegrar orðabókar, dregur saman sambönd sömu formgerðar, sameinar merkingarskyldar orðastæður og gerir merkingarbær orðasambönd sjálfstæð og leitarbær. Í gagnagrunni *Ís-lensks orðanets* er slík greining fyrir hendi í verulegum mæli eins og að nokkru leyti hefur verið gerð grein fyrir hér að framan. Í flettuskipan orðanetsins birtist hún m.a. í tvíliða lýsingarorðaflettum þar sem nefnimynd lýsingarorðsins er augin með dæmigerðum fulltrúa eða fulltrúum nafnorðs sem lýsingarorðið á við: **sterkur [band, þráður], sterkur [ljós, birta], sterkur [vindur], sterkur [tilfinning]**. Þannig eru merkingarblæbrigði lýsingarorðsins aðgreind, og að baki einstökum flettum er að finna sambönd (orðastæður, e. *collocations*) með breytilegum nafnorðum. Þar sameinar t.d. flettan **sterkur [tilfinning]** sambönd eins og *sterk tilfinning*, *sterk kennd*, *sterk samkennd*, *sterkar ástriður*, *sterk innlifun* og *sterk löngun*. Nafnorðspáttur flettunnar getur svo haft sameinandi hlutverk gagnvart lýsingarorðum í sömu stöðu: **frískur [vindur], snarpur [vindur], stífur [vindur]**. Á sama hátt geta fleiryrtar sagnaflettur sameinað orðastæður með merkingarskyldum nafnorðum, eins og flettan **borga <gjaldið, skuldina, launin>** þar sem m.a. er að finna sambönd eins og *borga burðareyri*, *borga brúartoll*, *borga inngangseyri*, *borga félagsgjald* og *borga meðlag*. Það svið má svo enn víkka út með því að líta til nafnorða sem fram koma undir flettunni **greiða <gjaldið, skuldina>**. Þau eru sem vænta má að nokkru leyti hin sömu en þar koma einnig fram sambönd eins og *greiða burðargjald*, *greiða árstíllag*, *greiða lóðargjald*, *greiða sekt*, *greiða tekjuskatt*. Þannig getur merkingarflokkun flettnanna fært saman merk-

2 Verkefnið er enn í vinnslu en leitaradgangur er nú þegar að hluta efnisins á vefsíðunni <http://mim.hi.is/> (rúmlega 17 milljón lesmálsorðum af þeim 25 milljónum sem fyrirhugað er að verði í málheildinni fullbúinni).

ingarskyld orð og veitt yfirsýn um sambönd og orðastæður á víðu merkingarsviði.

6 Íslenska sem markmál

Þau orðabókargögn og sú gagnagreining sem hér hefur verið lýst hefur ekki einhliða gildi fyrir íslensku sem viðfangsmál í tvímála orðabókum. Raunar miðast hún ekki síður við einmála orðabókargerð þar sem gerð er grein fyrir íslenskum orðaforða og íslenskri orðanotkun á íslensku. Báðar orðabókartegundirnar hafa þörf fyrir greiðan aðgang að umfangsmiklu gagnaefni sem þær síðan velja úr og skipa á ólíkan hátt eftir því hvernig viðkomandi orðabókarverk er hugsað og afmarkað. En íslensku er einnig lýst á sinn hátt þegar henni er stillt upp gagnvart öðru viðfangsmáli í erlend-íslenskri orðabók. Eins og fyrr var dregið á eru aðstæður gjörólíkar í tvímála orðabók af því tagi, þar sem gagnþörfin miðast ekki við sjálfstæða efnisskipan heldur ræðst íslenska orðabókarefnið og meðferð þess af því hvernig háttar til í efnisskipan og framsetningu viðfangsmáls orðabókarinnar. Meginverkefnið er í því fólgið að velja viðeigandi íslensk jafnheiti við þau flettiorð, merkingarbrigði og orðasambönd sem þar koma fram. Ákvarðanir í því efni þurfa að geta stuðst við viðeigandi gögn um íslenskan orðaforða og orðanotkun og þar kemur greining af því tagi sem hér hefur verið fjallað um m.a. að notum.

