

Orð og tunga

8



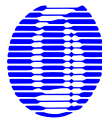
# Orð og tunga

## 8

Ritstjóri  
Guðrún Kvaran

Ritnefnd

Ásta Svavarsdóttir, Jón Hilmar Jónsson, Veturliði Óskarsson



Orðabók Háskólans  
Reykjavík 2006

© Orðabók Háskólans, 2006

Öll réttindi áskilin.

Bók þessa má eigi afrita með neinum hætti, svo sem ljósmyndun, prentun, hljóðritun eða á annan sambærilegan hátt, að hluta eða í heild, án skriflegs leyfis höfunda og útgefanda.

ISBN 9979-54-707-3

Umbrot: Bessi Aðalsteinsson.

Hönnun kápu: Björg Vilhjálmsdóttir.

Prentun og bókband: Leturprent.

# Efnisyfirlit

Formáli ritstjóra .....	vii
<b>Pemagreinar:</b>	
Birna Arnbjörnsdóttir: <b>Orðabækur, málfræðigrunnur og netkennsla</b> .....	9
Kristín Bjarnadóttir: <b>Málfræði í orðabókum</b> .....	27
Matthew Whelpton: <b>Argument Structure — For Mental Dictionaries Only?</b> .....	45
<b>Aðrar greinar:</b>	
Haraldur Bernharðsson: <b>Gás, gæs og Gásir, Gásar Brot úr hljóðsögu og beygingarsögu</b> .....	59
Katrín Axelsdóttir: <b>Hvað er klukkan?</b> .....	93
Margrét Jónsdóttir: <b>Um ærsl, busl og usl í orðasamböndum</b> .....	105
Þórdís Úlfarsdóttir: <b>Málfræðileg mörkun orðasambanda</b> ...	117
<b>Orðabókar- og rannsóknarverkefni:</b>	
<b>Beygingarlýsing íslensks nútímamáls</b> (Kristín Bjarnadóttir)	145
<b>ISLEX – Íslensk-norræn veforðabók</b> (Þórdís Úlfarsdóttir) ..	146
<b>Mörkuð íslensk málheild</b> (Sigrún Helgadóttir) .....	148
<b>Rannsókn á aðkomuorðum í Norðurlandamálum</b> (Ásta Svavarsdóttir) .....	151
<b>Spænsk-íslensk, íslensk-spænsk orðabók</b> (Guðrún H. Tulinius og Margrét Jónsdóttir) .....	153

Tilbrigði í setningagerð (Ásta Svavarsdóttir) .....	156
Tungutækni-verkefni sem Orðabók Háskólans tekur þátt í (Eiríkur Rögnvaldsson) .....	157
Bókafregnir .....	160

## Tilkynning

NFL - Norrænt félag um orðabókafraeði heldur 9. norrænu orðabókaráðstefnuna á Akureyri 6.–9. júní 2007.

Ráðstefnur þessar eru haldnar annað hvort ár og er þetta í annað sinn sem Íslendingar eru gestgjafar, í fyrra skiptið var ráðstefnan haldin í Reykjavík 7.–10. júní 1995. Ráðstefnur þessar eru vettvangur fyrir orðabókafólk á Norðurlöndunum til að kynna viðfangsefni sín og fræðast um orðabókaverkefni í nágrannalöndunum. Fyrirlestrarnir eru að jafnaði 30 til 40 og eru gefnir út í vönduðu ráðstefnuriti í flokknum *Nordiske studier i leksikografi*. Þarna gefst mönnum gott tækifæri til að koma rannsóknum sínum á framfæri og sjálfsagt fyrir íslenskt orðabókafólk að nýta sér að ráðstefnan er haldin hérlendis að þessu sinni.

Formaður félagsins er Daninn Lars Trap-Jensen og með honum í stjórn sitja Carsten Hansen frá Danmörku, Kristina Nikula frá Finnlandi, Dagfinn Worren frá Noregi og Birgit Eaker frá Svíþjóð ásamt undirritaðri.

Nánari upplýsingar um félagið er að finna á heimasíðu þess: <http://www.nordisk-sprakrad.no/nfl.htm>.

Halldóra Jónsdóttir, varaformaður NFL  
halldora@lexis.hi.is

# Formáli ritstjóra

*Orð og tunga* kemur nú út í annað sinn með breyttu sniði en ákveðið var á síðasta ári að breyta útliti ritsins annars vegar og taka upp ritrýningu greina hins vegar sem felst í því að allar greinar eru lesnar yfir af ónafngreindum sérfræðingum. Sú nýbreytni var einnig tekin upp á síðasta ári að binda fyrri hluta ritsins ákveðnu þema. Breytingarnar á tímaritinu hafa mælst vel fyrir og er því þessi 8. árgangur með sama sniði og síðast.

Þema þessa heftis var rætt á málstofu sem fram fór 1. apríl 2005 á vegum Orðabókar Háskólans undir heitinu *Málfræði og orðabækur*. Þau sem erindi fluttu voru Birna Arnbjörnsdóttir, Jóhannes Gísli Jónsson, Kristín Bjarnadóttir og Matthew Welpton. Í heftinu birtast þrjár greinar sem unnar voru upp úr málstofufyrirlestrunum.

Auk þemagreinanna birtast nú í ritinu fjórar greinar á sviði orðfræði og orðabókarfræði. Höfundar eru Haraldur Bernharðsson, Katrín Axelsdóttir, Margrét Jónsdóttir og Þórdís Úlfarsdóttir.

Ætlunin er framvegis að kynna árlega í heftinu helstu orðabókarverkefni og rannsóknarverkefni sem unnið er að og tengjast orðabókargerð eða styrkja hana á einhvern hátt. Að þessu sinni eru kynnt sjö verkefni.

Allur undirbúningur undir prentun *Orðs og tungu* fór fram á Orðabók Háskólans og var það verk eins og áður í höndum Bessa Aðalsteinssonar. Björg Vilhjálmsdóttir grafískur hönnuður, sem hannaði nýtt útlit fyrir 7. árganginn, vann einnig kápuna á þetta hefti. Háskólaútgáfan annaðist eins og áður prentun og dreifingu.

Guðrún Kvaran





Birna Arnbjörnsdóttir

# Orðabækur, málfræðigrunnar og netkennsla

## 1 Inngangur

Við Hugvísindastofnun Háskóla Íslands hefur á undanförunum fjórum árum verið í þróun kennsluforritið *Icelandic Online* sem er námskeið í íslensku sem erlendu máli með vefaðgangi. Kennsluforritið var gefið út í þremur hlutum á árunum 2004 og 2005. Fljótlega eftir að hönnun *Icelandic Online* hófst árið 2001 varð ljóst að þróa þyrfti hjálparmiðla svo sem málfræðigrunn og tvímála orðasöfn nemendum til stuðnings og að miðlana þyrfti að laga að málfærni nemenda sem eru byrjendur í íslenskunáminu sem og að innihaldi námskeiðsins. Þær hugleiðingar sem fram eru settar hér að neðan eru byggðar á rannsóknum og reynslu tengdri þróun *Icelandic Online* og umræddra hjálpargagna. Sérstaklega er hugað að því hvernig framsetning hjálpargagna á netinu getur komið byrjendum í námi beygingarmála sem best að gagni enda er óhjákvæmilegt að slík gögn eða hjálparmiðlar skipi hærrí sess í kennslu og námi mála þar sem orð hafa margar birtingarmyndir en við nám tungumála sem ekki teljast beygingarmál. Í framhaldi af því verður reynt að draga fram muninn á orða- og málfræðibókum sem hefðbundnum uppsláttarritum annars vegar og orða- og málfræðibókum sem handbókum í tungumálanámi hins vegar og vísað í því sambandi til hvernig hugmyndir um notkun hjálparmiðla í tungumálanámi samræmast nútíma kenningum um tileinkun

erlendra mála. Þá verður sagt frá því hvernig orðalistar og málfræðigrunnur voru sérhannaðir og aðlagðir málfræni nemenda og innihaldi *Icelandic Online* en hönnunin tók mið af kenningum um máltil-einkun fullorðinna og þróun í kennslu og námi tungumála á netinu sérstaklega. Að lokum verða færð rök fyrir því að haldgóðar upplýsingar um málfræði eigi heima í orðabókum hvort sem þær eru þróaðar fyrir notkun á netinu eða ekki og velt upp hugmyndum að málheild eða gagnabanka um íslenskt mál á Internetinu sem væri uppbyggður í ákveðnum lögum á svipaðan hátt og gert er í *Icelandic Online* þannig að úr honum mætti sækja upplýsingar við hæfi notandans hvort sem hann er málfræðingur að sinna rannsóknum eða nemandi að stíga sín fyrstu skref í tileinkun íslensku sem erlends máls.

## 2 Icelandic Online

*Icelandic Online I* er kennsluforrit með vefaðgangi sem ætlað er byrjendum í námi í íslensku sem erlendu máli. Námskeiðið, sem var styrkt af Lingua 2 áætlun Evrópusambandsins, Rannís, NordPlus, sjóðum Háskóla Íslands o.fl. var þróað við Háskóla Íslands og er samstarfsverkefni Hugvísindastofnunar Háskólans, Stofnunar Sigurðar Nordals og íslenskuskorar auk sex erlendra háskóla en höfundur er verkefnisstjóri. Námskeiðið er þróað með ákveðinn skilgreindan markhóp í huga, þ. e. háskólastúdenta sem eru að hefja íslenskunám á Íslandi eða erlendis en námskeiðið hefur einnig reynst vel í kennslu fullorðinna innflytjenda auk nemenda með annað móðurmál en íslensku í framhaldsskólum landsins. Námskeiðið er jafngildi 45 klukkustunda náms. Nemendur í *Icelandic Online I* geta valið um tvö þemu, náttúru og menningu, eða í allt um níu hundruð námsviðföng í sex bálkum. Þar sem markhópurinn var skýrt skilgreindur var unnt að gera námskeiðið markvissara og raunhæfara með tilliti til innihalds og markmiða, m.a. með því að tekið var mið af viðurkenndum kenningum um hvernig fullorðnir tileinka sér tungumál, áhugasviðum háskólanema og því málumhverfi sem líklegt var að nemendur þyrftu að bjarga sér í með því að nota íslensku. Í *Icelandic Online I* er að auki reynt að auðvelda nemendum skilning á uppbyggingu íslensks máls með nýrri aðferðafræði í framsetningu á málfræði. Aðferðin er ný nálgun í kennslu beygingarmála þar sem fallorð hafa mörg birtingarform og því óhjákvæmilegt að áhersla á málfræði sé meiri strax í upphafi náms en í

kennslu tungumála sem ekki eru þannig uppbyggð. Í því samhengi þarf sérstaklega að huga að því að setja fram málfræðiupplýsingar og orðskýringar í merkingarlegu samhengi í námsefninu þ. e. að nemandinn vinni með málfræðiupplýsingar í tengslum við málnotkun og að upplýsingarnar séu einfaldaðar og aðlagðar bæði að færni nemandans í tungumálinu og að námsefninu. Þetta á sérstaklega við á fyrstu stigum náms. Með notkun tölva má auðveldlega gera flókin málfræðileg atriði sýnileg á gagnvirkan hátt. Með aðstoð myndefnis og jafnvel hreyfimynda má skýra orð og orðasambönd þannig að byrjandinn átti sig á merkingu þeirra, sérstaklega þegar upplýsingarnar tengjast beint efninu sem nemandinn vinnur með hverju sinni. Hér er átt við framsetningu málfræði og orðskýringa í textanum eða námsefninu sjálfu en hið sama á reyndar við um framsetningu upplýsinga í hjálparmiðlum. Hér á eftir verða sýnd dæmi um slíka aðlögun og dreginn fram munurinn á framsetningu upplýsinga úr orðasöfnum, alfræðisöfnum og málfræðigrunnum á netinu sem nýtast sem handbækur á fyrstu stigum tungumálanáms og hefðbundinni framsetningu málfræði og orðskýringa í uppsláttarritum.

### **3 Orðabækur sem uppsláttarrit, orðabækur sem handbækur og kenningar um máltileinkun**

Markmið orðabóka sem uppsláttarrita er m.a. að gefa tæmandi lýsingu á merkingu orða ásamt nauðsynlegum upplýsingum um notkun þeirra. Oftast er þá gert ráð fyrir að notendur hafi góða færni í málinu sem unnið er með. Notkun uppsláttarrita af þessu tagi til kennslu samræmast markmiðum hugsmíðahyggjunnar (Bruner 1974) en aðalinntak þeirrar kenningar er að hafi nemandinn greiðan aðgang að upplýsingum muni hann byggja upp sinn eigin þekkingargrunn. Í tungumálanámi og -kennslu birtist hugmyndafræði hugsmíðahyggjunnar í því formi að nægilegt *ilag* í málumhverfinu, þ. e. greiður aðgangur að markmálinu, nægi til máltileinkunar. Þetta má sjá m.a. í aukinni áherslu á notkun gagnabanka eða málheilda í tungumálakennslu sem gefa nemandanum tækifæri til að uppgötva sjálfur formgerð málsins út frá málnotkunardæmum, sérstaklega klösom (Ellis 2003). Með tilkomu veraldarvefsins hefur aðgengi að rauntextum aukist til muna sem aftur hefur aukið möguleika nemandans á að efla

færni í erlendum tungumálum með því að vafra um veraldarvefinn í vefleiðöngnum og grúska í gagnagrunnum á markmálinu. Greiður aðgangur að rauntextum ásamt tækifærum til að spjalla á markmálinu á þar til gerðum rásum við innfædda og aðra nemendur og koma sér saman um merkingu orða og orðasambanda (negotiate meaning) er undirstaða nýrrar nálgunar í tileinkun tungumála gegnum tölvur sem kallast *Networked Language Learning* (Warschauer 1998; Warschauer og Kern 2000; Kern, Ware og Warschauer 2004). Þessi nálgun leggur aðaláherslu á tjáskipti sem undirstöðu máltileinkunar en sú nálgun hefur verið ríkjandi í tungumálakennslu undanfarin ár (Lantolf 2000). Munurinn á hefðbundnum tjáskiptamiðuðum kennsluáðferðum og *Networked Language Learning* er að síðari nálgunin á við samskipti manns og tölvu.

Samskiptanetanálgunin (*Networked Language Learning*) hentar illa byrjendum í tungumálanámi því þeir hafa ekki nægilega málfærni til gagnlegra samskipta. Hún hentar líka illa í kennslu beygingarmála þar sem margbreytileiki birtingarforma torveldar skilning og tjáningu. Það er erfitt að þjálfra flæði (auka sjálfvirkni) í tjáskiptum þegar útskýra þarf strax frá upphafi hvers vegna svo mörg orð hafi svo mörg birtingarform.

Dæmi um notkun gagnabanka og internetsins til tungumálanáms má sjá í aukinni notkun leitarvéla t.d. *Google* sem uppsláttarrits, m.a. til að sannprófa reglur og hefðir um málnotkun, t.a.m. í ritun (Hubbard 2004). En notkun gagnabanka og opinna alfræðibóka (*wikis*) og almennra leitarvéla er afar óreiðanleg og krefst í raun nokkurrar færni í tungumálinu til þess að hún gagnist nemandanum. Til dæmis mætti auðveldlega kynna sér hvaða birtingarform nafnorðið *kona* í fleirtölu með greini hefur með forsetningunni *frá*. Er það *frá konunum* eða *frá konurnar*? Notandinn slær þá inn á *Google*: '*frá konunum*' fyrst og í ljós kemur að það kemur fyrir 917 sinnum á íslenskum vefsíðum. En ef slegið er inn *frá konurnar* kemur það aldrei fyrir (leit gerð á [www.google.com](http://www.google.com) 25. ágúst 2005). Af þessu má draga þá ályktun að birtingarform nafnorðsins *kona* í fleirtölu með greini á eftir forsetningunni *frá* er *konunum*. Þessi aðferð er þó ekki alltaf jafn örugg og sést það á prófi ef slegið er inn *þau langar* og *þeim langar*. Hið fyrra kemur fyrir 70.600 sinnum á íslenskum vef- og blogg síðum en hið síðara á 82.200 síðum (leit gerð á [www.google.com](http://www.google.com) 25. ágúst 2005). Þó að þetta sé hugsanlega vísbending um raunverulega málnotkun er þessi leið síður lík-

leg til að hjálpa byrjanda í íslenskunámi til að byggja upp málkerfi íslenskunnar. Hér er gengið út frá því að markmið íslenskukennslu sé að kenna nemendum staðlað nútímamál.

Að sumu leyti virðist þessi nýja hugmyndafræði, þ.e. hugsmíða-hyggjan og samskiptanetanálgunin, ganga í berhögg við niðurstöður rannsókna á máltileinkun fullorðinna sem hafa sýnt að þrátt fyrir góðan vilja og næg tækifæri og, að því er virðist, mikið ílag s.s. þegar nemandi er að tileinka sér tungumálið sem talað er í málumhverfi hans, virðist fullorðið fólk ekki læra tungumál með því einu að nota það eða leika sér með það (Schmidt and Frota 1986). Það eitt og sér að hafa nægilegt ílag í umhverfinu, þ. e. nægilegan aðgang að tungumálinu virðist ekki nægja til að nemendur tileinki sér formgerð málsins, a. m. k. ekki fullorðnir nemendur (Schmidt og Frota 1986, Doughty og Williams 1998). Þetta hefur áratuga áhersla á flæði og tjáskipti í tungumálakennslu sýnt. Aukið flæði hefur fengist á kostnað hittni, þ. e. þó að nemendur hafi aukið færni sína til að tjá sig á markmálinu hefur það ekki að sama skapi aukið getu nemandans til að tala rétt. Þetta á sérstaklega við um beygingamál. Í ljósi þessarar reynslu hefur áhugi fræðimanna aukist á undanförunum árum á orsökum og eðli stöðunar (fossilization, end state) í tungumálanámi því að ljóst er að máltileinkun hjá langflestum fullorðnum staðnar áður en millimálið nær að endurspeglar málkerfi markmálsins fullkomlega (Bley-Vroman 1989, Lardiere 1998, Birdsong 1999).

Varhugavert er að yfirfæra kennslunálgun sem byggist á því að efla málfærni með því að auka tækifæri til tjáskipta manna í millum (augliti til auglitis) og þá oftast en ekki með kennara sem kann eitthvað fyrir sér í markmálinu beint á aðstæður þar sem tjáskipti eru milli nemanda og tölvu. Margir sem aðhyllast tjáskiptanálgun í tungumálakennslu hafa haldið því fram að málfræðikennsla væri næsta óþörf (Krashen 1989) og þar með mætti draga þá ályktun að óþarfi sé að hafa upplýsingar um málfræði í orðabókum. Þetta hljómar eins og ákveðin þversögn — en þegar hugsmíðahyggjumenn tala um að greiður aðgangur að tungumálinu sé forsenda máltileinkunar eiga þeir við tungumálið sjálft en ekki hefðbundnar málfræðiútskýringar — eins og dæmið af Google hér á undan sýndi. En þegar um sjálfsnám á netinu er að ræða, án kennara eða leiðbeinanda, verður það ennþá mikilvægt að nemandinn hafi aðgang að réttum upplýsingum um tungumálið og þurfi ekki að reiða sig á misgóðar heimasíður og opin alfræðisöfn

(wikis). En forsenda þess að geta nýtt sér slíka miðla er að nemandinn hafi grunnfærni í markmálinu.

Til þess að byrjandinn hafi sem mest gagn af hjálparmiðlum, s.s. orðabókum og málfræðigrunnum, þarf að laga upplýsingarnar sem í þeim er að finna að málfærni nemandans. Þetta er mögulegt gegnum tölvur. Tilgangur orðabóka sem handbóka er að vera leiðbeinandi um uppbyggingu og notkun tungumálsins en notendur eru nemendur í erlendum tungumálum. Notkun handbóka um málfræði og málnotkun samræmast kenningum um að fullorðnir hafi gagn af leiðbeiningum um uppbyggingu málsins, þ. e. hafi gagn af málfræðikennslu (Doughty and Williams 1998). Áhersla á málfræði verður sjálfkrafa meiri í beygingarmálum og því vert að skoða notkun hjálparmiðla sérstaklega þegar kemur að tileinkun slíkra tungumála, m.a. íslensku.

#### 4 Hlutur málfræði í orðasöfnum

Umræðan um máltileinkun fullorðinna, hvort hún er sérhæfð (Jackendoff 1994) eða hvort tungumál lærast eins og hver annar lærdómur (sambandshyggja/connectionism) (Rumelhart og McClelland. 1986) byggist að einhverju leyti á því hvernig maðurinn tileinkar sér beygingar. Pinker og Prince (1988, 1994) bentu á að ef til vill væru ekki allar beygingarendingar leiddar af reglum heldur þyrfti nemandinn að læra sumar utan að og væru þær þá merktar sérstaklega í orðaforða (Lexicon). Þetta ætti sérstaklega við um óreglulegar beygingar. Ef svo er leiðir það af sjálfu sér að kenna þarf reglulegar endingar á annan hátt en óreglulegar endingar, þ. e. kenna þarf reglur og hvernig á að beita þeim annars vegar en ganga út frá því að aðrar beygingar þurfi að læra utanbókar og haga kennslu í samræmi við það. Það gefur augaleið að það hjálpar nemandanum við námið að segja honum að það sé góð þumalputtaregla að þágufall fleirtölu nafnorða hefur alltaf endinguna *-um* í stað þess að bíða eftir því að hann komist að þessu sjálfur m.a. með því að leita á Google (Birna Arnbjörnsdóttir 2003, 2004). Annar háttur verður þá hafður á þegar aðstoða á nemandann við að tileinka sér óreglulegar beygingar. Ef upplýsingar um óreglulegar beygingar geymast með einstökum orðum í orðaforða og nemandinn verður að læra utan að, tengist tileinkun þeirra tíðni í flagi. Því oftast sem nemandinn hittir fyrir óreglulega beygingu, því líklegri er hann til að tileinka sér beyginguna. Óreglulegar endingar þurfa því

að koma oftar fyrir í námsefninu til að auðvelda nemandanum að læra þær utan að. Framsetning málfræði í orðabókum, málfræðigrunnum og í kennslu verður þá í samræmi við það. Hér skal það tekið fram að íslenskan er merkilegt rannsóknarefni frá sjónarmiði máltileinkunar fyrir þær sakir að veikar og sterkar beygingar íslenskra fallorða falla ekki sérlega vel að þessari almennu flokkun í reglulegar og óreglulegar beygingar. Clahsen og fleiri (2003) hafa stungið upp á því að skiptingin sé ekki svona skýr heldur séu algengar virkar endingar bundnar reglum en aðrar ekki. Sambandsbyggjumenn (Rumelhart og McClelland 1986) telja tileinkun beyginga betur lýst sem einu allsherjar námskerfi þar sem beygingarform orða safnist saman um kjarnaorð eða grunnmynd í orðasafni og að tengingin milli taugunga styrkist og treystist með aukinni notkun þar til notkunin er orðin sjálfvirk. Þessar kenningar hafa áhrif á framsetningu hagnýtrar málfræði og kennslu beygingarmála. Þess má geta að rannsóknir m.a. Kolbrúnar Friðriksdóttur (2004), sem hafa beinst að því að skoða það ferli sem nemendur fylgja við að tileinka sér nafnorðabeygingar í íslensku, benda til þess að það taki nemendur níu mánuði að átta sig á algengum fyrirsegjanlegum beygingum þrátt fyrir að þeir hafi fengið beina málfræðikennslu.

## 5 Málfræði og orðasafn lagað að málfærni nemandans og að námsefninu

Hér er að ofan hefur verið vitnað í rannsóknir sem sýna að fullorðnir hafa gagn af málfræðikennslu í tileinkun erlendra mála (Schmith og Frota 1986). Þetta á sérstaklega við um beygingar. Hér er hins vegar ekki verið að mæla með beinni málfræðikennslu eins og tíðkaðist á árum áður. Í *Icelandic Online* var valin sú leið að vekja athygli á málfræðinni í merkingarlegu samhengi. Mismunandi litir eru notaðir til að vekja athygli á áhersluatriðum í textanum með aðferð sem kölluð hefur verið *Fókus á form* (Doughty og Williams 1998). Upplýsingar um uppbyggingu málsins, merkingu orða og reglur um málnotkun eru settar fram þannig að nemandinn geti áttað sig á þeim þrátt fyrir litla færni í málinu (Birna Arnbjörnsdóttir 2004). Í þessum tilgangi var hannaður sérstakur málfræðigrunnur fyrir byrjendur (Gunnar Þorsteinn Halldórsson 2004). Hanna þurfti grunninn á þann

hátt að upplýsingarnar mætti kalla upp í ákveðnum þyngdarstigum. Nemendur geta þá nálgast málfræðina í þremur lögum eða vinnupöllum (scaffolding) (Peregoy og Boyle 1993) eftir því hversu djúpt nemandinn treystir sé til að kafa í málfræðina. Við hvert stig eða lag dýpkar umfjöllunin: Fyrst er vakin athygli á málfræðiatríðum sem lögð er áhersla á í námsefninu með því að lita yfir viðeigandi orð eða orðasambönd með mismunandi litum (sjá 1. á mynd 1). Þá kemur stutt málfræðileg skýring sem er ein lína (sjá 2. á mynd 1). Með því að smella á línuna kallar nemandinn fram kassa með frekari upplýsingum s.s. beygingarlýsingum (sjá 3. á mynd 1) og þeir allra áhugasömstu geta komist í gagnagrunninn sjálfan (4. á mynd 1). Tölurnar við hliðina á málfræðiskýringunum í kassanum (á mynd 1) vísa til þeirra kafla í málfræðigrunninum þar sem finna má nánari umfjöllun um efnisatriðið.