6.1 Yfirsýn um íslensk jafnheiti

Íslenskt orðabókarefni í stöðu markmáls er að sjálfsögðu ósamfelldara en þegar um er ræða sjálfstæða efnisskipan gagnvart tilteknu markmáli. Í prentuðum orðabókartexta fæst engin heildstæð yfirsýn um íslenska orðafarið, hvorki orðaforðann sem slíkan og einkenni hans né stöðu einstakra jafnheita, að hvaða marki þau koma fram gagnvart ólíkum orðum og samböndum viðfangsmálsins, hvar brugðið er á það ráð að umorða tiltekna merkingu í stað þess að tilgreina jafnheiti o.s.frv. Þegar orðabókartextinn er í stafrænu formi getur fengist skýrari mynd en heildarmyndin skilar sér ekki nema með sérstakri gagnavinnslu þar sem íslensku jafnheitunum er raðað innbyrðis og þeim stillt upp andspænis þeim einingum viðfangsmálsins sem þau eru látin samsvara í orðabókartextanum.

Fullyrða má að sá íslenski orðaforði sem jafnheitin mynda og í ljós

kæmi við gagnagreiningu erlend-íslenskra orðabóka sé með nokkuð öðrum svip en sá sem valinn er þegar íslenska er viðfangsmálið. Þegar mótuð jafnheiti eru ekki tiltæk getur þurft að bregðast við með nýmyndunum af einhverju tagi, tökuþýðingum eða nýgervingum. Glíman við erlenda málið getur vakið áleitnar spurningar um aðild og viðurkenningu aðkomuorða og jafnheitavalið þannig fengið málpólitískt gildi. Orðmyndunarleg sérkenni hvors máls um sig torvelða oft jafnheitavalið og við þær aðstæður er mikilvægt að geta haft hliðsjón af fyrri lausnum og eftir atvikum fetað þar troðnar slóðir. Allt þetta undirstrikar gildi þess að að geta betur en nú gefst kostur á gengið að og hagnýtt það íslenska orðabókarefni sem erlend-íslenskar orðabækur leggja til. Þá er ekki einungis horft til orðabóka þeirrar tegundar því gagnlegt er að bera þann orðaforða sem hér er á ferðinni að orðaforða íslensk-erlendrar orðabókar. Þótt svipurinn sé ólíkur má gera ráð fyrir að meðal jafnheitanna komi fram orð sem ástæða sé til að bæta við flettulistann hinum megin, svo sem gegnsæjar mótaðar samsetningar sem ekki hafa augljóst flettugildi frá íslensku sjónarhorni en verða gildari að því leyti í ljósi erlendrar samsvörunar.

6.2 Greining á merkingarvenslum

Vandað og hnitmiðað val á jafnheitum er meginatriði við gerð tvímála orðabókar. Til þess að tryggja það sem best þarf að gefast yfirsýn um orðafar sem til greina kemur og það merkingarsvið sem um er að ræða hverju sinni. Í því sambandi getur *Íslensk samheitaorðabók* (2012) vísað veginn til merkingarskyldra orða en víðtækari og fjölþættari merkingarflokkun af því tagi sem áður var lýst og *Íslenskt orðanet* býr yfir er einnig mikilvæg stoð. Að auki gefur orðanetið færi á að meta stöðu einstakra orða með tilliti til merkingarskyldleika við önnur orð á grundvelli setningarlegra tengsla þeirra innbyrðis. Þar er fyrst og fremst um að ræða þann vitnisburð sem lesa má úr tengslum orðapara með tengiorðinu og meðal nafnorða og lýsingarorða (*fátækt og allsleysi, allsleysi og nauð; hátækur og umkomulaus, umkomulaus og einmana*). Vísbendingin um merkingarlega nálægð tiltekinna tveggja orða er fólgin í því hversu mörg orð þeim eru sameiginleg sem hinn hluti orðaparsins. Notkunardæmin eru að mestu sótt á vefsetrið *Tímarit.is* sem býr yfir afar ríkulegu efni til greiningar. Við samanburð á orðunum *glundroði* og *ringulreið* kemur t.d. í ljós að meðal förunauta þessara orða í orðapörum orðanestins (196 með *glundroði* og 217 með *ringulreið*) eru 102 orð sameiginleg báðum (t.d. bæði *glundroði* og *fálm*

og *ringulreið* og *fálm*). Að þessu leyti eru orðin tengdari innbyrðis en þau eru hvort um sig við önnur orð. Þegar tengslatíðnin er rakin áfram á bilinu 35–90 sameiginlegir förunautar er niðurstaðan fyrir *glundroði* þessi: upplausn 88, stjórnleysi 86, óstjórn 71, öngþveiti 62, ofbeldi 57, spilling 39, agaleysi 37. Orðið *ringulreið* tengist að mestu leyti sömu orðum á þessu bili: stjórnleysi 84, upplausn 84, öngþveiti 68, óstjórn 67, ofbeldi 65, vonleysi 43, agaleysi 38, hik 37, blóðsúthellingar 35.

Um leið og samanburðurinn staðfestir náinn merkingarskyldleika orðanna *glundroði* og *ringulreið* sýnir hann hvaða orð standa þeim næst hvoru um sig. Sé hins vegar litið til orðalistans í heild sem förunautarnir mynda og slíkum listum jafnvel slegið saman þegar merkingarskyld orð eiga í hlut myndast tengingar sem að meira eða minna leyti spanna heil merkingarsvið. Í því safni geta gægst fram orð sem við nánari athugun reynist full þörf á sem jafnheitum þótt þau leiti ekki á við fyrstu umhugsun.

7 Hugtaksleg nálgun og gagnkvæmni í lýsingu

Í texta tvímála orðabókar koma tengsl málanna fram í orðbundnu samhengi, þar sem viðfangsmálið myndar umgjörðina og ræður ferðinni en markmálið fylgir eftir með því að tefla fram samsvörunum við orð og orðasambönd viðfangsmálsins. Í þessari síendurteknu glímu við einstök orð eru hugtaka- og merkingarvensl málanna sífellt til skoðunar án þess að sú skoðun skilji eftir sig mikil ummerki í orðabókartextanum. Það skýrist að miklu leyti af því ójafnvægi sem er með málunum og að sú hugtaka- og merkingarflokkun sem á sér stað í orðabókarferlinu snertir aðeins annað málið hverju sinni og nær ekki að mynda heildir sem taka til beggja málanna.

Hér er á ferðinni einn helsti veikleiki tvímála orðabóka sem takmarkar notagildi þeirra í ýmsu samhengi (sbr. Anna Helga Hannesdóttir og Jón Hilmar Jónsson 2001 og Hanks 2012:60–61). Áherslan á jafngildi málanna og bein jafnheiti því til staðfestingar veldur því að fjölbreytni og breytileiki í mál- og orðanotkun skilar sér ekki sem skyldi, og orð og aðrar orðasafnseiningar sem ekki falla með góðu móti að jafnheitavenslum standa höllum fæti og verða hæglega út undan. Þetta á eðlilega í ríkari mæli við markmálið en fjölbreytileika viðfangsmálsins eru einnig takmörk sett.

7.1 Andspæni málanna í stafrænu umhverfi

Í stafrænu umhverfi eru aðstæður til að láta hið merkingar- og hugtakslega samhengi njóta sín betur og byggja upp lýsingu þar sem málin eru að meira eða minna leyti í jafnvægi. Snertipunktur málanna eru þá ekki jafngildin ein í formi orða heldur einnig og ekki síður hugtakavensl af ólíku tagi sem veita orðum og orðafari aðgang óháð því hvernig háttar til um bein jafnheiti. Slík lýsing getur að einhverju leyti sýnt að einstök hugtakasvið búa að misríkulegum orðaforða í ólíkum málum. Setningarleg og orðmyndunarleg sérstaða hvors máls um sig getur einnig komið mun skýrar fram.

Hugtaksbundin tvímála orðabókarlýsing af þessu tagi helst í hendur við þá merkingarflokkun og merkingargreiningu í hvoru máli fyrir sig sem lögð hefur verið áhersla á hér að framan sem undirstöðu nýrrar tvímála orðabókar. Hún leysir ekki orðbundna lýsingu af hólmi enda er notkunarskírskotunin að nokkru leyti ólík en með henni má veita skýrari innsýn í orðafar og orðanotkun og styrkja málbeitingarhlutverk orðabókarinnar.

7.2 Tvímála samanburður í ólíkum myndum

Víðtæk og rækileg forgreining á íslenskum orðabókargögnum skapar forsendur til að bera íslensku að öðrum tungumálum frá þessu sjónarhorni. Slíkur samanburður hefur sjálfstætt orðfræðilegt og orðabókarfræðilegt gildi en getur um leið verið liður í orðabókarverkefnum sem jafnframt er ætlunin að skili sér í hefðbundnari búningi. Í orðabókunum *Orðaheimi* (2002) og síðar *Stóru orðabókinni um íslenska málnotkun* (2005) er lagt inn á þessa braut gagnvart hugtakaflokkun íslenskra orðasambanda, þar sem íslensku hugtakaheitin í flettulista orðabókanna eiga sér samsvarandi heiti á ensku, dönsku og þýsku (í *Orðaheimi* er aðeins um að ræða ensku) sem fram koma í stafrófsröðuðum lista fyrir hvert tungumál. Með þessu er hugsað til þarfa erlendra notenda, m.a. nemenda í íslensku, sem vilja ná tökum á íslenskri málbeitingu og orðanotkun en hafa ekki sömu leitarforsendur og íslenskir notendur sem betur geta ratað að hugtakinu út frá tilteknu orðasambandi.