Í dæminu hér að neðan er verið að kenna atviksorð. Með því að smella á grænan ferhyrning eru atviksorð lituð með grænum lit (1. á mynd 1). Um leið kemur upp setning sem segir: *Atviksorð tákna ýmist dvöl eða stefnu* (2. á mynd 1). Með því að smella á setninguna kemur fram kassi með nánari upplýsingum og notkunardæmum (3. á mynd 1). Takið eftir að aðeins eru útskýrð þau atviksorð sem koma fyrir í textanum. Tölurnar til vinstri við útskýringarnar í kassanum vísa svo til nánari umfjöllunar um atviksorð og notkun þeirra í málfræðigrunninum sjálfum (4. á mynd 1).

**Vinnupallar (scaffolding)**

1. Fókús á málfræðiatríði
2. Skýring
3. Stutt umfjöllun um málfræði
4. Málfræðigrunnur

**málfræði**

1.3.2	2.7
2.2.1	Aðalsorð tákna ýmist dvöl eða stefnu. Í mörgum
2.2.3	síðlum eru tí þó þar sem aðeins orðað
2.2.5	áskilnaði kemur lítið gætt af því hvernig átt.
2.4.2	Alþingi að engin marklagningar er á hér og
2.7	hérna en þarna er aðeins hægt að nota ef
	stærri lítið gætt er átt.
	dvöl emi lítið héra þar þarna
	alsha ann lítið héra þarna
	þingað

Þjómsdóttir  
Báskur 14. 2005

Mynd 1. Vinnupallar (scaffolding)



En það nægði ekki að hanna sérstakan málfræðigrunn eingöngu því engin netorðabók var til sem hentaði til notkunar við kennslu íslensku á netinu. Úr varð að samstarfsaðilar frá háskólanum í Wisconsin keyptu rétt til að nota *Íslensk-enska orðabók* (1989) eftir Sverri Hólmarsson, Chris Sanders og John Tucker. Þessi orðabók var síðan uppfærð til að útskýra þau orð sem koma fyrir í námskeiðinu með ákveðnum takmörkunum þó. Upphaflega stóð til að tengja hvert einasta orð í námsefninu við orðabókina en sú aðferð reyndist ófær þar sem slíkar tengingar þyngja vefsamskiptin mikið en það gerir almennum notendum erfitt um vik með niðurhal. Þá var gripið til þess úrræðis að útbúa glósulista til hliðar við námskeiðið. Í hverjum lista er skýrð merking allra þeirra orða sem koma fyrir í hverjum þræði. Orðalistinn gerði aðstandendum *Icelandic Online* I einnig kleift að aðlaga orðabókargrunninn í Wisconsin á svipaðan hátt og gert var með málfræðigrunninn, þ. e. valin er merking orða í því samhengi sem þau koma fyrir í námsefninu eingöngu en annarri merkingu sleppt. Þessari aðlögun fylgdu vandamál svo sem í hvaða röð hjálparmálin fjögur, sem verða notuð, ættu að koma en þó voru tvö vandamál sérstaklega erfið og tímafrek: Í fyrsta lagi var það sem kallað hefur verið *fossasigsvandamálið* en orðið *fossasig* hefur fengið mun stærra og víðfeðmari merkingu en bara það að síga í fossa. Vandamálið tengist því að reynt var að velja þemu fyrir hvern bálk sem höfða myndu til nemenda og væru spennandi og nýstárleg. Í fyrsta kafla í fyrsta bálki í *Icelandic Online* I fara persónurnar í námskeiðinu í *fossasig*. Þetta orð, *fossasig*, er ekki aðalinntak námsins og því ekki aðalatriði þegar kemur að því að vinna að kennslumarkmiðum fyrsta bálks. En orðið *fossasig* þurfti að skýra og orðabókarmenn voru eðlilega tregir að bæta sjaldgæfum samsettum orðum eins og *fossasig* í netorðabókina. Þetta orð er því notað yfir þetta vandamál almennt í verkefninu þ. e. hvernig eigi að koma merkingu mjög sérhæfðra orða og samsettra orða, klasa og orða úr óformlegu talmáli á framfæri án þess að íþyngja orðabókinni of mikið og án þess að trufla flæði námsins.

Hitt vandamálið var að við tengingu annars orðsins í námskeiðinu við orðabókina, þ. e. tengingu sagnarinnar *að heita* sbr. *Ég heiti*. Þar kom fyrst upp merkingin *Jón heitinn*. Augljóst er að vandamál af þessu tagi greiða ekki fyrir því að nemandinn átti sig á merkingu sagnarinnar *að heita* né virka sem hvatning til frekara náms. Vandamálið verður enn flóknara þegar um er að ræða t.d. algengar forsetningar þar

sem mismunandi merkingu og noktunarmöguleikum er lýst á mörgum blaðsíðum. Þá eru líkur á að upplýsingarnar verði öllum nema þeim allra áhugasömustu ofviða. Til að auðvelda nemendum *Icelandic Online 1* að finna viðeigandi merkingu orða og trufla ekki flæði var hvert einasta orð í námsefninu tengt aðeins þeirri merkingu orðsins sem við átti hverju sinni og öðru sleppt. Hönnun sérstakra orðalista, sem áður er getið, gerði mögulegt að útskýra orð eins og *fossasig* auk ýmissa klasa s.s. *góðan daginn* o.s.frv. sem ekki voru í orðabókargrunninum og nemendur höfðu ekki forsendur til að leita uppi vegna lítillar kunnáttu.

Enn standa þó eftir málfræðiupplýsingar í grunninum sem eiga við allar merkingar orðanna. Til dæmis kemur fram í orðabókinni að sögnin *að heita* geti stýrt nefnifalli og þágufalli þó svo að í umræddri merkingu standi hún með nefnifalli. Ef aðeins sú málfræði sem við á er valin riðlast öll orðabókin. Þetta vandamál er ennþá óleyst en ljóst er að sú vinna sem fer í að tengja orðin í námsefninu við viðeigandi merkingu eingöngu er afar tímafrek en vel þess virði. Nemendur þurfa ekki að eyða miklum tíma í að leita að merkingu orða né í leit að málfræðiupplýsingum sem þeir þurfa til að geta tileinkað sér náms efnið. Aðlögunin virkar því hvetjandi á nemendur og eykur líkur á því að þeir haldi áfram í námskeiðinu. Á mynd 2. má sjá útskýringar á merkingu og notkun forsetningarinnar *á* eins og hún kemur fyrir í textanum (*á Íslandi*):



Mynd 2. Merking forsetningarinnar á aðlöguð málfærni nemandans og námsefninu  
Textinn er tekinn úr bálki 2. í *Icelandic Online 1*

Upprunalega orðabókin og þær skýringar á forsetningunni *á* sem notandi sér þegar hann fer beint inn á orðabókina líta hins vegar svona út:

**2á** *prep / adv*

**A.** *prep (acc)*

1. (*um hreyfingu/of movement*) on, onto, in, into

- hlaupa ~ vegginn** run into the wall
- ganga ~ fjöll** go mountain climbing
- fara ~ bókasafnið** go to the library

2. (*um tíma/of time*) on, in

- ~ mánudaginn** on Monday
- ~ morgun** tomorrow
- ~ morgnana** in the morning(s)
- ~ veturna** in (the) winter
- tvisvar ~ dag** twice a day

**B.** *prep (dat)*

1. (*um stað/of place*) on, in, at

- ~ götunni** on the street
- ~ horninu** at the corner
- ~ Íslandi** in Iceland

2. (*um tengsl/of relations*)

- nafn ~ e-u** name of sth
- eyrun ~ mér** my ears
- gluggarnir ~ húsinu** the windows of the house
- skortur ~ timbri** lack of timber

3. (*um tíma/of time*) in, at

- ~ einni viku** in a week
- ~ jóluum** at Christmas
- hún er ~ niunda árinu** she is eight (going on nine)
- ~ þeim tíma** at that time
- gera e-ð ~ tíu dögum** do sth in the course of ten days

4. *phrases*

- bók ~ ensku** a book in English
- þekkja e-n ~ röddinni** know sby by his voice

**C.** *adv*

- myrkur var ~** darkness had fallen
- honum liggur ~** he's in a hurry
- það finnst ~, það sér ~** it shows, one can see it

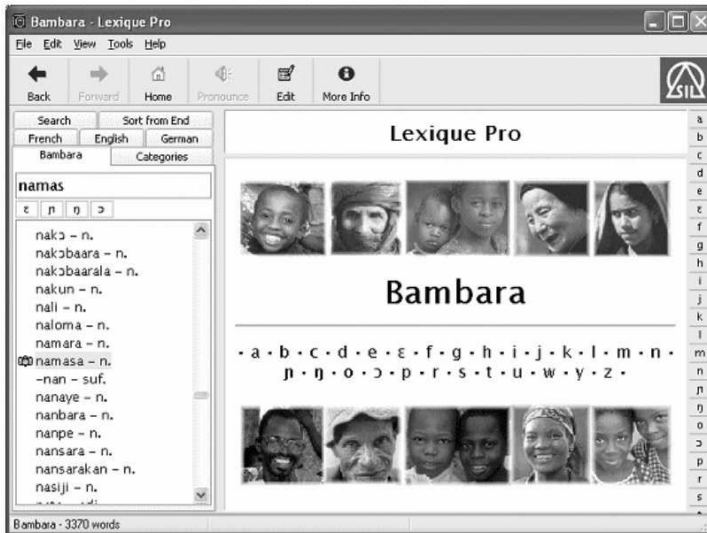
Mynd 3 sýnir hefðbundna orðabókaskýringu á merkingu forsetningarinnar *á*

Ljóst er að fyrra dæmið er mun skýrara og hentar betur þeim sem eru að hefja nám í íslensku.

## 6 Notkun hjálparmiðla á netinu

Ef til vill mætti segja að vegna þess hversu greiðan aðgang nemendur hafa að alls kyns orðabókum, alfræðisöfnum og málfræðilýsingum á netinu sé óþarfi að vera að leggja vinnu og fjármagn í hönnun upp-sláttarrita sem haldið er úti af sérfræðingum. Kennarar og nemendur sækja í síauknum mæli ókeypis forrit og fá aðgang að gjaldfrjálsum gagnabönkum á vefnum. Sem dæmi um frítt forrit til orðabókagerðar má nefna *Lexiquepro* þar sem setja má saman myndir, vídeó og hljóð-skrár.

## Features - Introduction



Mynd 4. sýnir dæmi um forrit sem sækja má á vefnum  
<http://www.lexiquepro.com/features.htm>

Gagnagrunnar eins og [www.wikipedia.com](http://www.wikipedia.com) verða sívinsælli sem og notkun *Google* og annarra leitarvéla eins og áður er lýst. *Wikipedia* er opið alfræðisafn þar sem hver sem er getur sett inn upplýsingar. En við nánari athugun t.d. á hvað þar stendur um íslenska beygingar- og setningafræði kemur í ljós að hér er alls ekki um lýsingu á íslenskri málfræði að ræða heldur afar stutta kynningu á almennri beygingarfræði.

Many [German](#) speakers will find Icelandic morphology familiar. Almost every morphological category in one language is represented in the other. [Nouns](#) are declined for [case](#), [number](#) and [gender](#), [adjectives](#) for case, number, gender and [comparison](#), and there are two [declensions](#) for adjectives, weak and strong. Icelandic possesses only the [definite article](#), which can stand on its own, or be attached to its modified noun (as in other [North-Germanic](#) languages). [Verbs](#) are [conjugated](#) for [tense](#), [mood](#), [person](#), number and [voice](#). There are three voices, active, passive and medial, but it may be debated, whether the medial voice is a voice or simply an independent class of verbs of its own. There are only two simple tenses, past and present, but to make up for that there are a number of auxiliary constructions, some of which may be regarded as tenses, other as [aspects](#) to varying degrees.

### Syntax

Icelandic is [SVO](#), generally speaking, but the inflectional system allows for quite some freedom in word order.

Mynd 5. Lýsing á íslenskri beygingarfræði á wikipedia  
[http://en.wikipedia.org/wiki/Icelandic\\_language#Morphology](http://en.wikipedia.org/wiki/Icelandic_language#Morphology)

Að sjálfsögðu ber að fagna því að hægt er að nálgast upplýsingar um nánast hvað sem er á vefnum en hins vegar vakna spurningar um gæði slíkra gagnabanka bæði varðandi innihald og svo viðhald. Verkefni á borð við NORDIT (2002) o. fl. sýndu fram á að til er ógrynni vefsíðna sem einhverjir velviljaðir hafa sett upp en síðan ekki haldið við og uppfært sjaldan. Glæsileg forrit tryggja ekki gæði innihaldsins.

Þegar kemur að notkun orðabóka og gagnabanka til tungumálanáms er ljóst að upplýsingaleitin krefst einhverrar tungumálakunnáttu og ekki síður kunnáttu í að leita, meta og velja upplýsingar á netinu — þ. e. tölvulæsi. Oft er talað um að ungir nemendur séu ákaflega færir í að nota tölvur til ýmissa hluta en það þýðir ekki að þeir séu færir um að meta réttmæti og nýta sér þær upplýsingar sem þeir hafa við höndina í tölvunum og á netinu sér til gagns (Sólveig Jakobsdóttir 2005). Þá má draga í efa hvort það að þjálfra málfærni sé sama ferli og það að byggja upp þekkingu t.d. í sögu eða landafræði eins og sambandshyggjumenn halda fram. Ef svo væri, væri ekki til sá greinarmunur sem allir eru sammála um að sé á því að kenna t.d. sögu eða landafræði á netinu og svo aftur að þjálfra nemendur í erlendum tungumálum þar sem þeir þurfa að afla sér þekkingar á því máli sem þeir eru að læra.

Það þarf líka að auka rannsóknir á hegðun þeirra sem nota netið til að læra tungumál og þá sérstaklega hvernig nemendur nota hjálparmiðla en rannsóknir á því hvort og þá hvernig nemendur nota orðabækur og málfræðigrunna eru afar fáar. Þær rannsóknir sem til eru miðast við háskólanema að læra þýsku eða finnsku og benda til þess að nemendur þurfi sérstaka hvatningu til að sækja sér upplýsingar, annars eru þeir ólíklegir til þess. Eftirgjöf og skýringar nýtast best ef þær fjalla um eitt atriði í einu og skýringar mega helst ekki vera nema 3 línur að lengd (Heift 2004, 2005; Chappelle 2003:4 kafli).

Þá hefur komið fram að notendur verða sífellt tregari til að greiða fyrir bjargir á netinu. Til dæmis hefur sala á tungumálanámskeiðum á CD tölvudiskum dregist saman um helming í Bretlandi undanfarin tvö ár (Davies 2005). Áður en ráðist er í viðamikla þróunarvinnu á gagnagrunnum á netinu til tungumálakennslu þurfum við því að þekkja betur hegðunarmynstur þeirra nemenda sem líklegir eru til að nýta sér síaukið upplýsingaflæði á vefnum. Gagnagrunnar verða að vera aðgengilegir sem flestum og í tungumálanámi einnig þeim sem ekki hafa mikla færni í tungumálinu.

## 7 Lokaorð

Í framhaldi af reynslunni af þróun hjálparmiðla við *Icelandic Online* er óhjákvæmilegt að ýmsar vangaveltur komi upp um hvernig byggja megi upp alhliða gagnabanka um tungumál eins og íslensku sem fáir tala og því mikilvægt að huga að skilvirkni og kostnaði slíkra verkfæna. Æskilegast væri að slíkur gagnabanki um íslenskt mál geti þjónað sem flestum þannig að ekki þurfi að leggja í kostnað við marga dýra gagnabanka. Slíkur gagnabanki innihéldi upplýsingar um íslenskt mál og væri þannig uppbyggður að úr honum mætti sækja upplýsingar við hæfi notandans hvort sem hann er fræðimaður í málfræðirannsóknum eða byrjandi í íslensku sem erlendu máli. Hugsanlega mætti byggja gagnabankann upp eftir sömu leiðum og hjálparmiðla *Icelandic Online* þ. e. í ákveðum lögum eftir magni og nákvæmni upplýsinganna. Efst sé þá einföld lýsing við hæfi þeirra sem eru að stíga fyrstu skrefin í íslenskunáminu ásamt myndum til að auðvelda skilning auk algengra málnotkunardæma. Eitt lag gæti verið byggt á beygingarlýsingum Orðabókar Háskólans <http://www.lexis.hi.is/beygingarlýsing>, með sérstakri áherslu á beygingar algengra orða. Annað lag gæti þjónað sama tilgangi og handbækur um ritun og almenna málnotkun fyrir alla notendur. En að baki liggja nákvæmur, alhliða gagnabanki með nákvæmri lýsingu á íslenskri máluppbyggingu í samræmi við nýjustu rannsóknir í málvísindum.

## Heimildaskrá

- Birdsong, D. 1999. *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates
- Birna Arnbjörnsdóttir. 2003. *Beygingamál og helstu kenningar um tileinkun erlendra mála: Tillaga að rannsóknaráætlun*. Erindi flutt á Hugvísindapingi Háskóla Íslands í Reykjavík 1. nóvember 2003.
- Birna Arnbjörnsdóttir. 2004. Teaching Morphologically Complex Languages Online: Theoretical Questions and Practical Answers. In Peter Juel Henriksen (ritstj.) *CALL for the Nordic Languages*. Copenhagen Studies in Language. 30. Copenhagen: Copenhagen Business School.
- Bley-Vroman, R. 1989. What is the logical problem of foreign language learning? Í: S. Gass & J. Schachter (ritstj.), *Linguistic perspectives on*

- second language acquisition*, bls. 41–68. Cambridge: Cambridge University Press.
- Bruner, J. 1974. *Going Beyond the Information Given*. New York: Norton.
- Chapelle, C. 2003. *English Language Learning and Technology*. Philadelphia: John Benjamins Publishing.
- Clahsen, H., I. Sonnenstuhl, og J. P. Blevins. 2003. Derivational morphology in the German mental lexicon: a dual-mechanism account. Í: H. Baayen & R. Schreuder (eds.), *Morphological Structure in Language Processing*. New York: Mouton de Gruyter.
- Davis, Graham. 2005. Í tölvuskeyti til höfundar 8. mars 2005.
- Doughty C. og J. Williams. 1998. *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- Ellis, N. C. 2003. Constructions, Chunking and Connectionism: The Emergence of Language Structure. Í: Doughty and Long (ritstj.) *The Handbook of Second Language Acquisition*, bls. 63–103. Oxford: Blackwell.
- Gunnar Þorsteinn Halldórsson. 2004. *Málfræðigrunnur með Icelandic Online*. Óprentað handrit.
- Heift, T. 2004. Corrective Feedback and Learner Uptake in CALL. *ReCALL Journal*, Cambridge University Press, Volume 16:416–431.
- Heift, T. 2005. Inspectable Learner Reports for Web-based Language Learning. *ReCALL Journal*, Cambridge University Press, 17:32–46.
- Hubbard, Phillip. 2004. *Using Google as Tool for Writing Instruction*. Erindi flutt á ráðstefnu EUROCALL í Vínarborg 6. september 2004. *Icelandic Online I*: <http://www.icelandic.hi.is>
- Íslensk-ensk orðabók. 1989. Sverrir Hólmarsson, Chris Sanders og John Tucker (ritstj.). Reykjavík: Iðunn.
- Jackendoff, R. 1994. *Patterns in the mind*. New York: Basic Books.
- Kern, R., P. Ware og M. Warschauer. 2004. Crossing frontiers: New directions in online pedagogy and research. *Annual Review of Applied Linguistics* 24: 243–260.
- Kolbrún Friðriksdóttir. 2004. *Íslenska sem annað mál: Hvernig lærist fallbeyging nafnorða í íslensku?* Óprentuð M.A.-ritgerð í íslenskri málfræði. Háskóli Íslands.
- Krashen, Stephen. 1989. *Principles and Practice in Second Language Acquisition*. Oxford: Pergammon Press.
- Lantolf, J. 2000. *Sociocultural Theory and Second Language Learning*. New York: Oxford University Press.

- Lardiere D. 1998. Case and tense in the 'fossilized' steady state. *Second Language Research*, 14:1–26.
- NORDIT : *Kvalitetsutveckling av IT läromaterial för nordiska språk*. 2002. Nordiska Ministerrådet.
- Norris, J. og L. Ortega, 2000. Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50:417–528.
- Peregoy S. og O. Boyle. 1993. *Reading, Writing and Learning in ESL: A Resource Book for K-12 Teachers* (2. útgáfa). New York: Longman.
- Pinker, S. og A. Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28:73–193.
- Pinker, S. og A. Prince. 1994. Regular and irregular morphology and the psychological status of rules of grammar. Í: Lima, S.; Corrigan, R. & Iverson, G. (ritstj.) *The Reality of Linguistic Rule*, bls. 321–351. Amsterdam: John Benjamins.
- Rumelhart, D. E., og J. L. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.
- Schmidt, R., og S. Frota. 1986. Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. Í: R. Day (ritstj.), *Talking to Learn: Conversation in Second Language Acquisition*, bls. 237–326. Rowley, MA: Newbury House.
- Sólveig Jakobsdóttir. 2005. <http://soljak.khi.is/netnot>. Sótt 15. ágúst 2005.
- Sverrir Hólmarsson, Chris Sanders og John Tucker. Íslensk-ensk orðabók. <http://digicoll.library.wisc.edu/IcelOnline/Search.TEId.html>.
- Warschauer, M. 1998. *Interaction, Negotiation and Computer-Mediated Learning*: [www.insa-lyon.fr/Departements/CDRL/interaction.html](http://www.insa-lyon.fr/Departements/CDRL/interaction.html). Sótt 20. ágúst 2005.
- Warschauer, M. og R. Kern (Ritstj.). 2000. *Network-based Language Teaching: Concepts and Practice*. Cambridge: Cambridge University Press.

## Abstract

The reflections presented in this article grew out of the author's experience in the development of an online grammar resource and glossary and their adaptation to the proficiency levels of the students and the content of *Icelandic Online*. *Icelandic Online* is an online tutorial for



learning Icelandic as a foreign language. This article describes how the online grammar source and dictionary entries were scaffolded or layered so that they could benefit the beginning learner. These resources play an especially important role in the learning of inflected languages where words can have many different forms and there is a focus on form in the early stages of learning. The distinction between grammars and dictionaries as encyclopedias and as handbooks for language learning is highlighted and the importance of this distinction in current theories of adult language learning and pedagogy is discussed. Finally the author reflects on how scaffolding could benefit small languages like Icelandic as information contained in a single online linguistic resource could be layered in a similar fashion as the resources attached to *Icelandic Online* and thus be used as a handbook for practical purposes but also as a corpus for conducting research.

**Keywords:**

Computer Assisted Language Learning, Scaffolding Auxiliary Resources in Online Language Learning

**Lykilorð:**

Tungumálakennsla með tölvum, aðlögun hjálparmiðla að málfærni nemenda

*Birna Arnbjörnsdóttir*  
*dósent, enskuskor*  
*Hugvísindadeild*  
*Háskóla Íslands*  
*birnaarn@hi.is*



Kristín Bjarnadóttir

# Málfræði í orðabókum

## 1 Inngangur

Í orðabókum er að finna lýsingu einstakra orða eða annarra þeirra einda sem mynda uppfletteiningar í verkinu, hvort heldur það er hefðbundin prentuð orðabók eða rafræn orðabók af einhverju tagi. Þær upplýsingar sem orðabókarlýsing byggist á og kemur til skila eru af ýmsu tagi en efni þessarar greinar er það hvernig málfræðiupplýsingar eru settar fram fyrir orðabókarnotendur, bæði að formi og efni.

## 2 Málfræði, orðasafnið og orðabækur

Málfræði er lýsing á málkerfinu, sett fram á ýmsan hátt eftir tíma og skólum. Viðfangsefni málfræðinga er að setja fram reglur og lögmál um það hvernig kerfið í heild eða einstakir hlutar þess virka; það má því segja að málfræðingar fáist við málið „ofan frá og niður“. Orðabækur lýsa kerfinu hins vegar „neðan frá og upp“ ef svo má segja, þ.e. út frá einstökum orðum eða flettum.

Bæði fyrirbærin, málfræði og orðabækur, eru flókin en hvort á sinn hátt. Orðabækur byggjast á málfræðilegri hugsun en orðabókarfræði er aðeins að hluta til málfræðileg. Miðlun efnisins sjálfs og framsetning þess er þar fullgildur þáttur. Þessi munur sést af þeirri flokkun sem gjarnan er viðhöfð þegar rætt er um þessar greinar.

Málfræði skiptist í fjölmargar undirgreinar, sbr. skilgreiningu Höskuldar Þráinssonar í *Handbók um málfræði* (1995:89):

MÁLFRÆÐI er samheiti yfir allar þær fræðigreinar sem lýsa eðli mannlegs máls og skýra það, svo sem hljóðfræði, beygingarfræði, orðmyndunarfræði, setningafræði, merkningarfræði.

Þarna má bæta við fjölmörgum greinum og reyndar má telja orðabókarfræði til málfræði líka.

Orðabækur eru af ýmsu tagi, en skilgreiningu á því fyrirbæri sem íslenskum orðabókarnotendum er tamast að hugsa til þegar minnst er á orðabók er að finna í dæmigerðustu íslensku orðabókinni:

**orðabók** KVK 1 bók með safni orða (í stafrófsröð) úr e-u máli og (venjulega) skýringum á merkingu (notkun) þeirra á sama eða öðru máli > *íslensk orðabók / færeysk-íslensk orðabók / Orðabók Sigfúsar Blöndals / stafsetningarorðabók*. 2 stofnun þar sem unnið er að samningu slíkrar bókar > *Orðabók Háskóla Íslands*

(*Íslensk orðabók* 2002)

Tæknibreytingar í orðabókargerð eru slíkar að orðabók er ekki lengur endilega bók, stafrófið er ekki eini hugsanlegi leitaraðgangurinn að efninu og merkingarlýsing er aðeins einn af þeim þáttum sem orðabókarlýsing nær til. Orðabækur eru af ýmsum gerðum og nægir þar að nefna almennar móðurmálsbækur (eins og *Íslensk orðabók* sem vísað er til hér á undan), tvímálaorðabækur, stafsetningarorðabækur, orðsifjabækur, samheitabækur, orðtengslabækur, þesárusa, málnotkunarbækur, iðorðabækur o.s.frv. Allar þessar orðabækur eiga það þó sameiginlegt að útgangspunkturinn er orðið sjálft eða flettan.

Í generatífri málfræði er það orðasafnið (orðasafn hugans) sem geymir allar upplýsingar um einstök orð og önnur fyrirbæri sem þar eru sambærileg við orðabókarflettur. Þar er kominn tengiliður málfræði sem er lýsing á málkerfinu og orðabókar sem er upplýsingaveita um einstakar flettur. Í stuttu máli má segja að í orðasafninu séu les ásamt öllu öðru sem ekki er fyrirsegjanlegt samkvæmt þeim reglum og lögmálum sem sett eru fram um málkerfið í málfræði, með öllum sínum undirgreinum. Í ljósi þessa mætti ætla að í góðri orðabók eigi að vera úrval lesa með nauðsynlegum upplýsingum um þá málfræði sem ekki er fyrirsegjanleg miðað við málfræðikunnáttu notandans, þ.e. sú málfræði sem hann er ekki sjálfbjarga með. Framsetning þessara upplýsinga hlýtur að mótast af gerð orðabókarinnar og notandahópnum

sem hún er ætluð, auk miðilsins sem hún birtist í, þar sem möguleikarnir í prentaðri bók og rafrænni (til dæmis) eru mjög mismunandi. Hér á eftir verður sjónum aðallega beint að almennri íslenskri orðabók, enda er nokkur hefð á framsetningu málfræðiupplýsinga þar.

### 3 Hvaða málfræði á að vera í orðabók?

Með málfræðiupplýsingum í orðabók er átt við kerfisbundna framsetningu tiltekinna málfræðilegra þátta sem á sér fastan stað í flettugreinunum. Gera þarf greinarmun á málfræði sem notuð er við skipulag á flettugreinum í orðabók og málfræði sem beinlínis er birt notandanum sem slík, skýrt afmörkuð innan flettugreinanna. Þá þarf einnig að gera greinarmun á upplýsingum sem leynast í flettugreinunum, t.d. í dæmum, útskýringum, og jafnvel í leiðbeiningum um málnotkun, og skipulega framsettum málfræðiatríðum. Skipulag flettugreina í orðabókum Jóns Hilmars Jónssonar (1994, 2002) ákvarðast t.d. algjörlega af málfræðilegum þáttum, af setningargerð í *Orðastað* og af merkingarlegum þáttum í *Orðaheimi*.<sup>1</sup>

Þegar aðrar íslenskar orðabækur eru skoðaðar kemur í ljós að málfræðiupplýsingar eru oftast fólgnar í orðflokksmerkingu flettna, ásamt vísbendingum um beygingu orða sem yfirleitt eru settar fram með því að gefa endingar kennifalla nafnorða eða kennimynda sagna, og heilar beygingarmyndir þar sem stofnbrigði krefjast þess. Þessi háttur er t.d. hafður á í *Íslenskri orðabók* (2002) þar sem beygingar grunnorða eru sýndar á þennan hátt:<sup>2</sup>

**hundur** KK -s, -ar  
**labba** S -aði, -að  
**fjörður** KK fjarðar, firðir

<sup>1</sup>Í *Stóru orðabókinni* (2005) hafa bæði þessi verk verið felld saman í eina heild. Í þessum verkum brýtur Jón Hilmar blað í íslenskri orðabókargerð; hefðin er sú að málfræðiupplýsingar séu fremur af skornum skammti og ekki endilega settar fram á kerfisbundinn hátt, nema þá helst upplýsingar um beygingu.

<sup>2</sup>Kenniföll og kennimyndir eru þær beygingarmyndir nafnorða og sagna sem taldar eru nægja til að gefa vísbendingu um beygingarflokk. Kenniföll nafnorða eru nefnifall og eignarfall eintölu og nefnifall fleirtölu, t.d. *köttur, kattar, kettir; nál, nálar, nálar*. Kennimyndir sterkra sagna eru fjórar: nafnháttur, þátíð framsöguháttar í 1. persónu eintölu og fleirtölu, og lýsingarháttur þátíðar, t.d. *ganga, gekk, göngum, gengið*. Kennimyndir veikra sagna eru þrjár: nafnháttur, þátíð framsöguháttar í 1. persónu eintölu og lýsingarháttur þátíðar, t.d. *æpa, æpti, æpt*.

Málfræðiupplýsingar af öðrum toga eru yfirleitt ekki settar fram með kerfisbundnum hætti í íslenskum orðabókum, þótt dálitlar upplýsingar um setningargerð sé einnig að finna í *Íslenskri orðabók*. Ekki er samt gerð tilraun til að sýna fallstjórn sagna og setja rökliði þeirra fram á kerfisbundinn hátt fyrr en í 3. útgáfu.<sup>3</sup>

Minna fer fyrir annars konar málfræðilegum upplýsingum í íslenskum orðabókum. Hljóðkerfisfræði er afskipt grein í íslenskri orðabókargerð og kerfisbundnar upplýsingar um framburð er yfirleitt ekki að finna, nema í Blöndalsorðabók þar sem hljóðritun fylgir uppfletti-orðum. Ekki er gerð grein fyrir orðmyndun á skipulagðan hátt nema í *Íslenskri orðsifjabók* (Ásgeir Blöndal Magnússon 1989), þótt stöku aðskeyti séu uppflettiorð í *Íslenskri orðabók*. Loks eru merkingarlegir þættir ekki settir fram á skipulagðan hátt þó að merkingarskýringar séu uppistaðan í verkum á borð við *Íslenska orðabók*; reyndar er þessi þáttur stundum svo allsráðandi að svo virðist sem gert sé ráð fyrir að notandinn viti bókstaflega allt um orðið sem hann leitar að, nema hvað það merkir. Oft er engar aðrar upplýsingar um uppflettiorðið að finna en merkingarskýringuna eina og notandinn er stundum engu nær um það hvernig orðið er notað.

Notendahópur íslenskra orðabóka er lítill og þess vegna þarf almenn íslensk móðurmálsorðabók að gegna mörgum hlutverkum í einu. Í hana leita allir, bæði skólafólk, innlent og erlent, almenningur og fræðimenn. Hún þarf því helst að vera nothæf bæði til skilnings og skrifta, þ.e. sem aktív og passív orðabók í senn. Lágmarkskröfur til almennrar íslenskrar orðabókar eru því talsvert miklar:

Til að orðabók komi að fullum notum þarf hún að innihalda nægilegar upplýsingar til að notandanum sé kleift að nota flettiorðin á réttan hátt. Því verður að spyrja: Hvaða þættir þurfa að koma fram í orðabókarflettu; og hvernig á að setja þá fram?

(Eiríkur Rögnvaldsson 1998:25)

Grein Eiríks sem þarna er vitnað til er byggð á erindi sem flutt var á málþingi um almenna íslenska orðabók sem haldið var í tilefni af því að endurskoðun á *Íslenskri orðabók* vegna 3. útgáfu var á mótunarstigi. Í greininni setur Eiríkur fram hugmyndir um það hvernig bæta

<sup>3</sup>Með *röklið* er hér átt við skyldubundna nafnliði sem fylgja sögn, t.d. í stöðu frumlags, andlags eða í forsetningarlið.

má ýmsa ágalla í framsetningu málfræðiupplýsinga í bókinni. Hann bendir á að beygingarlýsinguna í bókinni megi laga með því að samræma efnið betur og bæta þurfi inn bráðnauðsynlegum upplýsingum um grunnatriði í sambandi við setningarstöðu orða. Ábendingar Eiríks voru hafðar að leiðarljósi við endurskoðun *Íslenskrar orðabókar*, að svo miklu leyti sem unnt var vegna tímaskorts.<sup>4</sup>

Þótt stefnt sé að því takmarki að setja beyginga- og setningagerðarupplýsingar fram á kerfisbundinn hátt er ekki sjálfgefið hvaða leið á að fara. Þar þarf að athuga hvað gera má ráð fyrir að notandinn kunni og hvað komi honum að mestu gagni.

## 4 Málfræði í orðabókum: Þrjár höfuðleiðir

Í handbók um orðabókarfræði (Hausmann o.fl. 1989:518) setur Joachim Mugdan fram þrjár höfuðleiðir í framsetningu beygingarupplýsinga í almennum móðurmálsorðabókum. Hann tekur fram að valið milli þessara aðferða markist af málkunnáttu notendahópsins en orðabókarmaðurinn velur sér einnig leið eftir hugmyndum sínum um sambandið á milli málfræðibóka og orðabóka og verkaskiptingu þar á milli. Hér verða þessar hugmyndir Mugdans skoðaðar í ljósi íslenskra gagna. Fyrst eru beygingaratriði skoðuð en síðan er vikið að upplýsingum um setningargerð en áhugavert er að skoða hvort hægt er að beita sömu aðferðafræði þar.

### 4.1 Beygingarlýsing í orðabókum

Leiðir Mugdans eru miðaðar við framsetningu beygingarupplýsinga í almennri móðurmálsorðabók þar sem hægt er að gera ráð fyrir tiltölulega einsleitum hópi notenda. Leiðirnar þrjár eru þessar:

1. Beygingarlýsing á heima í málfræðibókum.
2. Lýsa þarf óreglulegri beygingu og undantekningum. Gert er ráð fyrir að notandinn kunni reglulegar beygingar eða geti fundið þær í málfræðibókum.
3. Orðabókin ein á að duga til að finna beygingu orða.

<sup>4</sup>Sjá formála 3. útgáfu *Íslenskrar orðabókar* og grein mína um endurskoðun á sagnlýsingunni í verkinu (Kristín Bjarnadóttir 2001).

Notendahópur *Íslenskrar orðabókar* er ekki af sama tagi og hópurinn sem Mugdan miðar við; bókin þarf að þjóna mun breiðari hópi, eins og fram kemur hér að ofan, enda er hún eina íslenska móðurmálsorðabókin sem gefin hefur verið út.

Hér á eftir eru þessar leiðir skoðaðar í ljósi íslenskra gagna. Tekið skal fram að grundvallaratriðin sem talin eru hér að ofan eru ættuð frá Mugdan; eftirfarandi staðfærsla er hins vegar mín eigin.

#### 4.1.1 Beygingarlýsing á heima í málfræðibókum

Jafnvel þótt henta þætti að hafa klára verkaskiptingu milli málfræðibóka og orðabóka þannig að orðabókarnotandinn þyrfti að sækja sér upplýsingar um beygingar í málfræðibækur væri ekki hægt að komast hjá því að merkja flettur þannig að notandinn kæmist á leiðarenda og fyndi orðinu stað í réttum beygingarflokki. Til þess að tryggt sé að uppflertimynd sé einkvæm þarf upplýsingar um orðflokk og kyn nafnorða, t.d. í orðunum *reiði* (hk og kvk) og *leiði* (hk og kvk). Þetta dugar ekki einu sinni í öllum tilfellum þar sem dálítið er til af orðum þar sem beygingarflokkarnir sjálfir skilja á milli orða, t.d. í samhljóma veikum og sterkum sögnum (*brenna*), og í nafnorðum eins og *ró* þar sem beyging er mismunandi eftir merkingu og kenniföllin ein duga til að skilja á milli orða. Í merkingunni 'kyrrð, friður' er *ró* eintöluorð og eignarfall eintölu er *róar/rór* en *ró* í merkingunni 'stálplata . . . með gati' er líka til í fleirtölu og kenniföllin eru *róar* (ef.et.) og *rær* (nf.ft).

Lágmarksupplýsingar um beygingar í íslensku til þess að hægt sé að **finna** upplýsingarnar í málfræðibókum eru því samkvæmt þessu annað tveggja:

1. Kennimyndir og kenniföll (ásamt orðflokki og kyni), a.m.k. í einsrituðum flettiorðum.
2. Tákn (fyrir beygingarflokka) sem vísa í málfræðibók.

Íslenska hefðin er að sýna kenniföll og kennimyndir, eins og gert er í *Íslenskri orðabók*, en dæmi eru um síðari aðferðina líka. Í *Rússnesk-íslenskri orðabók* eftir Helga Haraldsson (1996) er t.d. notað flókið tákna-kerfi, sem vísar í ítarlega beygingarlýsingu aftast í bókinni. Framsetningin er að því leyti óvenjuleg að beygingarupplýsingar fylgja báðum málunum, bæði rússnesku og íslensku.



Báðar þessar leiðir hafa galla. Táknakerfið gerir talsverðar kröfur til notandans vegna þess hve ógagnsætt það er og ósennilegt er að notendum verði það verulega tamt nema með mjög mikilli og stöðugri notkun. Táknakerfið vísar samt nákvæmlega á réttan stað, ef notandinn hefur í sér einurð til að fylgja því eftir og málfræðin sjálf er sett fram á nægilega ítarlegan máta. Kenniföll og kennimyndir gefa í verunni heldur meiri upplýsingar en ætla mætti ef allar málfræðiupplýsingar eiga heima í málfræðibók en án þessara upplýsinga er stundum ekki hægt að greina samhljóða orð í sundur í les. Samt duga kenniföll og kennimyndir nafnorða og sagna ekki alltaf til þess að hægt sé að ganga að allri beygingu uppflettiorðs vísri í málfræðibókum. Kenniföll og kennimyndir eru ekki einu beygingarmyndir orða sem eru ófyrirsegjanlegar út frá stofngerð og í málfræðibókum er yfirleitt aðeins gefið yfirlit um beygingarflokka, með mismörgum dæmum. Tæmandi er lýsingin aldrei og notandinn er því stundum engu nær um þá beygingarmynd sem hann vill fá. Nægir þar að nefna þágufall eintölu í sterkum karlkynsnafnorðum (með eða án *-i*) og *-na* eða *-a* í eignarfalli fleirtölu af veikum kvenkynsnafnorðum. Loks má geta þess að lítið stoðar að fara í málfræðibók til að finna ítarlega beygingarlýsingu íslenskra lýsingarorða; hana er hvergi að finna á bók.<sup>5</sup>

#### 4.1.2 Ófyrirsegjanleg beyging birt í orðabókinni

Ein leið til að nýta málkunnáttu notandans til að ákvarða hvaða beygingarupplýsingar eiga heima í orðabók er að gera ráð fyrir því að stærstu beygingarflokkar séu sjálfgildi í lýsingunni og upplýsingar um þá þurfi ekki að birta. Stærstu flokkarnir eru jafnframt opnir og ómarkaðir beygingarflokkar, sbr. greiningu Ástu Svavarsdóttur í bókinni *Beygingakerfi nafnorða í nútímáslensku* (1993), t.d. veikar sagnir sem fá endinguna *-aði* í þátíð, veik beyging karlkyns- og kvenkynsnafnorða o.s.frv. Útkoman verður þá þessi:

<sup>5</sup>Samkvæmt reynslunni við vinnu við *Beygingarlýsingu íslensks nútímamáls* (sjá vefsíðu Orðabókar Háskólans: [www.lexis.hi.is](http://www.lexis.hi.is)) þarf fimm beygingarmyndir til þess að skipta lýsingarorðum upp í beygingarflokka svo að ótvírætt sé. Þessar orðmyndir ættu því að geta gegnt sama hlutverki og kenniföll nafnorða og kennimyndir sagna. Orðmyndirnar eru nefnifall eintölu í öllum kynjum í sterkri beygingu, ásamt miðstigi og efastigi (kk.nf.et.sb.), t.d. *góður, góð, gott, betri, bestur; heitur, heit, heitt, heitari, heitastur; stuttur, stutt, stutt, styttri, stýstur; háður, háð, háð, háðari, háðastur*. Þessar upplýsingar hafa ekki verið settar fram á skipulegan hátt í orðabók og hér á eftir verða nafnorð og sagnir látin duga sem dæmi um beygingarupplýsingar í orðabókum.

1. Regluleg beyging er ekki birt:  
Beygingin er fyrirsegjanleg út frá orðflokki, kyni og stofngerð, t.d.

**hest | ur** KK, **penn | i** KK, **bjór |** KK, **mó | r** KK  
**rústa** S

2. Allar aðrar beygingar eru birtar í orðabókinni (á einhvern hátt), t.d.

**gest | ur (-s, -ir)** KK  
**fjörð | ur (fjarðar, firðir, þgftet. firði)** KK  
**keyra (keyrði keyrt)** S

Þarna er notendum ætluð talsverð þekking á beygingarkerfinu og ekki er víst að allir notendur geri sér ljóst hvaða merkingu það hefur þegar engar beygingarupplýsingar fylgja orði. Íslenskir notendur orðabókar þar sem þessi leið er farin eru reyndar ekki mjög líklegir til að koma að tómunum kofunum; það er fremur ósennilegt að þeir þurfi að fletta upp kenniföllum eða kennimyndum orða úr opnum (reglulegum) beygingarflokkum. Ófyrirsegjanlegar myndir þyrfti samt alltaf að birta, þ.m.t. þágufall eintölu af sterkum karlkynsorðum og nefnifall fleirtölu af veikum kvenkynsorðum. Þar virðast engar reglur duga til að ákvarða hver beygingin er.<sup>6</sup>

Þessi leið byggir m.ö.o. á því að notendur viti hvaða merkingu það hefur ef enginn beygingarlýsing fylgir orði. Málfræðiupplýsingarnar sem fylgja þessum orðum eru þá aðeins orðflokkur, kyn nafnorða og tákun á stofngerð, t.d. með lóðréttu striki á undan nefnifallsendingu í karlkynsnafnorðum.

#### 4.1.3 Orðabókin á að duga ...

Þriðja leið Mugdans í framsetningu beygingarupplýsinga í orðabókum er að gera kröfu um að orðabókin ein dugi til þess að notandinn fái fullnægjandi upplýsingar um beygingu orðs, án þess að gert sé ráð fyrir mikilli málkunnáttu. Í íslensku væri hægt að gera þetta á tvo vegu:

<sup>6</sup>Þetta kom skýrt í ljós við vinnu við *Beygingarlýsingu íslensks nútímamáls* á Orðabók Háskólans. Þar eru settar fram tilgátur um þessar beygingarmyndir orða, byggðar á talsverðum rannsóknnum, en rækileg úttekt á þessu er ekki til.

1. Með því að birta kenniföll og kennimyndir af öllum orðum, auk óreglulegra mynda
2. Með því að birta heil beygingardæmi . . .

Fyrri liðurinn hér er mjög svipaður síðari lið í næsta kafla hér á undan, þ.e. að birta allar kennimyndir og kenniföll og annað efni sem þarf til að gefa nægjanlegar upplýsingar um ófyrirsegjanlegar myndir. Munurinn er þó sá að hér eru birtar upplýsingar um alla beygingarflokka og ekki treyst á málkunnáttu notandans. Hin leiðin er að birta beygingardæmi í heild en sú leið hefur til skamms tíma ekki verið fær í orðabókum vegna plássleysis.

#### 4.1.4 . . . en fyrir hvern á orðabókin að duga?

Valið milli þeirra kosta sem hér hafa verið settir fram er e.t.v. ekki lengur jafnmikið lykiltríði í orðabókargerð og verið hefur. Orðabókarhefðin hefur að verulegu leyti mótast af þeirri nauðsyn að spara pláss eftir fremsta megni, stundum svo að það verður til verulegra vandræða fyrir notendur. Notkun á táknakerfi í framsetningu á beygingarupplýsingum sparar meira pláss en nokkur önnur leið, en hefur líka galla eins og Helgi Haraldsson bendir á í formála *Rússnesk-íslenskrar orðabókar* sem nefnd er sem dæmi um verk þar sem beygingarupplýsingar eru gefnar sem tilvísanir í sjálfstæða málfræði:

Margvíslegum brögðum er beitt til að spara rými, og kann þetta í fyrstunni að vefjast fyrir sumum.

(Helgi Haraldsson 2002:ii)

Víst er að það reynir á þolinmæðina að læra á táknakerfi eins og það sem Helgi notar, en hann hefur það til síns máls að rússneskumælandi fólk mun vera vant að nota táknakerfi af þessu tagi í orðabókum og vissulega sparar hann með þessu umtalsvert rými í bókinni sem er mjög stór samt.

Jafnvægislistin sem þarf til þess að koma öllum nauðsynlegum upplýsingum um beygingu orða fyrir í orðabók án þess að umfangið fari úr böndunum er mikil kúnst, sérstaklega þegar gera þarf ráð fyrir jafnfjölbreyttum notendahópi og við útgáfu almennrar íslenskrar orðabókar. Í rafrænni orðabók, hvort heldur er á diskum eða á vefnum, eru möguleikarnir allt aðrir. Þar er jafnvel hægt að koma efninu

Þannig fyrir að notandinn sjálfur geti ákveðið hvers konar upplýsingar henta honum best og sótt sér þær að vild. Möguleikarnir eru þá þeir sem taldir eru upp hér á undan, allt frá því að sleppa öllum beygingarupplýsingum öðrum en þeim sem nægja til að uppflettimynd sé einkvæm, yfir í að sýna heil beygingardæmi þar sem allar beygingarmyndir og öll afbrigði eru sýnd. Þar gefst jafnvel kostur á leiðbeiningum til notandans, um málfar, merkingarmun beygingarmynda o.s.frv. Skipting Mugdans er þá úr sögunni en í staðinn koma viðfangsefni fólgin í því að finna framsetningu sem er nægilega kerfisbundin og samt nægilega sveigjanleg til þess að þola breytingar á birtingarformi eftir því hvort notandinn er skólanemi, fræðimaður eða jafnvel tölva.

## 4.2 Setningafræði í orðabókum

Orðabókarnotendur á Íslandi ætlast til að finna upplýsingar um beygingu orða í orðabókinni sinni; það er okkar hefð. Upplýsingar um setningargerð eru mun gloppóttari í íslenskum orðabókum, eins og sjá má af gátlista í grein Eiríks Rögnvaldssonar (1998:32) sem einnig er vitnað í hér að framan (sjá 3. hluta). Á listanum eru helstu atriði sem Eiríkur taldi ástæðu til að endurskoða í *Íslenskri orðabók* vegna 3. útgáfu:

### Gátlisti ER við endurskoðun *Íslenskrar orðabókar*:

- Fjöldi rökliða (áhrifsgildi) sagna þarf að koma skýrt fram.
- Hvaða rökliðir eru skyldubundnir og hverjum má sleppa.
- Hvort og hvenær forsetningarliðir geta komið í stað andlaga.
- Fallstjórn sagna þarf að taka sérstaklega fram.
- Miðmynd þarf mjög oft að vera sjálfstætt flettiorð.
- Lýsingarháttum þarf að gera mun hærra undir höfði.
- Skerpa þarf notkun tákunarinnar ÓP.
- Takmarkanir á setningarstöðu orða verða að koma fram.

Meðal þeirra atriða sem *Íslensk orðabók* hafði einmitt helst verið gagnrýnd fyrir voru gloppóttar og ófullnægjandi upplýsingar um setningargerð, eins og Eiríkur bendir á. Í þessum lista eru þau atriði sem venjulegur málnotandi þarf að kunna skil á til þess að nota orðaforðann í bókinni, ekki bara til skilnings heldur líka til málnotkunar.

Talsverðar breytingar voru gerðar á framsetningu setningargerð-

arupplýsinga í 3. útgáfu *Íslenskrar orðabókar* og var reynt að samræma eftir bestu getu og bæta við efni eins og tími gafst til (sjá Kristín Bjarnadóttir 2001). Í ljósi þeirra breytinga er áhugavert að skoða höfuðleiðir Mugdans í framsetningu beygingarupplýsinga og hvort skipting af svipuðu tagi varpar ljósi á vandamál í framsetningu setningargerðarupplýsinga í íslenskum orðabókum.