Hér beinist orðabókarlýsingin að einu máli og því ekki um markmál að ræða en erlendum notendum er opnuð leið að orðabókartextanum með eins konar markmálslegri skírskotun. Slíkar tengingar má mynda við fleiri mál og þær geta eins átt við um almennari hugtakaorðabók

sem nær í ríkari mæli til stakra orða. Stærra skref er að byggja upp tvímála lýsingu á grundvelli íslenska efnisins þar sem samstæður hugtakalisti tengir málin saman. Unnið hefur verið að íslensk-sænskri lýsingu af því tagi sem að vísu er enn ófullgerð en hefur sýnt fram á gildi þess að nálgast orðabókarlegan samanburð tveggja mála frá hugtakslegu sjónarhorni (sjá Anna Helga Hannesdóttir o.fl. 2010). Ætla má að óháð hugtakaskipan, þar sem jafnt er skírskotað til málanna tveggja, myndi enn betur tryggja jafnvægi í orðabókarlýsingunni.

8 Niðurlag

Hér hefur verið fjallað um efni og aðföng tvímála íslenskra orðabóka og hvernig aðstæður við gerð þeirra hafa mótað efnisskipan þeirra og framsetningu. Skortur á efnisgreiningu og efnisflokkun hefur komið í veg fyrir að nýjar orðabækur njóti í nægilegum mæli góðs af fyrri verkum og sá vandi blasir einkum við þegar horft er til eldri orðabóka í prentaðri mynd. Gögn sem liggja að baki endanlegum orðabókar-texta eru lítt eða ekki aðgengileg, og reynsla og innsæi orðabókarhöfunda skilar sér ekki á frjóan hátt í nýjum verkum.

Með tilkomu stafrænna gagna og veforðabóka hafa aðstæður breyst til hins betra að þessu leyti. Hægara er að halda utan um gögn sem gerð orðabókar byggist á og margvíslegt orðabókarefni er betur innan seilingar en áður. En enn meira máli skiptir að hægt er að greina og flokka orðabókarefni óháð tilteknum orðabókarnotum og án þeirrar ströngu takmörkunar sem jafnan gildir um orðabækur í endanlegri gerð og nýta slíka forgreiningu við vinnslu einstakra orðabóka.

Þörfin fyrir almenna greiningu og flokkun lýtur sérstaklega að merkingarþættinum og það á í ríkum mæli við um tvímála orðabækur hvort sem litið er til viðfangsmáls eða markmáls. Skýr yfirsýn um orð og orðafar á einstökum merkingarsviðum stuðlar að markvissum efnistökum og áfangaskiptingu og þegar markmál er annars vegar eru merkingarvensl oft ein til leiðsagnar um viðeigandi jafnheiti.

Gagnvart notkunarumhverfi orða, lýsingu orðasambanda og vali og framsetningu notkunardæma er forgreining með áherslu á sjálfstæði fleiryrttra merkingarbærra eininga og samræmda framsetningu orðasambanda mikilvægur leiðarvísir. Í *Íslensku orðaneti* hefur farið fram umfangsmikil greining af því tagi sem tvímála íslenskar orðabækur geta notið góðs af. Samræmd framsetning er þar forsenda

fyrir árangursríkri röðun og flokkun efnisins burt séð frá því hvaða tilhögun gildir í orðabókarverkum sem kjósa að nýta sér orðanetið.

Veforðabækur búa yfir möguleikum til að jafna stöðu viðfangsmáls og markmáls í tvímála orðabókum, losa þær að nokkru undan hinum orðbundnu venslum þar sem annað málið er í forgrunni og byggja þess í stað upp jafnvægari lýsingu þar sem málin tengjast fyrst og fremst á hugtakslegum grunni. Slík lýsing leysir ekki hefðbundnari tvímála orðabækur af hólmi en hún mætir mikilvægum þörfum margra notenda, sérstaklega með tilliti til máltileinkunar og yfirsýnar um einstök svið orðaforðans.