#### 4.2.1 Setningafræði á heima í málfræðibókum

Það að vísa orðabókarnotendum algerlega á setningafræðibækur í leit að setningargerðarupplýsingum einstakra orða er einmitt stefnan sem Eiríkur Rögnvaldsson gagnrýnir með framsetningu gátlistans síns. Sem dæmi um þessa framsetningu má nefna ótölulegan fjölda sagna í 1. og 2. útgáfu *Íslenskrar orðabókar* (1963, 1983) þar sem engir rökliðir eru sýndir með sögnum, hvorki sem formlega skilgreindur setningafræðihluti né í dæmum. Af sama toga eru sagnir sem aðeins eru merktar með ÓP án þess að tiltekið sé hvort um aukafallsfrumlag eða gervifrumlag er að ræða.<sup>7</sup>

Bækur um setningafræði eru, ef eitthvað er, enn verri heimildir um málfræði einstakra orða en bækur um beygingarfræði. Í setningafræðibókum er sjónarhornið alltaf „ofan frá og niður“ og fjallað er um hverja setningargerð fyrir sig. Þess vegna er útilokað að finna tæmandi úttekt á rökliðum einstakra sagna og öðru efni af sama toga. Í þessu kemur fram eðlismunur á umfjöllun um beygingu og umfjöllun um setningargerð í málfræðibókum. Beygingardæmi eru venjulega skoðuð sem ein heild, þótt vissulega sé líka fjallað um einstök föll. Sögnum er hins vegar yfirleitt ekki skipt upp í setningargerðarflokka sem jafnast á við beygingarflokkana heldur er fjallað um hverja setningargerð fyrir sig. Það er því fremur lítil von um árangur við leit í málfræðibók þegar orðabókin gefur ekki fullnægjandi upplýsingar um rökliði sagna.

#### 4.2.2 Ófyrirsegjanlegar setningargerðir birtar í orðabókinni

Hér er hugmyndin sú að sleppa því að lýsa öllu sem er fyrirsegjanlegt út frá meginreglum íslenskrar setningargerðar og sleppa með því

<sup>7</sup>Sjá umfjöllun um sögnina *að fatra* í grein minni í *Orði og tungu* 5 (Kristín Bjarnadóttir 2001:91).

öllum upplýsingum um setningarfræðileg sjálfgildi. Þar er t.d. átt við að ekki þurfi að sýna formgerðarföll, en orðasafnsbundin föll þurfi að sýna.

Ef þessi leið væri notuð til hins ítrasta þyrfti ekki að gefa upplýsingar um frumlag í nefnifalli með sögnum, og eftir kenningunni ætti líka að vera óhætt að sleppa upplýsingum um andlag í þolfalli, þar sem þetta eru formgerðarföll. Þessi leið er fær með nefnifallið; hún var t.d. notuð í *Sýnihefti sagnorðabókar* (Ásta Svavarsdóttir o.fl. 1993) þar sem nafnháttur er notaður til að sýna setningargerðir þar sem frumlagið er sjálfgefið:

**þinga:** ...*halda fund e. ráðstefnu*

Í *Sýniheftinu* var hins vegar tekinn sá kostur að sýna þolfallsandlag með skammstöfun fyrir fallmynd rökliðarins þótt segi megi að fallið sé sjálfgildi þar sem gera þurfti greinarmun á áhrifssögn með þolfallsandlagi og áhrifslausri sögn. Þágufallsandlag þurfti svo að sýna, bæði sem orðasafnsfall og sem röklið með áhrifssögn:

**þeyta:** *hræra e. slá hratt*

**þeyta e-ð:** *hræra eða slá e-ð hratt ...*

**þeyta e-u:** *kasta e-u; losa sig við e-ð*

**þeyta e-m ...:** *kasta e-m, hrinda e-m*

Ekki er raunhæft að gera ráð fyrir að almennir notendur þekki muninn á formgerðarfalli og orðasafnsfalli og munurinn á áhrifssögn og áhrifslausri sögn gerir að engu möguleikann á því að sleppa því að sýna röklið sem er sjálfgilt fall, a.m.k. með þeirri aðferð sem hér er höfð til að sýna rökliði.

Í *Sýniheftinu* er einn merkingarþáttur í rökliðum sagna sýndur um leið og fallstjórnin, þ.e. +/-lifandi, eins og sjá má í dæmunum hér á undan. Sá háttur var tekinn upp að nokkru leyti við endurskoðun *Íslenskrar orðabókar* (2002), þannig að frumlagi er sleppt ef það er í nefnifalli og vísar til lifandi veru. Nokkur misbrestur er þó á því að svona upplýsingar séu fullkomlega samræmdar. Þarna er samt vísir að þeirri aðferð sem hér er verið að skoða og vert er að athuga hvernig upplýsingar af þessu tagi nýtast almennum notendum.

### 4.2.3 Og orðabókin á að duga ...

Samkvæmt þessari skipan á orðabókin ein að duga til að finna mögulega setningarstöðu orða. Til þess að þessu skilyrði sé fullnægt þarf að sjá til þess að a.m.k. öll atriðin á gátlista Eiríks sem vitnað er til hér að ofan komi fram. Orðtengslabækur Jóns Hilmars Jónssonar, *Orðastaðurur* (1994) og *Stóra orðabókin* (2005), leysa þennan vanda að verulegu leyti en almenn íslensk orðabók með merkingarskýringum þar sem á þessu er tekið að fullu er ekki til. Til glöggvunar fylgir hér tilbúið sýnishorn með lágmarksupplýsingum um alla rökliði þar sem notuð er sama framsetning og í *Sýniheftinu* en það var haft til fyrirmyndar þegar sagnlýsingin í *Íslenskri orðabók* var endurskoðuð:

**e-r þeytir:** *e-r hrærir e. slær hratt*

**e-r þeytir e-ð:** *e-r hrærir eða slær e-ð hratt ...*

**e-r þeytir e-u:** *e-r kastar e-u; e-r losar sig við e-ð*

**e-r þeytir e-m ...:** *e-r kastar e-m, e-r hrindir e-m*

Til þess að setja fram svona lýsingu svo að vel fari þyrfti líka að gera grein fyrir skyldubundnum atviksorðum, forsetningarliðum o.þ.h.

Höfuðleiðir Mugdans við lýsingu á beygingu orða í orðabókum bregður ljósi á vandamálin í framsetningu setningargerðarupplýsinga. Þar hefur glíman við plássleysið orðið til þess að málfræðihluti orðabókargreinanna er stórlega vanræktur og lausnin er ekki fólgin í málfræðibókum. Vandamál orðabókarmannsins er einnig sá að ekki er alltaf auðvelt að finna nauðsynleg gögn til þess að setja fram setningarlegar upplýsingar þannig að vel fari og gagn sé að.

## 5 Framsetning málfræðiupplýsinga í orðabókum

Mismunurinn á upplýsingum um beygingar og setningargerð í íslenskum orðabókum er að hluta fólgin í framsetningunni. Upplýsingar um beygingar eru alltaf settar fram á formlegan hátt og þær eiga sér fastan stað í flettugreininni. Setningargerðarupplýsingar birtast hins vegar ýmist sem svokölluð setningargerðardæmi, t.d. *gefa e-m e-ð*, eða sem eiginleg dæmi, t.d. *skrifa skýrt og rétt* (sjá Kristín Bjarnadóttir 2001 þar sem gerður er greinarmunur á eiginlegum dæmum, setningargerðardæmum og setningargerðarhausum sem hafa sjálfstætt uppflöttigildi). Oftar en ekki er setningargerðarlýsingin glöppótt

og ófullnægjandi, þ.e. hún nægir engan veginn til þess að orðabókarnotandinn geti notað orðið með stuðningi af orðabókarlýsingunni. Til þess að setningargerðarupplýsingar komi að sama gagni og beygingarupplýsingarnar þarf að setja þær fram á jafnkerfisbundinn hátt. Sjálf framsetningin skiptir einnig máli og tengist nokkuð aðferðum Mugdans sem hér hafa verið skoðaðar. Þar skiptir notendahópurinn meginmáli og viðhorf orðabókarmannsins til notandans.

Framsetning málfræðiatríða í orðabókum er einkum með þrennu móti. Í fyrsta lagi eru notaðir kótar eða tákn sem vísa í skýringar í leiðbeiningum í orðabókinni; í öðru lagi eru notaðar skammstafanir fyrir málfræðihugtök sem ætlast er til að notandinn hafi á valdi sínu; og í þriðja lagi eru notaðar skammstafanir fyrir orð sem gegna því hlutverki að sýna sama efni og felst í málfræðihugtökunum, t.d. *e-u* sem stytting á *einhverju* sem stendur í stað *þgf* eða *DAT*. Fjórða leiðin, og sennilega sú notendavænasta, er ekki algeng en hún er að sýna dæmigerða orðnotkun sem vísbendingu um setningargerð, eins og Jón Hilmar Jónsson gerir í bókum sínum. Með þessu móti fær notandinn mjög miklar upplýsingar um málnotkun, bæði setningarlegar og merkingarlegar, og ætti að vera fullfær um að nota orð sem svona eru sýnd.

Þessar leiðir gera mjög mismiklar kröfur til notandans, rétt eins og orðabókartexti sem byggður er á aðferðum Mugdans, og svigrúm fyrir málfræðiupplýsingarnar sem hægt er að koma fyrir er líka mjög misjafnt:

### Hve abstrakt má framsetningin vera?

Kótar, skammstafanir eða mannamál?

<i>Rússnesk-íslensk orðabók</i>	hneppa 2T
<i>Nimb</i> (2004)	formentlig_1 Dd1_PP
<i>tilbúið dæmi</i>	gefa DAT ACC
<i>Íslensk orðabók</i> (2002)	gefa e-m e-ð
<i>Orðastaður</i>	gefa <honum, henni><eitthvað; hlut, dýrgrip, peninga> (<fyrir greiðann, að launum, í þakklætis- skyni, að skilnaði>); ...

Tvö dæmi skera sig úr vegna þess hve miklar upplýsingar eru í þeim fólgnar, þ.e. dæmið frá Sanni Nimb sem sýnir eina gerð af setningu-



arstöðu danska atviksorðsins *formentlig* og dæmið úr *Orðastað*. Framsetningin er samt með gjörólíkum hætti; framsetningin í *Orðastað* miðar að því að gera notandanum lífið sem léttast og þar er að auki rúm fyrir blæbrigði í merkingu sem birtist í orðavali fyrir rökliði með sögninni sem þarna er sýnd. Dæmið frá Sanni Nimb er úr doktorsritgerð hennar um setningarstöðu atviksorða í dönsku (Nimb 2004:230) en í því verki er sett fram formleg greining á setningarstöðu af mikilli nákvæmni, eins og sést í flettugreininni fyrir orðið í heild:

formentlig\_1 Dd1\_PP

*Formentlig for første gang herhjemme, blev en reklamespot sendt ud i æteren til ære.*

formentlig\_2 Dd1\_PP

*Formentlig et andet forhold gør, at en række østeuropæiske mærker også trygt kan stå på gaden.*

formentlig\_3 Dd1FTe\_S

*Vi har formentlig kun set toppen af isbjerget af konsevenser endnu.*

Þetta dæmi er sýnt hér til þess að vekja athygli á tvennu, þó að efnið í þessari grein sé annars bundið við íslenska orðabókargerð. Í fyrsta lagi mótast framsetningin og greiningin af því að lýsingin er hluti af tungutækni-verkefni þar sem fengist er við öll svið setningafræðinnar og er efnið ætlað í gagnabanka sem síðar má vinna orðabókarflettur upp úr. Í öðru lagi nægir þetta dæmi til að benda á að verkefnin í orðabókarlýsingu fyrir íslensku eru mörg hver óunnin; í *Íslenskri orðabók* (2002) vottar til dæmis ekki fyrir kerfisbundinni lýsingu á setningarstöðu atviksorða.

Greining af því tagi sem Sanni Nimb setur fram er sennilega ekki sérlega vel til þess fallin að létta venjulegum orðabókarnotanda lífið; til þess er hún of tæknileg. Hún er hins vegar vísbending um það sem hægt er að gera með því að nota stór textasöfn með nákvæmri greiningu. Framsetning hverrar orðabókar er svo eftir sem áður á valdi orðabókarmannsins sem þarf að taka fullt tillit til notenda sinna og nýta allar þær aðferðir sem nú standa til boða. Orðabækur eru ekki lengur í því hafti sem prentuð bók skapar; nú er veigamesti þátturinn í allri framsetningu ekki að spara pláss heldur að ganga frá öllum upplýsingum á kerfisbundinn hátt.

## Heimildir

- Ásgeir Blöndal Magnússon. 1989. *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Ásta Svavarsdóttir. 1993. *Beygingakerfi nafnorða í nútímaíslensku*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Ásta Svavarsdóttir, Guðrún Kvaran, Jón Hilmar Jónsson og Kristín Bjarnadóttir. 1993. *Sýnihefti sagnorðabókar*. Rannsóknar- og fræðslurít 3. Reykjavík: Orðabók Háskólans.
- Eiríkur Rögnvaldsson. 1998. Málfræði í íslenskri orðabók: Hvernig og til hvers? *Orð og tunga* 4:25–32.
- Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta (eds.). 1989. *Dictionaries. An International Encyclopedia of Lexicography*. Berlin: Walter de Gruyter.
- Helgi Haraldsson. 1996. *Rússnesk-íslensk orðabók*. Reykjavík: Nesútgáfan.
- Höskuldur Þráinsson. 1995. *Handbók í málfræði*. Reykjavík: Námsgagnastofnun.
- Íslensk orðabók handa skólum og almenningi*. 1963. Ritstjóri: Árni Böðvarsson. Reykjavík: Bókaútgáfa Menningarsjóðs. [1. útg.]
- Íslensk orðabók handa skólum og almenningi*. 1983. Önnur útgáfa aukin og bætt. Ritstjóri: Árni Böðvarsson. Reykjavík: Bókaútgáfa Menningarsjóðs.
- Íslensk orðabók*. 2002. Þriðja útgáfa aukin og endurbætt. Ritstjóri: Mörður Árnason. Reykjavík: Edda hf.
- Jón Hilmar Jónsson. 1994. *Orðastaður. Orðabók um íslenska málnotkun*. Reykjavík: Mál og menning.
- Jón Hilmar Jónsson. 2002. *Orðaheimur. Íslensk hugtakaorðabók*. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2005. *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.
- Kristín Bjarnadóttir. Um sagnlýsinguna í 3. útgáfu Íslenskrar orðabókar. *Orð og tunga* 5:87–114.
- Mugdan, Joachim. 1989. Information on Inflectional Morphology in the General Monolingual Dictionary. Í: Hausmann o.fl. 1989, bls. 518–525.
- Nimb, Sanni. 2004. *Danske adverbier mellem leksikon og syntaks*. Doktorsritgerð við Kaupmannahafnarháskóla.

Sigfús Blöndal. 1920–1924. *Íslensk-dönsk orðabók*. Reykjavík. [Blöndals-orðabók.]

Sjá líka *Beygingarlýsingu íslensks nútímamáls* á vefsetri Orðabókar Háskólans: [lexis.hi.is](http://lexis.hi.is)

## Abstract

The topic of the paper is the scope and presentation of grammatical information in Icelandic dictionaries. Grammar (or linguistics) is the foundation on which dictionaries rest, but there is nevertheless a fundamental difference between a grammar and any kind of dictionary. The topic of a grammatical description is the structure of the language in question, focusing on the whole, i.e. top-down. The structure of a dictionary is based on individual entries, and the focus is bottom-up. Taking information on inflectional morphology and the transitivity and case assignment of verbs as examples, three different methods of division of labour between grammar and dictionary are discussed. The conclusion is that the dictionary needs to provide its own stand-alone description of grammatical features, with sufficient detail to provide for the user's needs, because of the difference in focus between a dictionary and a grammar.

## Lykilorð:

málfræði í íslenskum orðabókum  
beyging í íslenskum orðabókum  
rökliðir sagna í íslenskum orðabókum

## Keywords:

grammatical information in Icelandic dictionaries  
inflectional morphology in Icelandic dictionaries  
transitivity and case assignment in Icelandic dictionaries

*Kristín Bjarnadóttir*  
*Orðabók Háskólans*  
*Neshaga 16*  
*Reykjavík*  
*kristinb@lexis.hi.is*



Matthew Whelpton

# Argument Structure – For Mental Dictionaries Only?<sup>1</sup>

## 1 Introduction

There has been a long-standing division of interest between dictionary-makers and researchers in theoretical linguistics. This is reflected in the status of the grammatical information in the “dictionaries” developed by each group. Developments over the last 15 years have however opened up active areas of interaction. In this paper, I would like to look at that division and point to some of the interesting ways in which it has been breaking down.

## 2 Of lexicographers and linguists

### 2.1 The lexicon

At its minimum, a traditional dictionary is a list of citation forms and word senses. It may also include a range of other information, typically: broad phonemic transcription; information on inflectional morphology; part of speech; examples of use; and etymologies. The traditional conception of a dictionary is as an encyclopaedia of words.

---

<sup>1</sup>This paper is based on a lecture that I gave at the Orðabók Háskólans on Friday 1st April 2005. I would like to thank the participants in that seminar for their comments and insights. Needless to say, all faults and follies that remain are mine solely!

The core specifies the main semiotic coupling – form and sense. This is enriched with as much general information concerning the word as is possible given restrictions of space (the great proviso!).

Within modern linguistic theory, the role of the lexicon has been very different. In Chomsky's (1957) earliest characterisation, no lexicon was necessary at all. Words were inserted into syntactic trees by rewrite rules, as shown in Figure 1.

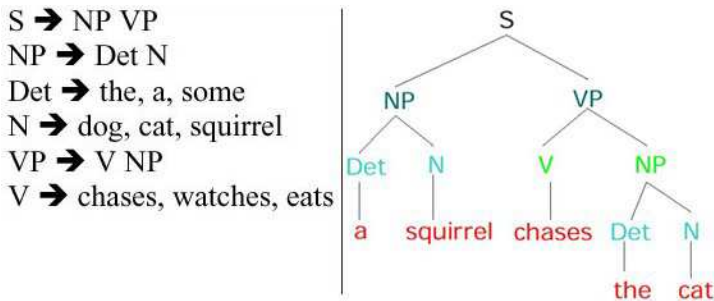


Figure 1: A simple rewrite rule grammar without lexicon

It was soon realised however that making the syntax do everything weakened rather than strengthened it. The role of the syntax was to express general principles of sentence organisation. Many aspects of sentence form, though, are related to the idiosyncratic properties of the words being used. In Chomsky's (1965) account, syntactic rules express exclusively syntactic generalisations (e.g. where do count nouns go) and a lexicon is introduced which lists all the words of the language, tagged with the syntactic information about them that must be learned (and cannot simply be deduced).

The lexicon was thus introduced as a separate module of the grammar, though its function was limited to that of a word store. From this point on within Chomskian work, the lexicon served to list the idiosyncratic grammatical information which the word carried into the sentence. Below are some examples of the kinds of grammatical information that have been suggested.

1. Rely [\_\_\_ on NP]
  - a. John relied on him.
  - b. \*John relied under him.
  - c. \*John relied against him.
2. Tell [\_\_\_ NP<sub>Acc</sub> SubC<sub>that</sub>]
  - a. John told him that Mary had left.
  - b. \*John told to him that Mary had left.
  - c. \*John told that Mary had left.
3. Say [\_\_\_ (to NP) SubC<sub>that</sub>]
  - a. \*John said him that Mary had left.
  - b. John said to him that Mary had left.
  - c. John said that Mary had left.

This information on the number and type of phrases that must follow the Verb is called a subcategorisation frame.

The kinds of information we see in subcategorisation frames illustrates very clearly what I am calling the division of interests between theoretical linguists and lexicographers. Traditional dictionaries do not require the listing of grammatical information and usually do not list it except for part of speech and examples of use – e.g. Give something to someone. The linguistic lexicon contains little but grammatical information and only those idiosyncratic elements of grammatical behaviour which are not predictable from general syntactic rule.

However, two general developments have changed this significant difference in focus between the two communities of language researchers: first, the rise of corpora and computational modelling in language research, esp. in dictionary development; and second, the rise of semantics, in particular a clear and well-articulated field of formal semantics alongside syntax.

## 2.2 Semantic Theory

The work of the philosopher Richard Montague (1974) began a tradition of linguistic analysis which has established the importance of compositionality for any serious formal theory of semantics – that is, the property that the semantic interpretation of a sentence can be read systematically off the syntactic structure of the sentence.

Independently, within the Chomskian tradition, researchers were arguing for the importance of general semantic classes of relations between items which helped to establish their syntactic behaviour – so-called thematic roles and related concepts (see Jackendoff 1972 and Fillmore 1966). Thematic roles were adopted in Chomsky's (1982) Government and Binding Theory as a way of regulating well-formedness of grammatical structures: and theta roles, like preposition selection, had to be listed in the lexicon.

Thematic role theory was often driven by discussions of “alternations”, alternative patterns of syntactic realisation.

4. Passive

- a. John hit Bill. Agent - Patient
- b. Bill was hit by John. Patient – Agent

5. Dative alternation.

- c. John gave a book to Bill. Source – Theme – Goal
- d. John gave Bill a book. Source – Goal – Theme

At the crudest level, this meant that the lexicon began to include some form of semantic information for syntactic purposes.

6. Argument structure – *give*

- a. [\_\_\_ NP<sub>Acc/Theme</sub> to NP<sub>Goal</sub>]
- b. [\_\_\_ NP<sub>Dat/Goal</sub> NP<sub>Acc/Theme</sub>]

It also however stimulated a strong tradition of research concerning the regularities expressed by labels such as “Theme” and “Goal” and the patterns of syntactic realisation with which they were associated. And this brings us to one of the most promising areas in which contemporary linguistic theory and lexicographic research can be seen to interact.

An excellent example for our purposes is the work of Beth Levin, as it explicitly relates syntactic behaviour and conceptual structure through the properties of individual verbs: especially her 1993 work on English verb classes. Levin (1993) classifies verbs in terms of the alternations that verbs can and cannot undergo. She then looks for the elements of conceptual meaning that verbs which behave the same way syntactically share and which differ systematically from the verbs which do not share their syntactic behaviour.

Consider for instance Saeed's (1997) excellent introductory discussion of Levin's analysis of the three constructions – Middle, Conative,



Body Part Ascension – as they apply to four verbs – *cut, break, touch, hit*. The results are summarised in Figure 2.

7. Middle

- a. The bread cuts easily.
- b. Crystal vases break easily.
- c. \*Cats touch easily.
- d. \*Door frames hit easily.

8. Conative

- e. Margaret cut at the bread.
- f. \*Janet broke at the vase.
- g. \*Terry touched at the cat.
- h. Carla hit at the door.

9. Body part ascension

- i. Margaret cut Bill’s arm // Margaret cut Bill on the arm.
- j. Janet broke Bill’s finger // \*Janet broke Bill on the finger.
- k. Terry touched Bill’s shoulder // Terry touched Bill on the shoulder.
- l. Carla hit Bill’s back // Carla hit Bill on the back.

	Touch	Hit	Cut	Break
<b>Conative</b>	No	Yes	Yes	No
<b>Body-part</b>	Yes	Yes	Yes	No
<b>Middle</b>	No	No	Yes	Yes

Figure 2: Verb-Construction Matching

Levin argues that the availability of a particular construction is linked to the presence of a conceptual component in the meaning of the relevant verb, as shown in Figure 3.

<b>touch</b>	Verb of pure contact	CONTACT
<b>hit</b>	Verb of contact by motion	CONTACT, MOTION
<b>cut</b>	Verb of caused change of state by contact in motion	CAUSE, CHANGE, CONTACT, MOTION
<b>break</b>	Verb of caused change of state	CAUSE, CHANGE

Figure 3: Conceptual Analysis of 4 Verbs

The specific component required by each of the constructions is shown in Figure 4:

CONSTRUCTION	COMPONENT REQUIRED
Middle	CAUSE, CHANGE
Conative	MOTION
Body-part Ascension	CONTACT

Figure 4: Conceptual component selections of 3 constructions

Here are the original examples, marked with the relevant conceptual components:

10. Middle = CAUSE, CHANGE
  - a. The bread cuts easily. = CAUSE, CHANGE, CONTACT, MOTION
  - b. Crystal vases break easily. = CAUSE, CHANGE
  - c. \*Cats touch easily. = CONTACT
  - d. \*Door frames hit easily. = CONTACT, MOTION
11. Conative = MOTION
  - e. Margaret cut at the bread. = CAUSE, CHANGE, CONTACT, MOTION
  - f. \*Janet broke at the vase. = CAUSE, CHANGE
  - g. \*Terry touched at the cat. = CONTACT
  - h. Carla hit at the door. = CONTACT, MOTION
12. Body part ascension = CONTACT
  - i. Margaret cut Bill's arm // Margaret cut Bill on the arm. = CAUSE, CHANGE, CONTACT, MOTION
  - j. Janet broke Bill's finger // \*Janet broke Bill on the finger. = CAUSE, CHANGE
  - k. Terry touched Bill's shoulder // Terry touched Bill on the shoulder. = CONTACT
  - l. Carla hit Bill's back // Carla hit Bill on the back. = CONTACT, MOTION

Levin's work is in effect refining the conceptual elements that might appear in a dictionary entry for a verb – and a dictionary that adopted such “primitive conceptual vocabulary” for its definitions would provide an invaluable resource for linguists following up on Levin's work.

Thus, semantic developments in generative grammar have led to analyses that allow a fruitful interaction between lexicography and

linguistics. Such work however is also related to the other development mentioned earlier – corpora. Levin’s work would not have been possible without the availability of computer corpora. Corpora make it possible to investigate directly the patterns in which particular words appear. A grammatically tagged corpus allows us to investigate the full range of “alternate syntactic environments” in which verbs appear.

### 2.3 Corpora and lexical analysis

This brings us to another example of work where both lexicographers and linguistics can share and interact – the WordNet project (<http://wordnet.princeton.edu/>).

“WordNet® is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets.”

- o The verb cut has 41 senses (first 10 from tagged texts)
- o 1. (58) **cut** -- (separate with or as if with an instrument; "Cut the rope")
- o 2. (18) **reduce, cut down, cut back, trim, trim down, trim back, cut, bring down** -- (cut down on; make a reduction in; "reduce your daily fat intake"; "The employer wants to cut back health benefits")
- o 3. (3) **swerve, sheer, curve, trend, veer, slue, slew, cut** -- (turn sharply; change direction abruptly; "The car cut to the left at the intersection"; "The motorbike veered to the right")
- o 4. (2) **cut** -- (make an incision or separation; "cut along the dotted line")
- o ...

Figure 5 - WordNet entry for verb “cut”

Such work is obviously of direct interest to lexicographers but as the description of WordNet shows, there is a strong linguistic motivation for the project, again in terms of the conceptual structures that underlie lexical organisation.

This development is taken much further by Charles Fillmore (Berkeley) in the FrameNet project (<http://framenet.icsi.berkeley.edu>). This is an attempt to model word properties within a cognitive grammar framework in which syntactic distributions and cognitive frames of reference for word senses come together.

As in WordNet, a verb is listed with a number of senses, though in this case, the senses are not defined in terms of synonym sets but

rather in terms of a much more complex framework of conceptual information, as shown in Figure 6.

- cut.v (Cause change of scalar position) **Created** Lexical entry
- cut.v (Experience bodily harm) **Finished** Initial Lexical entry Annotation
- cut.v (Cause harm) **Finished** Initial Lexical entry Annotation

Figure 6 - FrameNet entries for the verb *cut*

The number of senses of a word is as many as the number of frames of cognition within which it can be cognised or understood. For instance, one sense of the verb *cut* falls within the Cause\_harm frame. A frame represents a structured understanding of what is involved in a “harm-causing”, as shown in Figure 7.

### Cause\_harm

#### Definition:

The words in this frame describe situations in which an **Agent** or a **Cause** injures a **Victim**. The **Body part** of the **Victim** which is most directly affected may also be mentioned in the place of the **Victim**. In such cases, the **Victim** is often indicated as a genitive modifier of the **Body part**, in which case the **Victim** FE is indicated on a second FE layer.

Figure 7 – Frame for Causing\_harm

The frame is effectively a constellation of elements related in a systematic way which provides the context of understanding of particular instances. Here, within the Cause\_harm frame, we have an Agent/Cause and a Victim as the primary elements of the frame with provision made for the affect on the Victim being realised on a Body Part of the Victim rather than on the Victim holistically. Notice that there is a mix here of conceptual elements (Cause) and formal elements (“genitive modifier”). In this sense the frame looks like an elaborated thematic description and the frame elements look rather like traditional thematic labels, relating elements of conceptual structure to particular formal realisations. Within this framework however there is no attempt to limit the relation between a frame element and its syntactic realisations. Such relation simply are as they are.

An entry for the verb *cut* in this particular sense will then appear as in Figure 8.

**Frame: Cause\_harm**

**Definition**

COD: make an opening, incision, or wound in (something) with a sharp tool or object.

**Frame Elements and Their Syntactic Realizations**

The Frame elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realizations(s)
Agent	(40)	CNI.-- 5exx DNI.-- 1exx NP.Ext 34exx
Body_part	(36)	NP.Ext 4exx PP[on].Dep 1exx NP.Obj 31exx
Victim	(39)	2nd.-- 32exx NP.Ext 1exx INI.-- 2exx DNI.-- 1exx NP.Obj 3exx

Figure 8 – FrameNet entry for “cut.v”

This entry contains a specification of the items’s part of speech (*v.*), the frame within which it should be interpreted (which will also supply the frame elements to be mentioned in a moment), a sense which defines the specific conceptual delineation of this activity, and a grid relating the elements of the frame within which this word sense is embedded to the specific syntactic realisations of that element in uses of this particular lexical item. Notice that there are therefore two “sense” elements in this “definition” – the specific “micro” details of the activity itself (very reminiscent of a traditional dictionary definition) and the “macro” context of interpretation by which the individual words sense is related to others by a shared context of understanding.

The syntactic realisation information from Figure 8 shows the range of realisations for a particular frame element. However, the information can also be represented in a way more reminiscent of the kind of subcategorisation frame discussed earlier. Such labelled valence patterns are shown in Figure 9:

cut.v

**Valence Patterns:**

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns		
1 TOTAL	Agent	Body_part	
1 exx	NP Ext	NP Obj	
35 TOTAL	Agent	Body_part	Victim
4 exx	CNI --	NP Ext	2nd --
1 exx	CNI --	PP[on] Dep	NP Ext
1 exx	DNI --	NP Obj	2nd --
27 exx	NP Ext	NP Obj	2nd --
2 exx	NP Ext	NP Obj	INI --
4 TOTAL	Agent	Victim	
1 exx	NP Ext	DNI --	
3 exx	NP Ext	NP Obj	

Figure 9 – Valence patterns for “cut.v”

So far we have discussed the frame as the overarching interpretative umbrella under which particular word senses reside. But one frame is itself of course just one structured cluster of conceptual information which is related to others in a variety of ways. The frame Cause\_harm is thus embedded within a network of other frames, stretching out into a vast network of “world knowledge”, as shown in Figure 10.

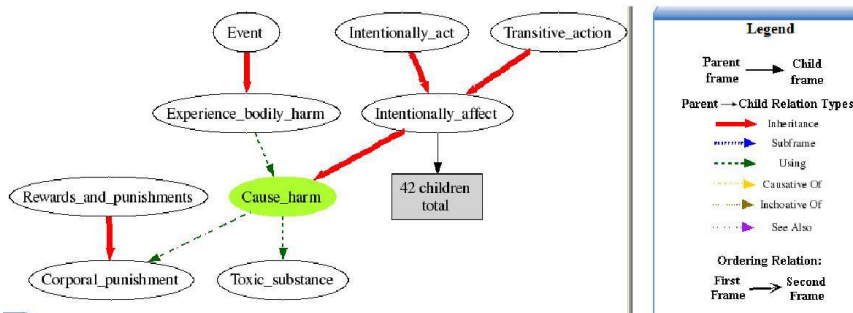


Figure 10 – Position of Cause\_harm frame in knowledge network

The parent frame of *Cause\_harm* is thus the frame *Intentionally\_affect* and inherits from it a particular set of conceptual properties. This frame can also bear other relations than inclusion/inheritance. For instance, it will be linked to particular structured frames which describe possible Causes or Agents of the harm – in Figure 10 for instance we see the frame *Toxic\_substance* listed. In this way, the representation attempts to address the problem faced by all who study lexical meaning that the sense of particular words in context seems ultimately to pull in or play on our total knowledge of how the world (and our culture) works.

The FrameNet project thus aims at the ultimate ambition of providing a computer-tractible model of human conceptual structure, within which language is embedded. I said earlier that traditional dictionaries are essentially encyclopedias of words – the FrameNet project effectively claims that an encyclopedia of words, properly pursued as a part of cognitive science, will be nothing less than an encyclopedia of the mind. . .

### **3 Conclusion**

So where does that leave us? The approaches of Levin, the WordNet project and Fillmore's FrameNet project show a range of possible ways that theoretical linguists and dictionary makers can interact. Certainly, linguists may still pursue explicit models of syntactic organisation and lexicographers may still pursue an encyclopaedic listing of the diverse properties of words. But linguists can pursue their work more effectively with the resources that a corpora-driven grammatically-sensitive dictionary can now offer. And lexicographers can expand and enrich the sophistication of the information they provide concerning a word's syntactic behaviour through the research efforts of linguists.

What kinds of information might a linguist find useful? A list of the different phrases that form a word's complementation; a list of the case assignments associated with these complementation patterns; a classification of verbs in terms of "conceptual components" or "sense type", e.g. Transfer, psychological experience, internal change of state, external change of state etc. A linguist can more easily investigate theoretical questions using data organised in this way – and I would

hope that in pursuing theoretical questions of this kind, the linguist will refine and throw further light on the kinds of conceptual characterisations and syntactic constructs that are needed in characterising the properties of words effectively – which is presumably what lexicographers want.

Certainly, recent developments should act as a spur for lexicographers, linguists, psychologists and computer scientists to work more effectively together!

## References

- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton & Co.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
- Fillmore, C. 1966. The case for case. *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston.
- Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago and London: University of Chicago Press.
- Montague, R. 1974. *Formal Philosophy*. New Haven: Yale University Press.
- Saeed, J. I. 1997. *Semantics*. Oxford: Blackwell.
- Websites accessed 31<sup>st</sup> March 2005

## Útdráttur

Greinin fjallar um stöðu rökliðagerðar (snertiflatar setningar- og merkingarfræði) í þróun orðabóka af ýmsum gerðum. Byrjað er á því að bera það gjörólíka viðhorf til merkingar og setningafræði sem kemur fram í eldri orðabókum og í fyrstu ritum málkunnáttufræðinga (generatífista). Síðan er talað um vaxandi skörun þessara tveggja sviða og möguleika á frjórri samvinnu milli orðabókafræði og málvísinda, með hliðsjón af kenningum Beth Levin á sviði málkunnáttufræði, af vinnu við gerð orðasafnslegra málheilda innan WordNet verkefnisins og af vinnu Fillmore á sviði hugrænna (kognítífra) málvísinda innan



FrameNet-verkefnisins. Niðurstaðan er sú að aldrei hafa möguleikarnir á samvinnu orðabókafræði og hinna ýmsu málvísindastefna til skýringa á rökliðagerð verið meiri.

**Keywords:**

argument structure, lexical sense, semantic frame, semantics, valency

*Dr. Matthew Whelpton*

*Senior Lecturer (Associate Professor) in English Linguistics*

*Department of English – Faculty of Humanities*

*University of Iceland*

*IS-101 Reykjavík, Iceland.*

*whelpton@hi.is*



Haraldur Bernharðsson

## *Gás, gæs og Gásir, Gásar*

### Brot úr hljóðsögu og beygingarsögu

#### 1 Inngangur

Orðið *gæs* var *gás* í fornu máli og hefur því fengið nýtt rótarsérhljóð, *æ* í stað hins eldra *á*. Í fleirtölu var það fyrst *gæss* og síðar *gæs(s)* en er nú *gásir* og hefur því áskotnast ný fleirtöluending. Örnefnið *Gásir* eða *Gásar* á hinum forna verslunarstað við ósa Hörgár í Eyjafirði og síðar bænum þar er af óvissum uppruna en ljóst er að snemma hefur það verið tengt gæsarorðinu. Ef sá skilningur örnefnisins er lagður til grundvallar að það sé í reynd gæsarorðið sýnir örnefnið enn gamla rótarsérhljóðið *á* og reyndar hefur það rutt úr vegi rótarsérhljóðinu *æ* þar sem það átti upphaflega heima. Samnafnið og örnefnið hafa því þróast hvort í sína átt. Hér verður rýnt í sögu þessara orða og leitast við að varpa ljósi á þessa þróun. Fyrst verður stuttlega hugað að forsögunni (§2) en síðan að samnafninu (§3) og loks örnefninu (§4).<sup>1</sup>

---

<sup>1</sup>Efni þessarar greinar var kynnt í fyrirlestri á Hugvísindafingi í Háskóla Íslands 23. október 2004 og þakka ég áheyrendum þar góðar umræður. Jóhannesi Bjarna Sigtryggssyni þakka ég gagnaöflun í söfnum Ordbog over det norrøne prosasprog í Kaupmannahöfn og nafnlausum ritrýni tímaritsins gagnlegar athugasemdir og ábendingar. Verk þetta var styrkt af Vísindasjóði Rannsóknamiðstöðvar Íslands.

## 2 Forsaga

Íslenska kvenkynsorðið *gæs* ‘sundfugl af andaætt (*anser*)’ á sér langa sögu og samsvaranir í fjölda annarra indóevrópskra mála. Meðal dæma úr grannmálum má nefna fær. *gás*, nno., sæ., d. *gås*, fe. *gōs*, ne. *goose*, frísn. *goes*, *guos* en upprunalega hefur verið nefhljóð í rót orðsins, eins og ljóst má vera af dæmum á borð við fhþ. *gans*, nhþ. *Gans* og holl. *gans*, enn fremur lat. *anser* (< *\*hanser*), gr. *χίψ*, sem öll merkja ‘gæs’ og skr. *hamsá-* ‘gæs, áltf’. Frumindóevrópsk mynd orðsins er því endurgerð *\*ǵ<sup>h</sup>ans-* með nefhljóði og stuttu rótarsérhljóði og í frumgermönsku hefur það verið *\*gans-*.<sup>2</sup> Í sumum germönsku málanna, þar á meðal þeim norðurgermönsku, hefur nefhljóðið fallið brott en skilið eftir, fyrst um sinn að minnsta kosti, nefkvæði á undanfarandi sérhljóði sem enn fremur hefur lengst við uppbótarlengingu (Noreen 1923:108 [§123], 168 [§233]).

Í íslensku hefur gæsarorðið beygingu kvenkyns samhljóðsstofna (rótarnafnorða) og í forníslensku hefur það því haft myndina *géss* í nf./þf. ft. Þegar í forníslensku höfðu kvenkyns samhljóðsstofnar *ō*-stofna beygingu í eintölu (að öllu eða mestu leyti) með *u*-hljóðvarpi í nf., þf. og þgf. þar sem það átti við og því gera höfundar handbóka oft ráð fyrir að á elsta skeiði hafi gæsarorðið haft nf./þf./þgf. et. *gōs*, ef. *gásar*, sbr. *ō*-stofna beyginguna nf./þf./þgf. et. *sōk*, ef. *sakar* (sjá til dæmis Noreen 1923:284–85 [§416 og athgr. 3]).

Uppruni orðsins í frg. *\*gans-* gefur, eins og áður sagði, tilefni til að ætla að rótarsérhljóðið í físl. *gōs* hafi verið nefkveðið en nefkvæði hefur verið aðgreinandi þáttur hjá löngum sérhljóðum í íslensku að minnsta kosti fram á tólftu öld, ef marka má vitnisburð höfundar Fyrstu málfræðiritgerðarinnar (útg. Hreinn Benediktsson 1972). Í kerfi (langra) nefkveðinna sérhljóða rann *ó* (þ.e. *ō*) saman við og *ó* (þ.e. *ō*), sbr. físl. *Óláfr* > *Óláfr*, *nótt* > *nótt*, en í kerfi langra munnkveðinna sérhljóða rann *ó* aftur á móti saman við *á*, sbr. físl. et. *sár* andspænis ft. *sōr* > *sár* (Hreinn Benediktsson 1959:292–93/2002:58–59, 1965:61–62). Eigi það við rök að styðjast að gæsarorðið hafi verið *gōs* — með nefkveðnu, *u*-hljóðverptu rótarsérhljóði — í elstu íslensku hefði hljóðrétt þróun þess því gefið af sér myndina *gōs*. Engin dæmi hafa þó varðveist, eftir því sem næst verður komist, um myndina *gōs* í íslensku,

<sup>2</sup>Ásgeir Blöndal Magnússon 1989:293 (*gæs*), de Vries 1962:157 (*gás 1*), Bjorvand og Lindeman 2000:333 (*gås*), Pokorny 1959, 1:412.

heldur aðeins *gás*, og það krefst skýringar. Forsendurnar fyrir því að búast við *gós* frekar en *gás* eru tvær, annars vegar nefkvæði rótarsérhljóðsins og hins vegar *u*-hljóðvarp, og rétt að huga nánar að þessu tvennu:

Í fyrsta lagi verður að hafa í huga að nefkvæði sérhljóða er almennt ekki auðkennt sérstaklega í stafsetningu í varðveittum íslenskum handritum (Hreinn Benediktsson 1972:130–31), þótt leifar slíkrar auðkenningar megi sjá í Íslensku hómilíubókinni, Sth. perg. 15 fol. frá um 1200, þar sem táknun nefkveðins *ó* (þ.e. *ø*) er í heildina frábrugðin táknun munnkveðins *ó* og *ó* (de Leeuw van Weenen 1993:60–61). Í Hómilíubókinni eru þó engin dæmi um gæsarorðið (de Leeuw van Weenen 2004) svo að hvorki þar né í stafsetningu annarra íslenskra handrita frá elsta skeiði fæst staðfesting á þeirri ætlun að sérhljóðið í *gós* hafi verið nefkveðið. Hún er því einvörðungu reist á samiburði við önnur mál, eins og áður var greint frá, og reyndar virðist ekki brýn ástæða til að efast um að brottfall nefhljóðsins í rótinni hafi valdið nefkvæði sérhljóðsins; eðlilegra virðist aftur á móti að spyrja hve lengi nefkvæði sérhljóðsins var kerfislega aðgreinandi og hvenær hið nefkveðna *ó* féll saman við nefkveðið *ó*. Til að mynda gæti það hugsast að í einhverri mállýsku eða mállýskum elstu íslensku hafi aðgreining hinna nefkveðnu *ó* og *ó* haldist alveg þangað til nefkvæði hætti að vera aðgreinandi í sérhljóðakerfinu; í þessari (íhaldssömu) mállýsku hefði hið upprunalega nefkveðna *ó* þá ekki fallið saman við *ó* (eins og algengast var) heldur væntanlega hið langa munnkveðna *ó* sem síðar féll svo (í öllum mállýskum) saman við *á*, sbr. físl. ft. *sór* > *sár*. Þar með hefði komið upp myndin *gás* sem með tímanum hefði getað rutt *gós* úr vegi. Þessi hugmynd verður þó ekki studd neinum sjálfstæðum heimildum og hér verður því að nægjast með getgátur einar.

Í öðru lagi hlýtur sú spurning að vakna hvort gæsarorðið hafi sannanlega haft *u*-hljóðvarp í nf., þf. og þgf. et., þ.e. hvort það hafi í raun haft fullkomna *ø*-stofna beygingu í eintölu. Því er til að svara að, eftir því sem næst verður komist, hafa engin dæmi varðveist er sýnt gætu að þetta orð hafi í elsta máli haft rótarsérhljóðið *ó*. Hinu hljóðverpta *ó* og óhljóðverpta *á* er haldið í sundur í stafsetningu á til að mynda Íslensku hómilíubókinni, Sth. perg. 15 4to, en hvorki þar né í öðrum handritum á elsta skeiði er að finna nein dæmi um gæsarorðið (sjá til dæmis orðasafn Larssons 1891). Kvenkyns samhljóðsstofnar (rótarnafnorð) hafa, eins og áður var getið, jafnan *ø*-stofna beyg-

ingu í eintölu (að mestu eða öllu leyti), eins og til dæmis físl. *rong* og *mork*, ft. *rengr* og *merkr*, og þar sem gæsarorðið hefur fleirtölu að hætti kvenkyns samhljóðsstofna, físl. nf./þf. ft. *gæss*, úr enn eldra *géss*, er ekki óeðlilegt að ætla að eintalan hafi beygst að hætti *ō*-stofna og því haft *u*-hljóðvarp í nf., þf. og þgf. et. í elstu íslensku (Noreen 1923:282–86 [§§412–17]). Þar sem sú ætlun verður ekki staðfest með dæmum í varðveittum handritum er ef til vill ekki hægt að útiloka að *u*-hljóðverpta nefkveðna rôtarsérhljóðið *ó* hafi við áhrifsbreytingu vikið fyrir hinu óhljóðverpta *á*, sem til dæmis var í ef. et. *gásar* (sbr. ef. et. *sakar* af *søk*). Þar með væri komin skýring á því hvers vegna gæsarorðið varð *gás*, en ekki *gós*, þar sem hið nefkveðna *á* féll saman við munnkveðið *á* þegar nefkvæði hætti að vera aðgreinandi í sérhljóðakerfinu. Þó mælti það ef til vill gegn slíkri skýringu að þar sem eignarfall er markað væri síður við því að búast að það hefði haft áhrif með þessum hætti á nefnifallið (og reyndar einnig þolfallið og þágufallið) sem er ómarkað, eins og vikið verður að síðar (§3.2 og §4.3).<sup>3</sup>

Þróun kvenkynsorðsins *ást* er að hluta til sambærileg við þróun gæsarorðsins. Einnig þar sýnir samanburður við skyld mál að nefhljóð hefur fallið brott í rót með uppbótarlengingu og nefjun undanfarandi sérhljóðs: fhþ. *anst* ‘gleði, hylli, þakklæti’, gotn. *ansts* ‘ást’ úr frg. *\*ansti-* (Ásgeir Blöndal Magnússon 1989:28 [1 *ást*]). Þetta orð er að vísu *i*-stofn í bæði fornháþýsku og gotnesku og vissulega hefur það *i*-stofna fleirtölu í íslensku, nf./þf. ft. *ástir*, en í elstu íslensku hefur rôtarsérhljóðið tekið *u*-hljóðvarpi í nf., þf. og þgf. et., *óst*, andspænis ef. et. *ástar*, og því hefur eintalan *ō*-stofna beygingu. Merki *u*-hljóðvarpsins í nf., þf. og þgf. andspænis hljóðvarpsleysi eignarfallsins koma til að mynda vel fram í Íslensku hómilíubókinni þar sem sérhljóðunum *ó* og *á* er haldið skýrt aðgreindum í stafsetningu (de Leeuw van Weenen 1993:56–61, 2004:190 [ó*st*]). Hér hefði því hljóðrétt þróun verið sú að hið nefkveðna *ó* hefði runnið saman við nefkveðið *ó* er síðar féll saman við munnkveðið *ó* en stafsetning í Hómilíubókinni bendir ekki til þessarar þróunar og það gerir heldur ekki mynd þessa orðs í yngra máli þar sem físl. *óst* varð *ást* en ekki *óst*.

Samanburður við þróun físl. *óst* í *ást* í stað *óst*, að svo miklu leyti sem á honum er byggjandi, virðist því geta rennt stoðum undir þá

<sup>3</sup>Í *Lexicon poeticum* (1931:175 [gás]) lætur Finnur Jónsson í ljós efasemdir um tilvist hljóðverptu myndarinnar því að við flettiórðið *gás* bætir hann í sviga: „eller *gós?*, men ordet er måske aldrig blevet egl. *ō*-st.“

skoðun að físl. *gás* hafi þróast úr eldra *gós* með *u*-hljóðverptu rótarsérhljóði en þar hafi *ó*, sem í öndverðu var nefkveðið, ekki runnið saman við *ó*, eins og búist var við, heldur *á*. Slíkt samfall hefur þá ekki verið hljóðrétt í þessum orðum, heldur verður að gera ráð fyrir að áhrifsbreyting hafi verið liður í þróuninni. Eftir hljóðrétt samfall hinna nefjuðu *ó* og *ó* (og síðar samruna hins nefjaða *ó* við munnkveðið *ó*) voru í þessu orði og fleiri orðum af þessu tagi víxl *ó* og *á*, *nf./þf./þgf. gós* : *ef. gásar*. Ekki er ólíklegt, eins og Hreinn Benediktsson (1979:321–22/2002:237–38) hefur bent á, að þessi víxl, sem voru fremur óvenjuleg í beygingarkerfinu, hafi orðið fyrir áhrifum frá víxlum munnkveðins *ó* og *á* (eins og í *ótt*, *ef. áttar*) og því hafi hið opna *ó* verið endurreist við áhrifsbreytingu í gæsarorðinu og fleiri orðum með sams konar víxlum *ó* og *á*. Síðar, við samruna *á* og *ó* í *á*, urðu *nf./þf./þgf. gós* : *ef. gásar* reglulega að *nf./þf./þgf. gás* : *ef. gásar*. Til glöggvunar má sýna þessa þróun eins og gert er í (1) þar sem skaftlaus ör (>) táknar hljóðbreytingu en skeft ör (→) áhrifsbreytingu.