Ritaskrá

- Anna Helga Hannesdóttir og Jón Hilmar Jónsson. 2001. Að hafa í sig og á. Ísländsk fraseologi i ett isländskt-svenskt perspektiv. *LexicoNordica* 8:67–91.
- Anna Helga Hannesdóttir, Sofia Tingsell og Jón Hilmar Jónsson. 2010. Mot en begreppsaserat isländsk och svensk fraseologisk ordbok. Reflektioner kring pragmatiska idiom. Harry Lönnroth & Kristina Nikula (ritstj.), *Nordiska studier i lexikografi* 10. Rapport från Konferensen om lexikografi i Norden, Tammerfors 3–5 juni 2009, bls. 140–149. Tammerfors.
- Anna B. Nikulásdóttir. 2012. Tölvutækur merkingarbrunnur fyrir íslenska máltækni. *Orð og tunga* 14:19–28.
- Det danske sprog- og litteraturselskab*. Vefsíður um tungumál: www.dsl.dk/sprog.
- Dönsk-íslensk orðabók*. 1992. Ritstjórar: Hrefna Arnalds og Ingibjörg Johanne-sen. Reykjavík: Ísafoldarprentsmiðja. (Einnig á vefnum: http://snara.is/bls/um/_dais.aspx.)
- Ensk-íslensk orðabók með alfræðilegu ívafi*. 1984. Reykjavík: Örn og Örlygur. (Einnig á vefnum: http://snara.is/bls/um/_enisal.aspx.)
- Ensk-íslensk skólaorðabók*. 1986. Jón Skaptason, ritstjóri. Reykjavík: Örn og Örlygur.
- Ensk-íslenska orðabókin*. 2006. Jón Skaptason, ritstjóri. Reykjavík: JPV útgáfa. (Einnig á vefnum: http://snara.is/bls/um/_enisjpv.aspx.)
- Frönsk-íslensk orðabók*. 1995. Þór Stefánsson, ritstjóri. Reykjavík: Örn og Örlygur. (Einnig á vefnum: http://snara.is/bls/um/_fris.aspx.)
- Hanks, Patrick. 2012. Corpus evidence and electronic lexicography. Í: Sylviane Granger & Magali Paquot (ritstj.), *Electronic Lexicography*, bls. 57–82. Oxford: Oxford University Press.
- Helgi Haraldsson. 1996. *Rússnesk-íslensk orðabók*. Reykjavík: Nesútgáfan.
- ISLEX-orðabókin*. (e.d.) Þórdís Úlfarsdóttir, aðalritstjóri. Stofnun Árna Magn-ússonar í íslenskum fræðum. Á vefsíðunni www.islex.is. (Sótt 11. febrúar 2013.)