- (1) *nf./þf./þgf. et. gós* > *gós* → *gós* > *gás*  
*ef. et. gásar*

Þróun beygingardæmisins er sýnd í (2). Í (2a) er beygingin eins og vænta má að hún hafi verið á tólftu öld eða um 1200. Jafnframt þróuninni frá *gós(-)* í *gás(-)*, sem rædd hefur verið, varð *nf./þf. ft. géss* að *gæss* við samruna *é* og *ó* í *æ* á þrettánda öld (Hreinn Benediktsson 1959:296–97/2002:61) og það gefur beygingardæmið í (2b) nálægt 1300 eða á fyrri hluta fjórtánda aldar. Í bakstöðu tók *ss* að styttest í *s* á fjórtánda öld (Björn K. Þórólfsson 1925:xxx–xxxi) og eftir þá styttingu lítur beygingardæmið út eins og í (2c), um 1400 eða á fimmtánda öld.

(2) Þróun beygingardæmisins frá elstu íslensku fram á fimmtánda öld

		a.	b.	c.
		<i>um 1200</i>	<i>um 1300</i>	<i>um 1400</i>
<i>et.</i>	<i>nf.</i>	<i>gós</i>	<i>gás</i>	<i>gás</i>
	<i>þf.</i>	<i>gós</i>	<i>gás</i>	<i>gás</i>
	<i>þgf.</i>	<i>gós</i>	<i>gás</i>	<i>gás</i>
	<i>ef.</i>	<i>gásar</i>	<i>gásar</i>	<i>gásar</i>
<i>ft.</i>	<i>nf.</i>	<i>géss</i>	<i>gæss</i>	<i>gæs</i>
	<i>þf.</i>	<i>géss</i>	<i>gæss</i>	<i>gæs</i>
	<i>þgf.</i>	<i>gósum</i>	<i>gásum</i>	<i>gásum</i>
	<i>ef.</i>	<i>gása</i>	<i>gása</i>	<i>gása</i>

Ef hugað er að beygingu orðsins í nútímamáli má sjá að á því hafa orðið ýmsar breytingar. Nútímamálsbeygingin íslenska er sýnd í (3a) og til samanburðar beygingin í færeysku í (3b) en hún samsvarar, að breyttu breytanda, þeirri beygingu sem var í íslensku um 1400 og sýnd var í (2c). Enn fremur eru til samanburðar orðin *mús* (3c) og *lús* (3d) sem hafa sams konar beygingu í nútímaíslensku og færeysku (*Føroysk orðabók* 1998:1431; Höskuldur Þráinsson o.fl. 2004:89).

(3) Íslenskt og færeyskt nútímamál

		a.	b.	c.	d.
		<i>íslenska</i>	<i>færeyska</i>	<i>ísl. og fær.</i>	<i>ísl. og fær.</i>
<i>et.</i>	<i>nf.</i>	<i>gæs</i>	<i>gás</i>	<i>mús</i>	<i>lús</i>
	<i>þf.</i>	<i>gæs</i>	<i>gás</i>	<i>mús</i>	<i>lús</i>
	<i>þgf.</i>	<i>gæs</i>	<i>gás</i>	<i>mús</i>	<i>lús</i>
	<i>ef.</i>	<i>gæsar</i>	<i>gásar</i>	<i>músar</i>	<i>lúsar</i>
<i>ft.</i>	<i>nf.</i>	<i>gæsir</i>	<i>gæs</i>	<i>mýs</i>	<i>lýs</i>
	<i>þf.</i>	<i>gæsir</i>	<i>gæs</i>	<i>mýs</i>	<i>lýs</i>
	<i>þgf.</i>	<i>gæsum</i>	<i>gásu</i>	<i>músum</i>	<i>lúsum</i>
	<i>ef.</i>	<i>gæsa</i>	<i>gása</i>	<i>músa</i>	<i>lúsa</i>

Í *nf./þf. ft. mýs* og *lýs* hefur átt sér stað sams konar stytting *ss* og áður var lýst fyrir *nf./þf. ft. gæss* > *gæs*, þ.e. *nf./þf. ft. mýss* > *mýs* og *lýss* > *lýs*; samsvarandi stytting hefur einnig átt sér stað í færeysku þar sem *nf./þf. ft.* er *gæs*, *mýs* og *lýs*. Í færeysku fylgja orðin *gás*, *mús* og *lús* því öll sama mynstri, með óhljóðverptu rótarsérhljóði í allri eintölunni og *þgf.* og *ef.* fleirtölunnar en *i*-hljóðvarpi (*á* > *æ*, *ú* > *ý*) og núllendingu (-Ø) í *nf./þf.* fleirtölunnar: *gæs*-Ø, *mýs*-Ø og *lýs*-Ø. Í íslensku fylgja *mús* og *lús* þessu mynstri en beyging *gæsar*orðsins íslenska hefur tekið allnokkrum breytingum. Breytingarnar á beygingu íslenska *gæsar*orðsins andspænis beygingu orðanna *mús* og *lús* bæði í íslensku og færeysku og einnig færeyska orðsins *gás* eru tvíþættar, eins og lýst er í (4).

(4) Ísl. *gæs* andspænis fær. *gás* og ísl./fær. *mús* og *lús*

- Nf./þf. ft.* hafa fengið nýtt fleirtölumyndan (-*ir* í stað -Ø).
- Víxlum tveggja rótarsérhljóða (*á* : *æ*, sbr. *ú* : *ý*) hefur verið útrýmt (og *æ* alhæft).

Við styttingu *ss* í bakstöðu, bæði í íslensku og færeysku, í *nf./þf. ft. gæs*-s, *mýs*-s, *lýs*-s vék fleirtölumyndanið -s fyrir núllendingu og greinarmunur *nf./þf. et. gás*-Ø, *mús*-Ø, *lús*-Ø og *nf./þf. ft. gæs*-Ø, *mýs*-Ø,



*lýs*-Ø hvíldi eftir það alfarið á víxlum rótarsérhljóðanna. Ekki var því hætt á samfalli eintölu og fleirtölu og reyndar er sama mynstur að finna í nútímaensku, eins og sýnt er í (5).

(5) Enska

- a. et. *goose* : ft. *geese*
- b. et. *mouse* : ft. *mice*
- c. et. *louse* : ft. *lice*

Í íslensku hefur gæsarorðið farið aðra leið og það krefst skýringar. Meginspurningin sem leitað verður svara við í næsta kafla er því þessi:

(6) Hvers vegna þróaðist ísl. *gás* á aðra leið en *mús* og *lús*?

### 3 Samnafnið: *gás, gæs og gæsir*

#### 3.1 Aldur breytinganna

Fleirtalan nf./þf. *géss* > *gæs(s)* > *gæs* í samnafninu virðist einhöfð í elsta máli og allar götur fram á sextánda öld. Hún birtist til að mynda í þf. ft. „*gæs*“ í Járnsíðu á Staðarhólsbók, AM 334 fol. (101vb2) frá um 1271–81 (ljóspr. útg. Ólafur Lárusson 1936), þf. ft. „*gíæs*“ í Göngu-Hrólfs sögu á AM 589 f 4to (35v23) frá um 1450–1500 (ljóspr. útg. Loth 1977) og „*gæs*“ á AM 152 fol. (115va20) frá um 1500–1525. Elstu þekktu dæmi um nýtt beygingarmyndan nf./þf. ft. í samnafninu eru frá miðri sextánda öld og sautjándu öld, eins og sýnt er í (7). Dæmi um nýtt beygingarmyndan í nf./þf. ft. í örnefninu, *Gásir*, eru nokkru eldri eða frá miðri fimmtánda öld, eins og síðar verður vikið að (§4.2).<sup>4</sup>

(7) Elstu þekktu dæmi um nýtt beygingarmyndan nf./þf. ft.

- a. nf. ft. „*gæsser*“ í Búalögum á AM 128 4to (90r11) frá um 1544 (útg. Jón Þorkelsson 1915–33:51.186)<sup>5</sup>

<sup>4</sup>Hér og annars staðar í dæmum sem tekin eru stafrétt úr handritum er ekki gerður greinarmunur á „s“ og „l“.

<sup>5</sup>Í *Katalog* 1:417 segir Kálund AM 128 4to frá síðari helmingi fimmtándu aldar og Jón Þorkelsson (1915–33:46) segir bókina skrifaða um 1490–1500. Jón Helgason (1932:147–48 neðanmáls), sem álitur að bókinn sé vart svo gömul, bendir á að á henni megi finna ártalið 1544 og telur ekki fráleitt að bókinn geti verið frá þeim tíma; þá aldursákvörðun er að finna í *ONP Registre* 1989:41.

- b. nf. ft. „Grägiæser“ í Búalögum á GKS 3671 b 8vo (76r18) frá 1641 (útg. Jón Þorkelsson 1915–33:181.141)<sup>6</sup>
- c. nf. ft. „grägiæser“ í Búalögum á JS 343 II 8vo (bls. 47.16–17) frá um 1650 (útg. Jón Þorkelsson 1915–33:189.220)
- d. þf. ft. „gięsir“ í Kormáks sögu á NKS 1147 fol. (24rb21) frá sautjándu öld
- e. þf. ft. „Giæser“ í Kormáks sögu á AM 554 g 4to (18r31) frá sautjándu öld
- f. nf. ft. „gragaser“ í Búalögum á Lbs. 66 4to (bls. 147a.5) frá 1661 (útg. Jón Þorkelsson 1915–33:197.283)
- g. nf. ft. „giæser“ í Búalögum á Lbs. 869 4to (143v11) frá um 1670 (útg. Jón Þorkelsson 1915–33:68.176)
- h. nf. ft. „Grägiæsir“ í Búalögum á AM 60 8vo (210v1 [84v1]) frá síðari hluta sautjándu aldar (útg. Jón Þorkelsson 1915–33:220.765)
- i. nf. ft. „giæser“ í Þiðriks sögu á Sth. papp. fol. 100 (165r11) frá síðari hluta sautjándu aldar (sbr. útg. Bertelsens 1905–11, 1:99 lesbrigðaskrá B)
- j. nf. ft. „giæser“ í Þiðriks sögu á AM 177 fol. frá 1690–91 (útg. Bertelsen 1905–11, 1:99 lesbrigðaskrá B)<sup>7</sup>

Við þetta má bæta að í málfræði Runólfs Jónssonar 1651 er „Gaas“ skipað í beygingarflokk (aðra beygingu kvenkynsorða) með orðinu *drós*. Runólfur (1651:15) sýnir beygingu orðsins *drós* í heild en það hefur nf./þf. ft. *drósir*, og meðal annarra kvenkynsorða í þessum flokki eru *krás*, *krús*, *rás* og *rós*. Ekki er því um að villast að orðið *gás* hefur haft endinguna *-ir* í nf./þf. ft. í máli Runólfs. Ekki er þó jafnaugljóst hvert rótarsérhljóð nf./þf. ft. hefur verið þar sem Runólfur sýnir ekki beygingu orðsins í heild. Úr því að hann getur ekki víxla rótarsérhljóða mætti ætla að það hafi verið *á* og nf./þf. ft. þá *gásir*. En það

<sup>6</sup>Í næstu línu fyrir ofan kemur fyrir „grágasar egg“ (GKS 3671 b 8vo, 76r17) þar sem ef. et. af *gás* hefur enn rótarsérhljóðið *á*, sbr. einnig dæmið í (7f).

<sup>7</sup>Við listann má bæta nf. ft. „giæser“ í Búalögum á Lbs. 6 fol. (323r24) frá 1623 (útg. Jón Þorkelsson 1915–33:152.311) en Jón Þorkelsson (1915–33:144) telur að textinn kunni að vera ritaður af dönskum manni „því að sumt í ritshættinum hér í Búalögum bendir á það. Ritarinn hefir að vísu skilið allvel íslenzku, en misritanir ýmsar í textanum eru þess eðlis, að þær geta tæplega verið eptir íslenzkan mann“. — Í *Íslenzku fornbréfasafni* 2:738.24 (nr. 479) er prentað nf. ft. „gæser“ í máldaga Þykkvabæjarklausturs eftir AM 263 fol. sem ritað hefur verið 1598; í handritinu er þó ótvírætt ritað „gæss“ (bls. 83).

er þó varasöm ályktun því að í þennan flokk setur hann orðið *nös* án nokkurra athugasemda um víxl rótarsérhljóða sem hljóta að hafa verið í beygingu þess þá sem nú (et. *nös*, ft. *nasir* en varla *\*nösir*).

Gamla myndin nf./þf. ft. *gæs* lifir þó áfram í uppskriftum eldri texta. Í 57. kap. búnaðarbálks Jónsbókar kemur fyrir þf. ft. *gæss* (útg. Ólafur Halldórsson 1904:190.14) og þá fleirtölumynd (með núllendingu eftir styttingu *ss*) er að finna í mörgum Jónsbókarhandritum frá sextándu og sautjándu öld og í útgáfunni 1578 er prentað þf. ft. „*Giæs*“ (ljóspr. útg. Ólafur Lárusson 1934:x.ij). Í ritmálsskrá Orðabókar Háskólans er aðeins að finna eitt dæmi um nf./þf. ft. *gæs(s)* og er það í ljóði eftir Matthías Jochumsson frá 1910 (1956–58, 1:557.25); þar er vísast um fornryði að ræða en ekki daglegt mál á síðari hluta nítjándu aldar eða fyrri hluta þeirrar tuttugustu.

Rótarsérhljóðið *á* í *gás(-)* virðist allsráðandi á þrettándu, fjórtándu og fimmtándu öld, til dæmis þf. et. „*gas*“ í Ólafs sögu helga á Sth. perg. 2 4to (27v8), frá um 1250–1300 (ljóspr. útg. Jón Helgason 1942; útg. Johnsen og Jón Helgason 1941, 1:217.2), þf. et. „*gas*“ í Kormáks sögu á Möðruvallabók, AM 132 fol. (127va41), frá um 1330–70, og þf. et. „*gasina*“ (127vb7 og 127vb13) (útg. van Arkel-de Leeuw van Weenen 1987) og þf. et. „*gas*“ í Ólafs sögu helga á Tómasskinnu, GKS 1008 fol. (115v17) frá um 1450–1500 (ljóspr. útg. Loth 1964). Elstu þekktu dæmi um alhæfingu rótarsérhljóðsins æ úr nf./þf. ft. *gæs(s)* á kostnað *á* eru frá síðari hluta sextándu aldar og sautjándu öld, eins og sýnt er í (8).

(8) Elstu þekktu dæmi um alhæfingu rótarsérhljóðs nf./þf. ft.

- a. nf. et. „*Giæs*“, þf. et. „*Giæs*“ í Guðbrandsbiblíu 1584 (Bandle 1956:263)
- b. þf. et. „*gięsina*“ í Kormáks sögu á NKS 1147 fol. (þrjú dæmi: 24rb18–19, 24rb22, 24rb25) frá sautjándu öld (sbr. útg. Einars Ól. Sveinssonar 1939:282–83)
- c. þf. et. „*gięsina*“ í Kormáks sögu á AM 554 g 4to (18v5–6) frá sautjándu öld<sup>8</sup>

<sup>8</sup>Meðal dæma í ritmálssafni Orðabókar Háskólans er þf. et. *gæs* sem sagt er frá fyrsta þriðjungi sextándu aldar, nánar tiltekið 1508. Dæmið er í orðtakinu *þá skal grípa gæs er gefst* sem Páll Eggert Ólason (1944:53) eignar Ögmundi Pálssyni Skálholtsbiskupi og tengir atburði sem sagður er hafa átt sér stað 1508 (eða 1505; sjá Jón G. Friðjónsson 1993:210). Ekki er ljóst á hvaða heimildum þetta er byggt og því ekki hægt að reiða sig á að Ögmundur hafi notað myndina *gæs* frekar en *gás*.

Myndir með rótarsérhljóðinu *á* eru þó lífseigar í uppskriftum gamalla texta. Í Kormáks sögu á NKS 1147 fol. er jafnframt dæmunum í (8b) eitt dæmi um þf. et. „Gäs“ (24rb18) og í AM 554 g 4to eru jafnframt dæminu í (8c) þrjú dæmi um þf. et. með *á*: „gaas“ (18r26), „heim-gaasina“ (18r27) og „gaasina“ (18v1–2).<sup>9</sup> Guðmundur Andrésón (d. 1654) hefur bæði „Gäs“ og „Giæs“ sem flettiorð í orðabók sinni sem prentuð var 1683 (útg. 1999:56, 59) en segir að *gæs* sé venjulega notað um fuglinn en *gás* aftur á móti um ‘kvensköp’.<sup>10</sup> Björn Halldórsson (1724–94) hefur einnig bæði „Gás“ og „Gæs“ í orðabók sinni sem prentuð var 1814 (útg. 1992:171, 195). Aftur á móti þýðir Jón Árnason lat. *anser* aðeins með „Giæs“ í *Nucleus Latinitatis* 1738 (útg. 1994:10) og í sömu flettu koma einnig fyrir þgf. et. „Giæs“ og ef. ft. í samsetningunni „Giæsa feiti“. Í málfræði Jóns Magnússonar (1997:146–147), sem lokið hefur verið 1738, er getið um bæði nf. et. „gäs“ og „giæs“ en því miður er þar engin umfjöllun um beygingu orðanna (og ekki greint frá merkingu þeirra). Myndir með rótarsérhljóðinu *á* hafa því lífað lengi jafnframt myndum með *æ* og slíkar myndir er að finna í ritmálssafni Orðabókar Háskólans að minnsta kosti fram á nítjándu öld. Til að mynda lifir *gás* í orðasambandinu að *gjalda einhverjum gagl fyrir gás* ‘launa einhverjum eitthvað illa, svíkja einhvern í viðskiptum’ sem kemur fyrir þegar í fornu máli en verður þó að teljast sjaldgæft í nútímamáli (Jón G. Friðjónsson 1993:174). Þá er og að finna í ritmálssafninu nokkur dæmi um nf./þf. ft. *gásir* þar sem rótarsérhljóðið *á* hefur rutt *æ* úr vegi, sbr. (7f) að framan, en þau dæmi eru reyndar ekki mörg.

Ef marka má þau dæmi sem fundist hafa er upphafs beggja breytinganna í samnafninu að leita á sextándu öld. Ýmislegt virðist benda til að breyting fleirtöluendingarinnar (nf./þf. ft. -Ø → -ir, sbr. 4a) sé

<sup>9</sup>Auk þess er í AM 554 g 4to ef. ft. „Gaasa“ (18v9) rímskorðað í vísu (útg. Einar Ól. Sveinsson 1939:283) en þeirri vísu er sleppt í NKS 1147 fol. Það vekur athygli að í þessum kafla Kormáks sögu (útg. Einar Ól. Sveinsson 1939:282–83) þar sem eru fjögur dæmi um þf. et. orðsins *gás* er það fyrsta dæmið af fjórum í textanum sem hefur rótarsérhljóðið *á* í NKS 1147 fol. og fyrstu þrjú í AM 554 g 4to: skrifarnir fylgja forritum sínum framan af en svo tekur þeirra eigin mál völdin og *gás*-víkur fyrir *gæs*-.

<sup>10</sup>„Gäs / f.g. *Anser, Gansa* [svo]; item *Vulva muliebris*, aliàs **giæs** solet vocari *Avis, gäs* Os *Vulvæ*, porro“ (Guðmundur Andrésón útg. 1999:56). Guðmundur vísar hér til þess að jafnframt *gás* ‘sundfugl af andaætt’ er til kvenkynsorðið *gás* er hefur merkinguna ‘kvensköp’ og síðar einnig ‘gála, lauslát kona’. Síðarnefnda orðið, sem heldur rótarsérhljóðinu *á*, hefur ekki þróast á sama hátt og fuglsorðið og er því ekki til umræðu hér.

ívið eldri en útbreiðsla rótarsérhljóðsins *æ* (*á* → *æ*, sbr. 4b) eða að fyrrnefnda breytingin hafi að minnsta kosti fyrr náð fótfestu en sú síðarnefnda. Meðal þess sem nefna mætti til stuðnings þeirri skoðun er þetta:

- (9) a. Elsta dæmi um nýju fleirtöluendinguna *-ir* (7a) er ívið eldra en elsta dæmi um útbreiðslu rótarsérhljóðsins *æ* (8a).
- b. Í málfræði Runólfs Jónssonar 1651 er nf. et. *gás* en ending nf./þf. ft. er *-ir*
- c. Sautjándu aldar skrifarar NKS 1147 fol. og AM 554 g 4to rita bæði *gás(-)* og *gæs(-)* í þf. et., sbr. (8b–c), en þf. ft. hefur endinguna *-ir*, sbr. (7d–e).
- d. Í ritmálssafni Orðabókar Háskólans eru nokkur dæmi um varðveislur rótarsérhljóðsins *á* fram á nítjándu öld en engin raunveruleg dæmi um nf./þf. ft. *gæs(s)*.
- e. Dæmi um nýtt beygingarmyndan í nf./þf. ft. í örnefninu *Gásir* koma fram þegar um miðja fimmtánda öld (sjá §4.2).

Þessar röksemdir eru auðvitað misveigamiklar og nauðsynlegt að hafa fyrirvara á. Nefna mætti að minnsta kosti þessa:

- (10) a. Aldursmunur á elstu dæmunum (7a) og (8a) er óverulegur og aldur handrita eins og í (7a) verður ekki ákvarðaður með mikilli nákvæmni.
- b. Það gæti verið hrein tilviljun að breyting nf./þf. ft. birtist fyrr í rituðum heimildum en breyting rótarsérhljóðsins í eintölunni vegna þess að dæmi um eintöluna eru fá; þetta orð er einfaldlega miklu algengara í fleirtölu en eintölu.
- c. Nýtt beygingarmyndan í örnefninu nf./þf. ft. *Gásir* hefur takmarkað gildi fyrir þróun samnafnsins því að breytingin í örnefninu er að nokkru leyti annars eðlis, eins og síðar verður rætt (§4.3).

Niðurstaðan um aldur breytinganna í samnafninu er þá þessi: Þau dæmi sem fundist hafa í rituðum heimildum benda til að þessar breytingar hafi farið af stað á sextánda öld. Á grundvelli þessara heimilda verður ekki með óbyggjandi hætti úr því skorið hvor breytingin muni eldri en svo virðist sem breyting endingar nf./þf. ft. *-Ø* í *-ir* (4a) hafi fyrr náð fótfestu en breyting rótarsérhljóðsins *á* í *æ* í eintölumyndum orðsins (4b). Ekki er því ósennilegt að fyrrnefnda breytingin sé eldri en sú síðarnefnda.

### 3.2 Eðli breytinganna

Í allri beygingarþróun gegna innbyrðis vensl ólíkra beygingarmynda veigamiklu hlutverki. Margvísleg rök hafa verið færð fyrir því að beygingarmyndir innan eins beygingardæmis séu „missterkar“, ef svo má að orði komast, að ein mynd geti verið eins konar grundvallarmynd sem aðrar myndir beygingardæmisins séu byggðar á. *Mörkun* (e. *markedness*) er gjarna notuð sem mælikvarði á þessi vensl og grundvallarmyndin þá sögð *ómörkuð* (e. *unmarked*) en hinar „afleiddu“ aftur á móti *markaðar* (e. *marked*). Þegar fyrir miðja síðustu öld benti pólski málfræðingurinn Jerzy Kuryłowicz (1945–49:23) á að við áhrifsbreytingar væri það yfirleitt „grunnmynd“ sem hefði áhrif á „afleidda mynd“ og samband grunnmyndar og afleiddrar myndar væri afleiðing af notkunarviði þeirra. Rannsóknir á máltöku barna benda til hins sama: börn tileinka sér ómarkaðar myndir fyrr en markaðar og við myndun mörkuðu myndanna eru hinar ómörkuðu lagðar til grundvallar (sjá til dæmis Hooper 1980, Bybee og Brewer 1980 og Bybee 1985).

Allajafna er litið svo á að eintala sé ómörkuð andspænis fleirtölu og því er algengt að eintölumyndir hafi áhrif á gerð fleirtölumynda við áhrifsbreytingar. Nokkur dæmi úr íslenskri málsögu er sýna þessa þróun eru í (11).

- (11) Eintölumynd lögð til grundvallar við myndun nýrrar fleirtölumyndar
- |    |                           |   |                                    |
|----|---------------------------|---|------------------------------------|
| a. | et. <i>brún-Ø</i>         | : | ft. <i>brún-n</i> → <i>brún-ir</i> |
| b. | et. <i>glóð-Ø</i>         | : | ft. <i>glóð-r</i> → <i>glóð-ir</i> |
| c. | et. <i>qlpt-Ø, álft-Ø</i> | : | ft. <i>elpt-r</i> → <i>álft-ir</i> |

Nafnorðin í (11) hafa í nútímaíslensku öll fengið nýjar fleirtölumyndir sem ýmist lifa við hlið eldri myndanna eða hafa komið alveg í stað þeirra. Þarna hefur átt sér stað áhrifsbreyting á leið til nútímamáls og í öllum tilvikum hefur nýja fleirtölumyndin verið mynduð á grunni eintölumyndarinnar: ný fleirtölumynd hefur orðið til með því að bæta fleirtöluendingunni *-ir* við eintölustofninn (*brún-*, *glóð-*, *álft-*) og um leið víkja eldri fleirtölumyndirnar (*brún-n*, *glóð-r*, *elpt-r*) sem allar höfðu annað rótarsérhljóð en eintölumyndirnar. Eintalan er þarna ómörkuð andspænis fleirtölunni og þar sem börn á máltökuskeiði tileinka sér eintöluna á undan fleirtölunni er ekki óeðlilegt að upp komi tilhneiging til að mynda fleirtöluna á grundvelli eintölunnar. Þar með

hverfa ýmis ófyrirsegjanleg frávík fleirtölustofnsins frá eintölustofnum, eins og til dæmis *i*-hljóðvarpsvínlin *ú : ý, ó : ó (æ)* og *á (< o) : e*.

Tiersma (1982) hefur sýnt fram á að við ákveðnar kringumstæður geti þær myndir sem allajafna eru taldar *markaðar* verið grundvöllur fyrir breytingar á þeim myndum sem venjulega eru *ómarkaðar*. Tiersma (1982:834) sýnir dæmi um víxl hnígandi og rísandi tvíhljóða í eintölu og fleirtölu nafnorða í frísnesku. Þessi víxl eru ófyrirsegjanleg og því er rík tilhneiging til að útrýma þeim. Langalgengast er að tvíhljóð eintölunnar sé alhæft og komi í stað upprunalegs tvíhljóðs fleirtölunnar, eins og í dæmunum í (12).

- (12) a. et. *hoer* : ft. *hworren* → *hoeren* 'vændiskona'  
 b. et. *koal* : ft. *kwallen* → *koalen* 'kol'  
 c. et. *miel* : ft. *mjillen* → *mielen* 'mál, nyt (mjólkurkúa)'  
 d. et. *poel* : ft. *pwollen* → *poelen* 'laug'

Fáein orð sýna þó gagnstæða þróun þar sem fleirtalan er lögð til grundvallar og tvíhljóð fleirtölunnar er alhæft, eins og sýnt er í (13):

- (13) a. et. *earm* → *jerm* : ft. *jermen* 'armur'  
 b. et. *goes* → *gvos* : ft. *gwozzen* 'gæs'  
 c. et. *hoarn* → *hwarne* : ft. *hwarnen* 'horn (á dýri)'  
 d. et. *hoas* → *vjazze* : ft. *vjazzen* 'sokkur'  
 e. et. *kies* → *kjizze* : ft. *kjizzen* 'tönn'  
 f. et. *spoen* → *spwon* : ft. *spwonnen* 'spónn, flís'  
 g. et. *toarn* → *twarne* : ft. *twarnen* 'þyrnir'  
 h. et. *trien* → *trjin* : ft. *trjinnen* 'tár'

Útjöfnunin sem á sér stað í dæmunum í (13), þar sem fleirtalan er lögð til grundvallar, bendir til þess að þar sé fleirtalan í raun *ómörkuð* andspænis eintölunni. Þessi dæmi eru því undantekning frá almennum lögmálum um mörkun.

Tíðni orðmynda er einn af nokkrum mælikvörðum á mörkun. Rannsóknir Greenbergs (1966:32) á tíðni eintölu og fleirtölu (og tvítölu ef því er að skipta) í nafnorðum í nokkrum indóevrópskum málum sýna að eintölumyndir eru miklum mun algengari en fleirtölumyndir (og tvítölumyndir langsaldgæfastar þar sem þær koma á annað borð fyrir). Ef tekið er tillit til þess að eintölumyndir nafnorða eru að jafnaði miklu algengari en fleirtölumyndirnar kemur ekki á óvart að eintölumyndirnar skuli yfirleitt verða fleirtölumyndunum yfirsterkari við áhrifsbreytingar.

Tiersma (1982:834–35) bendir á að orðin í (13), sem virðast víkja frá þessum almennu lögmálum um mörkun, vísa öll til para eða hópa: sokkar, armar og horn (á dýrum) eru oftast í þörum og gæsir, spæningar, tár, tennur og þyrnar eru oftast í „hópum“, ef svo má segja. Öll eru þessi orð því algengari í fleirtölu en eintölu og að dómi Tiersma skýrir það hvers vegna fleirtölumyndirnar eru ómarkaðar andspænis eintölunni; tilvik af þessu tagi kallar Tiersma *local markedness* sem nefna mætti *sérmörkun* á íslensku:

(14) Sérmörkun (e. *local markedness*), sbr. Tiersma 1982:835

Nafnorð er ómarkað í *fleirtölu* ef það merkir eitthvað sem frá náttúrunnar hendi er oftast í þörum eða hópum.

Tiersma sýnir tíðnitölur máli sínu til stuðnings úr hollensku, ensku, þýsku og spænsku og þær sýna glögglega að orð þau í þessum málum er merkja ‘armur’, ‘horn (á dýri)’, ‘sokkur’, ‘tönn’, ‘spónn, flís’, ‘þyrnir’ og ‘tár’ eru mun meira notuð í fleirtölu en eintölu.

Tölurnar fyrir gæsarorðið í þessum málum sýna reyndar ekki jafnskýra yfirburði fleirtölu yfir eintölu og eru eintölumyndir þess jafnvel algengari en fleirtölumyndir í sumum málanna. Skýringin er væntanlega sú að í borgarsamfélagi nútímans eru gæsir ekki jafnalgengar í stórum hópum eins og var í landbúnaðarsamfélagi fyrri alda; málnotkunin hefur því breyst samfara breyttum lifnaðarháttum. Þróun frísneska orðsins í (13b) sýnir þó að fleirtalan hefur verið lögð til grundvallar — hún hefur verið ómörkuð andspænis eintölunni — og vitnar þá líkast til um eldri málnotkun landbúnaðarsamfélagsins.

Þegar fleirtölumynd er ómörkuð andspænis eintölumynd getur það haft í för með sér eins konar *vangreiningu* (e. *under-analysis*) á fleirtölumyndinni þar sem fleirtölumerki hennar glatar merkingu sinni. Börn á máltökuskeiði heyra eintölumyndina sjaldan og tileinka sér því fleirtölumyndina sem grunnmynd í stað þess að skynja hana sem markaða mynd með fleirtölumerki. Þá getur komið upp tilhneiging til að bæta við skýru fleirtölumerki til að tákna með ótvíræðum hætti fleirtölumerkingu grunnmyndarinnar og verður það viðbót við eldra fleirtölumerkið sem glatað hefur merkingu sinni; úr verður það sem nefnt er *tvöföld fleirtala* (d. *double plural*). Tiersma (1982:837–39) sýnir dæmi um tvöfalda fleirtölu, meðal annars úr vesturfrísnesku (15a–f) þar sem fleirtöluendingunni *-s* hefur verið bætt ofan á fleirtöluendinguna *-en* (en hvor tveggja endingin er virk í frísnesku) og úr hollensku þar sem fleirtöluendingunni *-en* hefur verið bætt ofan á fleirtöluend-



inguna *-er* (15g–h). Svipuð dæmi eru enn fremur þekkt úr enskri mál-sögu, svo sem þegar ft. *cȳ* (Campbell 1959:254 [§628.2]; Brunner 1965: 228 [§284, athgr. 4]) var aukin með fleirtöluendingunni *-ne* en aðgreining eintölu og fleirtölu fólst þar áður í víxlunum *ū* : *ȳ* (15i).

- (15)
- |    |                          |   |  |
|----|--------------------------|---|--|
| a. | vfrísn. et. <i>boei</i>  | : | ft. <i>boeijens</i> 'handjárn'                 |
| b. | vfrísn. et. <i>lears</i> | : | ft. <i>learzens</i> 'stígvél'                  |
| c. | vfrísn. et. <i>reed</i>  | : | ft. <i>redens</i> 'skautar'                    |
| d. | vfrísn. et. <i>trep</i>  | : | ft. <i>treppens</i> 'þrep'                     |
| e. | vfrísn. et. <i>weach</i> | : | ft. <i>weagens</i> 'bylgja'                    |
| f. | vfrísn. et. <i>wolk</i>  | : | ft. <i>wolkens</i> 'ský'                       |
| g. | holl. <i>blad</i>        | : | ft. <i>blader</i> → <i>bladeren</i> 'laufblað' |
| h. | holl. <i>ei</i>          | : | ft. <i>eier</i> → <i>eieren</i> 'egg'          |
| i. | fe. et. <i>cū</i>        | : | ft. <i>cȳ</i> → <i>kine</i> 'kýr'              |

Snúum okkur þá aftur að íslenska gæsarorðinu. Ekki verður séð að fleirtölumyndir orðsins *gæs* séu algengari en eintölumyndirnar í nútímaíslensku. Dæmin í *Íslenskri orðtíðnibók* (Jörgen Pind o.fl. 1991:174) eru reyndar fá, aðeins sex úr fjórum textum, og skiptast þannig að tvö eru um þf. et. og eitt um þgf. et., nf. ft., þf. ft. og þgf. ft.; jafnmörg dæmi eru því um eintölu og fleirtölu. Í seðlasafni *Ordbog over det norrøne prosasprog* í Kaupmannahöfn eru 26 dæmi um *gás*: fimmtán eru fleirtölumyndir en sex eru eintölumyndir og af eintölumyndunum sex eru fjórar úr einum og sama kaflanum í Kormáks sögu.<sup>11</sup> Að svo miklu leyti sem þessar tölur gefa einhverja vísbendingu um tíðni gæsarorðsins í fornu máli sýna þær allmikla yfirburði fleirtölu yfir eintölu og gæsarorðið virðist því geta fallið undir skilgreiningu Tiersma á sérmörkun.

Tvenns konar breytingar áttu sér stað á beygingu gæsarorðsins íslenska, eins og lýst er í (16), sbr. (4):

- (16) Ísl. *gás*: tvær breytingar á sextánda öld
- Nf./þf. ft. fengu nýtt fleirtölumyndan (*-ir* í stað *-Ø*).
  - Víxlum tveggja rótarsérhljóða (*gás-* : *gæs-*) var útrýmt (og *æ* alhæft).

Eins og rætt var að framan (§3.1) er erfitt að skera úr um það með óyggjandi hætti hvor breytingin muni eldri en ýmislegt þykir þó

<sup>11</sup>Ekki eru talin með þrjú dæmi um örnefnið *at/frá Gásum*, tvö dæmi um viðurnefnið *gás* (sem ekki er til í fleirtölu) og eitt dæmi um *gás* 'kvensköp'.

benda til þess að nýja fleirtöluendingin *-ir* (16a) sé eldri en alhæfing rótarsérhljóðsins *æ* (16b).

Freistandi er að rekja viðbót fleirtöluendingarinnar *-ir* til þess að fleirtalan var ómörkuð andspænis eintölu. Á fjórtánda öld varð nf./þf. ft. *gæss* að *gæs* við hljóðréttu styttingu *ss* í bakstöðu (*ss* > *s*). Þar með vék fleirtöluendingin *-s* fyrir endingunni *-Ø* í gæsarorðinu, eins og sýnt var í (2), og þar með fólst beygingarfræðileg aðgreining eintölu og fleirtölu í víxlum rótarsérhljóðanna *á* og *æ*. Sama þróun átti sér stað í orðunum *mús* og *lús*. Til glöggvunar er þetta sýnt aftur í (17):

(17) Stytting *ss* í bakstöðu á fjórtánda öld

	a.	b.	c.
et. nf./þf.	<i>gás-Ø</i>	<i>mús-Ø</i>	<i>lús-Ø</i>
ft. nf./þf.	<i>gæs-s</i> > <i>gæs-Ø</i>	<i>mýs-s</i> > <i>mýs-Ø</i>	<i>lýs-s</i> > <i>lýs-Ø</i>

Ef eintala gæsarorðsins var sjaldheyrð gátu börn á máltökuskeiði hætt að skynja þetta hlutverk hljóðavíxlanna *á* og *æ* og bætt við nýju fleirtölumerki: *gæs-Ø* → *gæs-ir*. Aðgreining eintölu og fleirtölu byggðist í þessu tilviki á víxlum rótarsérhljóðanna *á* og *æ*; rótarsérhljóðin höfðu hvort um sig engu beygingarfræðilegu hlutverki að gegna, heldur var það *samband* þeirra sem hafði þetta aðgreiningarhlutverk. Ef eintalan var sjaldheyrð og fleirtalan ríkjandi gat þetta *víxlsamband* sérhljóðsins *á* í eintölu og *æ* í fleirtölu rofnað; þar með skapaðist þörf fyrir nýtt fleirtölumerki líkt og í dæmunum í (15). Ekki kemur á óvart að endingin *-ir* skyldi verða fyrir valinu þar sem sú ending virðist býsna virk meðal kvenkynsorða. Þegar í forníslensku gætir ríkrar tilhneigingar kvenkynsorða með nf./þf. ft. *-ar* eða *-r* til að taka upp endinguna *-ir* (Björvand 1972) og í nútímaíslensku er *-ir* algengasta fleirtöluending kvenkynsorða en *-Ø* er aftur á móti afar fátíð hjá kvenkynsorðum (Ásta Svavarsdóttir 1993:104–105, 110–11).

Hafi fleirtala gæsarorðsins verið ómörkuð andspænis eintölu skýrir það enn fremur hvers vegna rótarsérhljóð fleirtölunnar (*æ*) var alhæft á kostnað rótarsérhljóðs eintölunnar (*á*); sú þróun er þá fyllilega sambærileg við þróunina í frísnesku orðunum í (13). Þróun beygingardæmisins á sextánda öld hefur þá verið eins og sýnt er í (18).

## (18) Próun beygingardæmisins á sextánda öld

		a.	b.	c.
		<i>um 1400</i>	<i>sextánda öld</i>	<i>sextánda öld</i>
<i>et.</i>	<i>nf.</i>	<i>gás</i>	<i>gás</i>	→ <i>gæs</i>
	<i>þf.</i>	<i>gás</i>	<i>gás</i>	→ <i>gæs</i>
	<i>þgf.</i>	<i>gás</i>	<i>gás</i>	→ <i>gæs</i>
	<i>ef.</i>	<i>gásar</i>	<i>gásar</i>	→ <i>gæsar</i>
<i>ft.</i>	<i>nf.</i>	<i>gæs</i>	→ <i>gæsir</i>	<i>gæsir</i>
	<i>þf.</i>	<i>gæs</i>	→ <i>gæsir</i>	<i>gæsir</i>
	<i>þgf.</i>	<i>gásu</i>	<i>gásu</i>	→ <i>gæsum</i>
	<i>ef.</i>	<i>gása</i>	<i>gása</i>	→ <i>gæsa</i>

Eins og áður var getið verður sérmörkun hjá gæsarorðinu rakin til þess að gæsir eru oftast í hópum og því oftast talað um þær í fleirtölu (þótt reyndar hafi líkast til dregið úr fleirtölunotkuninni í þéttbýlissamfélagi nútímans). Nokkuð öðru máli gegnir um orðin *mús* og *lús*. Mýs og lýs birtast mönnum ekki í hópum með sama hætti og því er þar ekki um sérmörkun að ræða. Orðin *mús* og *lús* hafa því ekki þróast á sama veg og gæsarorðið þar sem fleirtalan var lögð til grundvallar og reyndar örlar frekar á tilhneigingu í gagnstæða átt hjá mýsarorðinu: þar er tilhneiging til að endurskapa fleirtöluna á grundvelli eintölunnar. Sú þróun er sýnd í (19a) og felst í því að rótarsérhljóð eintölunnar, *ú*, breiðist út á kostnað rótarsérhljóðs *nf./þf. ft., ý*, og einnig kemur endingin *-ir* í stað núllendingarinnar (*-Ø*) í *nf./þf. ft.*; þessi þróun er þá fyllilega sambærileg við hina almennu tilhneigingu sem lýst var í (11) og (12). Í ritmálssafni Orðabókar Háskólans er að finna eitt dæmi um *nf. ft. músir* en það er úr ljóðmælum séra Bjarna Gissurarsonar í Lbs. 838 4to frá síðari hluta sautjándu aldar eða fyrsta þriðjungi átjándu aldar, sjá (19b). Slík dæmi koma einnig fyrir í nútímamáli, eins og (19c–f) sem fengin eru úr íslenskum vefritum með aðstoð leitarvélanna *Google* (<http://www.google.com>) og *AltaVista* (<http://www.altavista.com>).<sup>12</sup>

<sup>12</sup>Tiersma (1982:839) nefnir reyndar dæmi þess að í miðensku hafi komið upp mállyskubundið fleirtalan [gīzn], [fītn] og [mīzn] í stað *geese, feet* og *mice*. Þar hafa þá orðin fyrir 'fót' og 'mús' þróast á sama hátt og gæsarorðið og kýrorðið (*kine*).

- (19) a. et. *mús-Ø* : ft. *mýs-Ø* → *mús-ir*  
 b. „enn dūsuna atu rid og **músir** [nf. ft.]“ í Lbs. 838 4to (bls. 214.11; leturbr. hér)  
 c. stýristautar, minnisspjöld, hljóðstyrksstjórar, **músir** (nf. ft.) og aðrir hlutir  
 d. **Músir** (nf. ft.) fyrir tölvuna  
 e. Gulli á sæta hunda sem heita Snotra og Perla og eru algerar **músir** (nf. ft.) en hann hefur aldrei leyft mér að sjá þær  
 f. engar hreyfingar í músagildrum, allt eins og það átti að vera og ég laus við allar **músir** (þf. ft.) í bili

Niðurstaðan um eðli breytinganna í samnafninu og um leið svarið við spurningunni í (6) er þá þessi: Bæði breyting endingarinnar nf./þf. ft. -Ø → -ir og alhæfing rötarsérhljóðsins æ á kostnað á má að öllum líkindum rekja til þess að gæsarorðið var algengara í fleirtölu en eintölu og fleirtalan var því ómörkuð gagnvart eintölunni; þetta er það sem hér hefur verið nefnt *sérmörkun* (e. *local markedness*) að hætti Tiersma (1982). Aftur á móti hefur fleirtala orðanna *mús* og *lús* verið mörkuð gagnvart eintölunni, eins og algengast er meðal nafnorða, og því hefur samsvarandi þróun ekki átt sér stað þar.

## 4 Örnefnið: at Gásu, á Gás(a)eyri, Gásir, Gásar, Gæsir

### 4.1 Uppruni

Vel er þekktur verslunarstaður sá við ósa Hörgár í Eyjafirði sem í forn-um heimildum er nefndur at *Gásu*. Uppruni þessa heitis er þó ekki fullkomlega ljós en varðveittar myndir þess í elstu heimildum eru allveg samhljóða gæsarorðinu og ljóst að snemma hefur örnefnið verið tengt við fuglsheitið. Hér er ekki rúm til umræðu um uppruna örnefnisins enengið verður út frá því að þar sé á ferð gæsarorðið.<sup>13</sup>

<sup>13</sup>Sjá um verslunarstaðinn hjá Birni Þorsteinssyni og Guðrúnu Ásu Grímsdóttur 1989:150–52; enn fremur hjá Margréti Hermannsdóttur 1987 og í safni greina hjá Christophersen og Dybdahl (ritstj.) 1999.

## 4.2 Yfirlit um helstu dæmi og beygingarþróun

Verslunarstaðarinn við Hörgárósa er víða getið í fornum heimildum (sbr. Hagland 1999 og Björn Vigfússon 2002). Hann er nefndur í nokkrum Íslendinga sögum og þáttum frá fyrri hluta þrettánda aldar og fram á miðja fjórtánda öld, eins og sýnt er í (20), en einvörðungu í þágufalli (fleirtölu), *at/á/frá Gásum*, þrettán sinnum, eða í eignarfalli (fleirtölu), *til Gása*, einu sinni; alls eru það því fjórtán dæmi. Aðeins í yngstu sögunum, Grettis sögu frá um og upp úr 1300 og Þorleifs þætti jarlaskálds frá því um 1300 eða fyrri hluta fjórtánda aldar, kemur fyrir heitið *á Gáseyri*.

(20) Verslunarstaðurinn nefndur í Íslendinga sögum og þáttum

- a. Víga-Glúms saga (frá fyrri hluta þrettánda aldar; útg. Jónas Kristjánsson 1956): þgf. *Gásur* (1 dæmi)
- b. Sneglu-Halla þáttur (frá fyrri hluta þrettánda aldar; útg. Jónas Kristjánsson 1956): þgf. *Gásur* (1 dæmi í Morkin-skinnugerð, 1 dæmi í Flateyjarbókargerð)
- c. Ljósvetninga saga (frá miðri þrettánda öld; útg. Björn Sigfússon 1940): þgf. *Gásur* (1 dæmi)
- d. Reykdæla saga (frá miðri þrettánda öld; útg. Björn Sigfússon 1940): þgf. *Gásur* (3 dæmi), ef. *Gása* (1 dæmi)
- e. Ögmundar þáttur dytts (frá síðari hluta þrettánda aldar eða um 1300; útg. Jónas Kristjánsson 1956): þgf. *Gásur* (1 dæmi)
- f. Grettis saga (frá um eða upp úr 1300; útg. Guðni Jónsson 1936): þgf. *Gásur* (3 dæmi), *á Gáseyri* (1 dæmi)
- g. Þórðar saga hreðu (frá miðri fjórtánda öld; útg. Jóhannes Halldórsson 1959): þgf. *Gásur* (2 dæmi)
- h. Þorleifs þáttur jarlaskálds (frá lokum þrettánda aldar eða fyrri hluta fjórtánda aldar; útg. Jónas Kristjánsson 1956, Österholm 1987): *á Gáseyri* (1 dæmi)

Í Sturlunga sögu, safni þrettánda aldar sagna sem steypt var saman í eitt rit um 1300 (Guðrún Nordal 1992:310), eru alls 24 dæmi: sextán um þágufallið *at/frá Gásur* og átta um eignarfallið *til Gása*, eins og sýnt er í (21). Aðeins einu sinni er verslunarstaðurinn nefndur *á Gásaeyri* en það er í Þorgils sögu skarða í klausu sem aðeins er varðveitt í BL Add. 11.127, sautjándu aldar eftirriti Reykjarfjarðarbókar, AM 122 b fol. (útg. Kálund 1906–11, 1:303.11).

(21) Verslunarstaðurinn nefndur í Sturlunga sögu (steyppt saman um 1300; útg. Kálund 1906–11)

- a. Prestssaga Guðmundar Arasonar: þgf. *Gásum* (3 dæmi)
- b. Guðmundar saga dýra: þgf. *Gásum* (2 dæmi), ef. *Gása* (1 dæmi)
- c. Íslendinga saga: þgf. *Gásum* (6 dæmi), ef. *Gása* (5 dæmi)
- d. Þórðar saga kakala: þgf. *Gásum* (4 dæmi), ef. *Gása* (1 dæmi)
- e. Þorgils saga skarða: þgf. *Gásum* (1 dæmi), ef. *Gása* (1 dæmi), *á Gásaeyri* (1 dæmi)

Í biskupa sögum eru sex dæmi um þágufallið *at Gásum* og eitt um eignarfallið *til Gása*, eins og sýnt er í (22), en notkun *Gáseyrar*-nafnsins færir þar mjög í vöxt og í Lárentíus sögu frá miðri fjórtándu öld er verslunarstaðurinn hartnær alltaf *á/af/við Gáseyri* og einu sinni aðeins *á Eyri*.

(22) Verslunarstaðurinn nefndur í biskupa sögum

- a. Jóns saga H (C) (frá þrettándu öld, í sautjándu aldar uppskriftum frá Norðurlandi; útg. Foote 2003a–b): *Gáseyri* (1 dæmi)
- b. Árna saga (frá fyrri hluta fjórtándu aldar; útg. Guðrún Ása Grímsdóttir 1998): ef. *Gása* (1 dæmi)
- c. Guðmundar saga A („elsta sagan“, frá fyrri hluta fjórtándu aldar; útg. Stefán Karlsson 1983): þgf. *Gásum* (6 dæmi)
- d. Guðmundar saga Arngríms (D) (frá miðri fjórtándu öld; útg. *Biskupa sögur* 1858–78, 2): *Gáseyri* (1 dæmi), *Gáseyrar* (1 dæmi)
- e. Lárentíus saga (frá miðri fjórtándu öld; útg. Guðrún Ása Grímsdóttir 1998): *Gáseyri* (13 dæmi í A-gerð, 13 dæmi í B-gerð), *í Gáseyraskipinu* (1 dæmi í A-gerð), *á Eyri* (1 dæmi í A-gerð)

Annálarnir frá fjórtándu öld og síðar sýna svipaða mynd: þar er oft talað um verslunarstaðinn *á Gáseyri* en að *Gásum*.