- Íslensk samheitaorðabók*. 2012. 3. útgáfa. Svavar Sigmundsson, ritstjóri. Styrktarsjóður Þórbergs Þórðarsonar og Margrétar Jónsdóttur. Reykjavík: Forlagið. (Einnig á vefnum: http://snara.is/bls/um/_samh.aspx.)
- Íslensk-ensk orðabók*. 1989. Sverrir Hólmarsson, Cristopher Sanders og John Tucker, ritstjórar. Reykjavík, Íðunn. [Önnur útgáfa, augin og endurskoðuð. 2009. Christopher Sanders, ritstjóri. Reykjavík: Forlagið.] (Einnig á vefnum: http://snara.is/bls/um/_isen.aspx.)
- Íslensk-ítölsk orðabók*. 1994. Paolo Maria Turchi, ritstjóri. Reykjavík: Íðunn. (Einnig á vefnum: http://snara.is/bls/um/_isit.aspx.)
- Íslensk-spænsk orðabók*. 2011. Guðrún Tulinius o.fl., ritstjórar. Reykjavík: Forlagið. (Einnig á vefnum: http://snara.is/bls/um/_issp.aspx.)
- Íslensk-pýsk orðabók*. 1993. Björn Ellertsson, ritstjóri. Reykjavík: Íðunn.
- Íslensk-frönsk orðabók*. 1950. Gerard Boots, ritstjóri. Reykjavík: Ísafoldarprentsmiðja.
- Íslensk-rússnesk orðabók*. 1962. Valerij P. Bérkov, ritstjóri. Moskva.
- Íslenskt orðanet*. (e.d.) Jón Hilmar Jónsson og Þórdís Úlfarsdóttir. Stofnun Árna Magnússonar í íslenskum fræðum. Á vefsíðunni <http://ordanet.is/>. (Sótt 28. febrúar 2013.)
- Íslenskt textasafn*. Stofnun Árna Magnússonar í íslenskum fræðum. Á vefsíðunni <http://www.lexis.hi.is/corpus/leit.pl>.
- Íslenskur orðasjóður* 2012. Á vefsíðunni: http://wortschatz.uni-leipzig.de/ws_isl/.
- Ítölsk-íslensk orðabók*. 1999. Paolo Maria Turchi, ritstjóri. Reykjavík: Íðunn. (Einnig á vefnum: http://snara.is/bls/um/_itis.aspx.)
- Jón Hilmar Jónsson. 1995. Verbgrammatikk i islandske tospråklige ordbøker med islandsk som kildespråk. *LexicoNordica* 2:65–78.
- Jón Hilmar Jónsson. 2005. Aðgangur og efnisskipan í íslensk-erlendum orðabókum – vandi og valkostir. *Orð og tunga* 7:21–40.
- Jón Hilmar Jónsson. 2012. Að fanga orðaforðann: orðanet í þágu orðabókar. *Orð og tunga* 14:39–65.
- Mörkuð íslensk málheild*. Stofnun Árna Magnússonar í íslenskum fræðum: http://www.arnastofnun.is/page/morkud_islensk_malheild.
- Norsk-islandsk ordbok*. 1987. Hróbjartur Einarsson, ritstjóri. Oslo: Universitetsforlaget.
- Orðaheimur*. 2002. Jón Hilmar Jónsson, ritstjóri. Reykjavík: JPV útgáfa.
- Sanders, Cristopher. 2005. Bilingual Dictionaries of Icelandic: Types of Users and their Different Needs – a Discussion. *Orð og tunga* 7: 41–57.
- Snara vefbókasafn*. Á vefsíðunni: <http://snara.is/>.
- Spænsk-íslensk orðabók*. 2007. Guðrún Tulinius o.fl., ritstjórar. Reykjavík: Forlagið. (Einnig á vefnum: http://snara.is/bls/um/_spis.aspx.)
- Stóra orðabókin um íslenska málnotkun*. 2005. Jón Hilmar Jónsson, ritstjóri. Reykjavík: JPV útgáfa. (Einnig á vefnum: http://snara.is/bls/um/_stor.aspx.)
- Svensén, Bo. 2004. *Handbok i lexikografi*. Stockholm: Norstedts.

- Svensk-isländsk ordbok*. 1982. Gösta Holm og Aðalsteinn Davíðsson, ritstjórar. Lund: Bröderna Ekstrands Tryckeri AB.
- Tímarit.is*. Landsbókasafn Íslands – Háskólabókasafn. Á vefsíðunni: timarit.is.
- Þórdís Úlfarsdóttir. 2013. ISLEX – norræn margmála orðabók. Orð og tunga 15. 41–71. (Þetta hefti.)
- Þýsk-íslensk orðabók*. 2008. Heimir Steinarsson, ritstjóri. Reykjavík: Opna. (Einnig á vefnum: http://snara.is/bls/um/_deis.aspx.)
- Þýsk-íslensk/íslensk-þýsk orðabók*. 2004. Steinar Matthíasson, ritstjóri. Reykjavík: Iðnú.

Lykilorð

tvímála orðabækur, jafnheiti, merkingarleg vensl, hugtakaorðabók

Keywords

bilingual dictionaries, equivalents, semantic relations, thesaurus

Abstract

The paper describes the collection and presentation of Icelandic lexicographic material in bilingual Icelandic dictionaries, especially where Icelandic plays the role of the source language. The composition of the Icelandic lemma list is discussed where the lack of semantic analysis and classification has made it difficult to detect internal relations and to concentrate on words and phrases belonging to a specific concept or semantic field. Here the lexicographic description of Icelandic can benefit from a comprehensive analysis and presentation of the structural and semantic relations of the Icelandic vocabulary as found in the project *Íslenskt orðanet*. On the basis of a large collection of collocations the patterns of use reveal the interaction of semantic and syntactic relations, building groups of semantically coherent words and phrases. In accordance with the condition of monosemic lemmas the lemma list of *Íslenskt orðanet* contains many multi-word units, showing structural behavior of the lexical elements involved.

Sorting and classification of such lexicographic material which has gone through a proactive analysis, together with corpus evidence, will improve the usefulness of new bilingual Icelandic dictionaries in the future.

Jón Hilmar Jónsson

Stofnun Árna Magnússonar í íslenskum fræðum / Háskóli Íslands

Neshaga 16

107 Reykjavík

jhj@hi.is