## (23) Verslunarstaðurinn nefndur í annálum

- a. Lögmansannáll (frá miðri fjórtándu öld; útg. Storm 1888 [VII]): *Gásu*m (2 dæmi), *á Gáseyri* (4 dæmi)
- b. Annálsbrot frá Skálholti (frá síðari hluta fjórtándu aldar; útg. Storm 1888 [VI]): *á Gásaeyri* (1 dæmi)
- c. Flateyjarannáll (frá síðari hluta fjórtándu aldar; útg. Storm 1888 [IX]): þgf. *Gásu*m (2 dæmi), *á Gáseyri* (4 dæmi)
- d. Oddaverjaannáll (frá sextándu öld; útg. Eiríkur Þormóðsson og Guðrún Ása Grímsdóttir 2003): þgf. *Gásu*m (3 dæmi), *á Gáseyri* (1 dæmi)
- e. Gottskálksannáll (frá síðari hluta sextándu aldar; útg. Storm 1888 [VIII]): *Gásu*m (1 dæmi), *á Gáseyri* (2 dæmi)

Þessar elstu heimildir um verslunarstaðinn sýna sem sagt einungis þgf. *Gásu*m og ef. *Gása* en engin dæmi um nefnifall eða þolfall örnefnisins. Á fjórtándu öld er verslunarstaðurinn oft nefndur *á Gáseyri* og virðist það heiti því yngra en *at Gásu*m (sbr. einnig Finn Jónsson 1908). Ef sá skilningur er lagður til grundvallar að örnefnið sé í reynd gæsarorðið má reikna með því að beygingarþróun örnefnisins á þrettándu og fjórtándu öld hafi verið sú sama og hjá samnafninu. Þessi þróun er sýnd í (24): þróun samnafnsins er endurtekin úr (2) og hafa verður hugfast að strangt til tekið er nefnifall og þolfall örnefnisins óþekkt og því rétt að hafa fyrirvara þar á.

## (24) Fleirtölumyndir samnafnsins og örnefnisins

<p>a.</p> <p><i>um 1200 &gt; um 1300 &gt; um 1400</i></p> <p><i>ft. nf. géss &gt; gæss &gt; gæs</i></p> <p><i>þf. géss &gt; gæss &gt; gæs</i></p> <p><i>þgf. gósum → gásu</i>m</p> <p><i>ef. gása</i></p>	<p>b.</p> <p><i>um 1200 &gt; um 1300 &gt; um 1400</i></p> <p><i>Géss &gt; Gæss &gt; Gæs (?)</i></p> <p><i>Géss &gt; Gæss &gt; Gæs (?)</i></p> <p><i>Gósum → Gásu</i>m</p> <p><i>Gása</i></p>
---	--

Í fornbréfum frá fimmtándu og sextándu öld birtist nefnifalls- og þolfallsmyndin *Gásir*, eins og sýnt er í (25), en þar er alls staðar átt við jörðina fremur en verslunarstaðinn sjálfan, enda bendir ýmislegt til að verslun hafi þá að mestu verið aflögð á *Gásu*m.<sup>14</sup>

<sup>14</sup>Finnur Jónsson (1908:3) telur hugsanlegt að nafnið *Gáseyr(r)* hafi sprottið af þörf fyrir að greina verslunarstaðinn eða höfnina frá bænum sem fyrst er nefndur í bréf-

## (25) Bæjarnafnið í fornbréfum (leturbr. hér)

- a. „Gasir“ (1446, skrá yfir eignir Munkaþverárklosturs, *DI* 4:699 [nr. 732])
- b. „ad aboten sele ... iordina **gaser** er ligur j glæsebæiar kirkiusokkn“ (1452 á Hólum, *DI* 5:89 [nr. 80])
- c. „hier i mote gaf abotinn iordina **gaser** er liggur i glæsi-bæiar kirkiu sokn“ (1460 á Munkaþverá, *DI* 5:91 [nr. 82])
- d. „Þetta j eignum ... hraukbær, gardzhorn, **gæsir**. dynhagi. biarger“ (1525, elsti hlutinn af Sigurðarregistri; eignir Möðruvallaklosturs, *DI* 9:318 [nr. 266–78])

## (26) Bæjarnafnið í nokkrum yngri heimildum

- a. „*Gaaser*, kallast nú almennilega *Gæser*.“ Jarðabók Árna Magnússonar og Páls Vídalín 1712 (Árni Magnússon og Páll Vídalín 1943, 10:179)
- b. „Í Apríl seint fórst Jón Björnsson á Gásum norður...“ Mælifellsannáll við árið 1735 (*Annálar 1400–1800*, 1:652)
- c. Málnotkun um miðbik tuttugustu aldar: *að fara í Gæsi* (þf.) eða *í Gási* (þf.) (Nilsson 1975:56)
- d. *Gásir* (flettiorð). „Nafn bæjarins hefur einnig verið ritað Gásar og Gæsir.“ *Byggðir Eyjafjarðar* 1990, 2:574–75.

Eins og þessar heimildir sýna birtist nefnifall og þolfall örnefnisins (og þá haft um bæinn fremur en verslunarstaðinn) fyrst í heimildum á fimmtánda öld og þá með rötarsérhljóðinu *á* í stað hins eldra *æ* og endingunni *-ir* í stað eldri núllendingarinnar: *Gásir*. Á næstu öldum og fram til nútíma birtist örnefnið ýmist sem *Gásir*, *Gásar* eða *Gæsir*. Kyn örnefnisins hefur einnig verið nokkuð á reiki (26c) og eru heimildir frá sjöunda áratug tuttugustu aldar um að það hafi í máli sumra verið karlkyns og þolfallsmyndin þá bæði *Gási* og *Gæsi* (Nilsson 1975:56) og væntanlega hefur *Gása* einnig verið til. Þessi þróun er sýnd til frekari glöggvunar í (27) og þróun samnafnsins er þar endurtekin úr (18).

---

um um miðja fimmtánda öld, sbr. dæmin í (25). Ekki er ljóst hvenær menn tóku að hafa vetursetu eða fasta búsetu at *Gásum* en þó bendir eftirfarandi frásögn úr 2. kap. Sneglu-Halla þáttar til að það hafi ef til vill verið snemma: „Bárðr hét maðr ok var hirðmaðr Haralds konungs. Hann sigldi til Íslands ok kom út at Gásum ok vistaðisk þar um vetrinn“ (útg. Jónas Kristjánsson 1956:264). Ég þakka Orra Vésteinssyni fyrir að vekja athygli mína á þessari frásögn.



## (27) Beygingarþróun samnafnsins og örnefnisins

	a.	b.
<i>ft. nf.</i>	<i>gæs</i> → <i>gæsir</i>	<i>Gæs</i> (?) → <i>Gásir, Gásar, Gæsir</i>
<i>þf.</i>	<i>gæs</i> → <i>gæsir</i>	<i>Gæs</i> (?) → <i>Gási(r), Gása(r), Gæsi(r)</i>
<i>þgf.</i>	<i>gásu</i> m → <i>gæsu</i> m	<i>Gásu</i> m
<i>ef.</i>	<i>gása</i> → <i>gæsa</i>	<i>Gása</i>

Í þróun örnefnisins er því að finna þrennt sem ekki sést í þróun samnafnsins:

- (28) a. Rótarsérhljóðið *á* hefur komist inn í *nf./þf. ft.* örnefnisins (*Gás-*) í stað hins upprunalega *æ* og lifir þar (jafnframt *æ*) til nútímamáls (fáein dæmi fundust um *á* í *nf./þf. ft.* samnafnsins en það virðist ekki hafa verið algengt eða langlíf þar).
- b. Nýtt beygingarmyndan *nf./þf. ft.* örnefnisins er *-ir* (*Gásir, Gæsir*) en einnig *-ar* (*Gásar*) (samnafnið fékk *-ir* en engar heimildir eru um að það hafi fengið *-ar*).
- c. Örnefnið þekkist einnig í karlkyni og fær því karlkynsbeygingu (engar heimildir eru um að samnafnið hafi skipt um kyn).

Því er enn ósvarað hvers vegna örnefnið hefur orðið viðskila við samnafnið og þróast á annan veg. Leitast verður við að skýra þetta í næsta kafla og verður lögð höfuðáhersla á muninn sem lýst er í (28a) og meginspurningin er þessi:

- (29) Hvers vegna er rótarsérhljóðið í örnefninu oft *á* (*Gásir, Gásar*) í nútímamáli en ekki eingöngu *æ* (*Gæsir*) eins og í samnafninu (*gæsir, gæsir, gæsum, gæsa*)?

### 4.3 Eðli breytinganna

Áður var á það minnst (§3.2) að vensl einstakra beygingarmynda innan beygingardæmis gegndu mikilvægu hlutverki í beygingarþróuninni: ómarkaðar myndir hefðu oftast áhrif á markaðar myndir. Venjulega er gert ráð fyrir að nefnifall sé ómarkað andspænis aukaföllum (sjá til dæmis Greenberg 1966:37–38 og um íslensku hjá Eiríki Rögnvaldssyni 1990:64–65 og Ástu Svavarsdóttur 1993:27, 49–53). Ef tillit er

tekið til þess mætti búast við því að við áhrifsbreytingar réði nefnifallið oft ferðinni og hefði áhrif á form hinna fallanna. Dæmi um slíkt er sýnt í (30) þar sem fornum víxlum rótarsérhljóðanna *y* og *u* í orðinu *dyr* (físl. *dyrr*) hefur verið eytt við útjöfnun: þar verður *y*, rótarsérhljóð nf./þf. ft., ofan á og ryður út *u*, rótarsérhljóði þgf./ef. ft.

(30) Nefnifall lagt til grundvallar: rótarsérhljóðið *u* víkur fyrir *y*

<i>ft. nf.</i>	<i>dyrr</i>
<i>þf.</i>	<i>dyrr</i>
<i>þgf.</i>	<i>durum</i> → <i>dyrum</i>
<i>ef.</i>	<i>dura</i> → <i>dyra</i>

Mańczak (1958:388–401) og Tiersma (1982) hafa þó bent á að í örnefnum geti þær fallmyndir er tákna dvöl á stað og hreyfingu til staðar og frá verið *ómarkaðar*; þetta er þá önnur tegund af sérmörkun:

(31) Sérmörkun (e. *local markedness*), sbr. Tiersma 1982:843

Í örnefnum geta þær fallmyndir sem tákna dvöl á stað og hreyfingu til staðar og frá verið *ómarkaðar*.

Í íslensku eru það fyrst og fremst *þágufall* og *eignarfall* sem þjóna þessu hlutverki. Þetta sást glögglega á dæmum þeim sem áður voru rædd um örnefnið *at Gásu* í (20)–(23): flest dæmin eru um þágufallsmyndina *Gásu*, næstflest um eignarfallsmyndina *Gása* en nefnifall og þolfall koma ekki fyrir í elstu heimildum. Tölur í *Íslenskri orðtíðnibók* (Jörgen Pind o.fl. 1991:1156–57) sem sýna tíðni einstakra falla í safni 500.000 lesmálsaða úr 100 ólíkum textum gefa sömu mynd. Þar kemur fram að í örnefnum er þágufall mest notað (32a) en í öðrum nafnorðum (báðum tölum og öllum kynjum) er nefnifallið algengast.

(32) Tíðni einstakra falla örnefna og nafnorða eftir *Íslenskri orðtíðnibók*

	a.	b.
	<i>örnefni</i>	<i>nafnorð</i>
<i>nefnifall</i>	10,4%	<b>31,2%</b>
<i>þolfall</i>	16,4%	27,9%
<i>þágufall</i>	<b>56,2%</b>	29,0%
<i>eignarfall</i>	17,1%	11,9%

Ef marka má þessar tölur um tíðni einstakra fallmynda í nútímamáli eru íslensk örnefni langmest notuð í þágufalli og því getur það talist *ómarkað* gagnvart öðrum föllum (sjá einnig umræðu hjá Haraldi Bernharðssyni 2004:25–32). Þetta þýðir með öðrum orðum að barn

á máltökuskeiði heyrir langoftast þágufallsmynd örnefna, eins og til dæmis (*at/á/frá*) *Gásu*m, og því verður þágufallsmyndin, *Gásu*m, grunnmynd örnefnisins í huga barnsins en ekki nefnifallsmyndin *Gæs*(s).

Áhrif sérmörkunar af þessu tagi í örnefnum veldur því að í örnefnum sjást breytingar sem allajafna sjást ekki hjá venjulegum nafnorðum: vensl einstakra fallmynda eru annars eðlis í örnefnum en í venjulegum nafnorðum. Þegar þágufallsmyndin (*at/á/frá*) *Gásu*m er lögð til grundvallar kemur ekki á óvart að rótarsérhljóðið *á* breiðist út á kostnað rótarsérhljóðsins *æ*, eins og sýnt var í (27b) og (28a). Í samnafninu var nefnifall fleirtölu aftur á móti lagt til grundvallar, eins og áður var rætt, og því breiðist rótarsérhljóð þess, *æ*, út á kostnað *á* í samnafninu, sbr. (27a).

Þágufallsmyndin (*at/á/frá*) *Gásu*m er óræð hvað kyn og beygingarflokk varðar því að sama beygingarmyndanið, *-um*, er notað í öllum kynjum og öllum beygingarflokkum í íslensku. Barn á máltökuskeiði getur því ekki auðveldlega greint kyn eða beygingarflokk örnefnisins á grundvelli þágufallsmyndarinnar (*at/á/frá*) *Gásu*m og í örnefnum er sjaldnast hægt að reida sig á ákvæðisorð (eða greini) sem tekið gætu af vafa um málfræðilegt kyn. Leiðin er því greið fyrir umtúlkun af ýmsu tagi: örnefnið getur hæglega skipt um kyn (orðið karlkyns frekar en kvenkyns, sbr. 28c) og tekið beygingu eftir fleiri en einum beygingarflokki (fengið nf. *-ir* eða *-ar* og þf. *-i(r)* eða *-a(r)*, sbr. 28b). Þess vegna verða til nf. ft. *Gásir*, *Gásar* og þf. *Gásir*, *Gásar* eða jafnvel *Gási*, *Gása*. Þróun af því tagi má sjá hjá nokkrum fjölda örnefna og vísast til frekari umræðu um það hjá Haraldri Bernharðssyni 2004.

## 5 Lokaorð

Hér hefur verið borin saman þróun samheitisins físl. *gós*, *gás*, ft. *géss* og örnefnisins sem birtist í verslunarstað þeim sem nefndist í fornu máli *at Gásu*m (*Gósu*m) en samheitið og örnefnið hafa þróast hvort í sína átt. Í báðum tilvikum er um að ræða svokallaða *sérmörkun* (e. *local markedness*) en þó með ólíkum hætti. Í samheitinu felst sérmörkunin í því að fleirtalan er ómörkuð andspænis eintölunni og því er fleirtalan lögð til grundvallar við áhrifsbreytingar á eintölumyndunum; þannig fær fleirtalan nýja endingu (*-ir* í stað *-Ø*), eintalan þiggur rótarsérhljóð frá nf./þf. fleirtölu og þróast á annan veg en til dæmis orðin *mús* og *lús* sem upprunalega höfðu sömu beygingu og físl. *gás* en aftur á móti

fleirtölu sem var mörkuð gagnvart eintölunni. Í örnefnum felst sér-mörkunin í því að þágufall fleirtölu er ómarkað andspænis nefnifalli og þolfalli og því fá nefnifall og þolfall rótarsérhljóð þágufallsins.

## Ritaskrá

- Annálar 1400–1800* 1–8. 1922–2002. Reykjavík: Hið íslenska bókmenta-félag.
- van Arkel-de Leeuw van Weenen, Andrea. 1987. *Möðruvallabók* 1. Index and Concordance. 2. Text. Leiden: E.J. Brill.
- Árni Magnússon og Páll Vídalín. 1943. *Jarðabók* 10. [Útg. Jakob Benediktsson.] Kaupmannahöfn: Hið íslenska fræðafjelag.
- Ásgeir Blöndal Magnússon. 1989. *Íslensk orðsifjabók*. Reykjavík: Orða-bók Háskólans.
- Ásta Svavarsdóttir. 1993. *Beygingakerfi nafnorða í nútímaíslensku*. Mál-fræðirannsóknir 5. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Bandle, Oskar. 1956. *Die Sprache der Guðbrandsbiblía*. Orthographie und Laute, Formen. Bibliotheca Arnamagnæana 17. Kopenhagen: Ejnar Munksgaard.
- Bertelsen, Henrik (útg.). 1905–11. *Piðriks saga af Bern* 1–2. Samfund til udgivelse af gammel nordisk litteratur 34. København.
- Biskupa sögur* 1–2. 1858–78. Kaupmannahöfn: Hið íslenska bókmenta-félag.
- Bjorvand, Harald. 1972. Zu den altwestnordischen Pluralendungen -ar, -ir und -r bei femininen Substantiva. *Norwegian Journal of Linguistics — Norsk Tidsskrift for Sprogvidenskap* 26:195–215.
- Bjorvand, Harald, og Fredrik Otto Lindeman. 2000. *Våre arveord*. Etymologisk ordbok. Oslo: Novus forlag.
- Björn Halldórsson. 1992. *Orðabók*. Íslensk — latnesk — dönsk... Fyrst gefin út árið 1814 af Rasmusi Kristjáni Rask. Jón Aðalsteinn Jónsson sá um útgáfuna. Orðfræðirit fyrri alda 2. Reykjavík: Orðabók Háskólans.
- Björn Sigfússon (útg.). 1940. *Ljósvetninga saga*. Íslenzk fornrit 10. Reykjavík: Hið íslenska fornritafélag.
- Björn Vigfússon. 2002. Gásakaupstaður í íslenskum fornritum. *Súllur* 28 [nr. 41]:37–60.
- Björn Þorsteinsson og Guðrún Ása Grímsdóttir. 1989. Norska öldin —

- með viðaukum eftir Sigurð Líndal. Sigurður Líndal (ritstj.): *Saga Íslands* 4:59–258. Reykjavík: Hið íslenska bókmenntafélag, Sögufélag.
- Björn K. Þórólfsson. 1925. *Um íslenskar orðmyndir á 14. og 15. öld og breytingar þeirra úr fornmalinu*. Með viðauka um nýjungar í orðmyndum á 16. öld og síðar. Reykjavík. [Endurprentun: Rit um íslenska málfræði 2. 1987. Reykjavík: Málvísindastofnun Háskóla Íslands.]
- Brunner, Karl. 1965. *Altenglische Grammatik*. Nach der angelsächsischen Grammatik von Eduard Sievers. Dritte, neubearbeitete Auflage. Sammlung kurzer Grammatiken germanischer Dialekte A3. Tübingen: Max Niemeyer Verlag.
- Bybee, Joan L. 1985. *Morphology*. A Study of the Relation Between Meaning and Form. Typological Studies in Language 9. Amsterdam: John Benjamins.
- Bybee, Joan L., og Mary Alexandra Brewer. 1980. Explanation in morphophonemics: changes in Provençal and Spanish preterite forms. *Lingua* 52:201–42.
- Byggðir Eyjafjarðar 1990* 1–2. 1993. Ritstj. Guðmundur Steindórsson, Jóhannes Sigvaldason og Kristján Sigfússon. Búnaðarsamband Eyjafjarðar [Akureyri].
- Campbell, A. 1959. *Old English Grammar*. Oxford: At the Clarendon Press.
- Christophersen, Axel, og Audun Dybdahl (ritstj.). 1999. *Gásir — en internasjonal handelsplass i Nord-Atlanteren*. Trondheim: Tapir. *DI = Diplomatarium Islandicum*. Sjá Íslenzkt fornbréfasafn.
- Einar Ól. Sveinsson (útg.). 1939. *Vatnsdæla saga*. Íslenzk fornrit 8. Reykjavík: Hið íslenska fornritafélag.
- Eiríkur Rögnvaldsson. 1990. *Íslensk orðhlutafræði*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Eiríkur Þormóðsson og Guðrún Ása Grímsdóttir (útg.). 2003. *Oddaannálar og Oddaverjaannáll*. Rit 59. Reykjavík: Stofnun Árna Magnússonar á Íslandi.
- Finnur Jónsson. 1908. Hinn forni kaupstaður „at Gásum“. *Árbók Hins íslenska fornleifafélags* 1908:3–8.
- Foote, Peter (útg.). 2003a. *Jóns saga Hólabyskups ens helga*. Editiones Arnarnagæanæ A14. Copenhagen: C.A. Reitzels Forlag.
- Foote, Peter (útg.). 2003b. Sjá Sigurgeir Steingrímsson o.fl.

- Føroysk orðabók*. 1998. [Ritstj. Jóhan Hendrik W. Poulsen o.fl.] Tórs-havn: Føroya Fróðskaparfelag.
- Greenberg, Joseph H. 1966. *Language Universals with Special Reference to Feature Hierarchies*. Janua Linguarum, Series Minor 59. The Hague: Mouton.
- Guðmundur Andrússon. 1999. *Lexicon Islandicum*. Orðabók Guðmundar Andrússonar. Ný útgáfa. Gunnlaugur Ingólfsson og Jakob Benediktsson önnuðust útgáfuna. Orðfræðirit fyrri alda 4. Reykjavík: Orðabók Háskólans.
- Guðni Jónsson (útg.). 1936. *Grettis saga Ásmundarsonar*. Íslenzk fornrit 7. Reykjavík: Hið íslenska fornritafélag.
- Guðrún Ása Grímsdóttir (útg.). 1998. *Biskupa sögur 3*. Íslenzk fornrit 17. Reykjavík: Hið íslenska fornritafélag.
- Guðrún Nordal. 1992. Sagnarit um innlend efni — Sturlunga saga. Vésteinn Ólason (ritstj.): *Íslensk bókmenntasaga 1:309–44*. Reykjavík: Mál og menning.
- Hagland, Jan Ragnar. 1999. Om Gásar i mellomalderens litterære kjelder. Axel Christophersen og Audun Dybdahl (ritstj.): *Gásir — en internasjonal handelsplass i Nord-Atlanteren*, bls. 95–102. Trondheim: Tapir.
- Haraldur Bernharðsson. 2004. Um Moldhaugnaháls út í Fjósa og Fjörður. Af áhrifsbreytingum í nokkrum fleirtöluörnefnum. *Íslenskt mál* 26:11–48.
- Hooper, Joan Bybee. 1980. Child morphology and morphophonemic change. Jacek Fisiak (ritstj.): *Historical Morphology*, bls. 157–87. Trends in Linguistics. Studies and Monographs 17. The Hague: Mouton.
- Hreinn Benediktsson. 1959. The Vowel System of Icelandic: A Survey of Its History. *Word* 15:282–312. [Endurprentun: Hreinn Benediktsson 2002:50–73.]
- Hreinn Benediktsson. 1965. *Early Icelandic Script as Illustrated in Vernacular Texts from the Twelfth and Thirteenth Centuries*. Íslenzk handrit, series in folio 2. Reykjavík: The Manuscript Institute of Iceland.
- Hreinn Benediktsson (útg.). 1972. *The First Grammatical Treatise*. University of Iceland Publications in Linguistics 1. Reykjavík: Institute of Nordic Linguistics.
- Hreinn Benediktsson. 1979. Relational Sound Change: ‘vá’ > ‘vo’ in Icelandic. Irmengard Rauch og Gerald F. Carr (ritstj.): *Linguistic*

- Method. Essays in Honor of Herbert Penzl*, bls. 307–26. The Hague, Paris, New York: Mouton Publishers. [Endurprentun: Hreinn Benediktsson 2002:227–42.]
- Hreinn Benediktsson. 2002. *Linguistic Studies, Historical and Comparative*. Ritstj. Guðrún Þórhallsdóttir, Höskuldur Þráinsson, Jón G. Friðjónsson og Kjartan Ottosson. Reykjavík: Institute of Linguistics.
- Höskuldur Þráinsson [= Þráinsson], Hjalmar P. Petersen, Jógvan í Lon Jacobsen, Zakaris Svabo Hansen. 2004. *Faroese. An Overview and Reference Grammar*. Tórshavn: Føroya fróðskaparfélag.
- Íslenzkt fornbréfasafn 1–. 1857–. Reykjavík og Kaupmannahöfn: Hið íslenzka bókmenntafélag.
- Johnsen, Oscar Albert, og Jón Helgason (útg.). 1941. *Saga Óláfs konungs hins helga — Den store saga om Olav den hellige* 1–2. Oslo: Norsk historisk kjeldeskrift-institutt.
- Jóhannes Halldórsson (útg.). 1959. *Kjalnesinga saga*. Íslenzk fornrit 14. Reykjavík: Hið íslenzka fornritafélag.
- Jón G. Friðjónsson. 1993. *Mergur málsins. Íslensk orðatiltæki, uppruni, saga og notkun*. Reykjavík: Örn og Örlygur.
- Jón Helgason. 1932. Nokkur íslenzk handrit frá 16. öld. *Skírnir* 106: 143–68.
- Jón Helgason (útg.). 1942. *Óláfs saga ens helga*. Corpus Codicum Islandicorum Medii Aevi 15. Copenhagen: Einar Munksgaard.
- Jón Magnússon. 1997. *Grammatica Islandica — Íslenzk málfræði*. Jón Axel Harðarson gaf út með inngangi, þýðingu og athugasemdum. Reykjavík: Málvísindastofnun Háskóla Íslands og Háskólaútgáfan.
- [Jón Þorkelsson (útg.).] 1915–33. *Búalög um verðlag og allskonar venjur í viðskiptum og búskap á Íslandi*. Sögurit 13. Reykjavík: Sögufélag.
- Jónas Kristjánsson (útg.). 1956. *Eyfirðinga sögur*. Íslenzk fornrit 9. Reykjavík: Hið íslenzka fornritafélag.
- Jörgen Pind (ritstj.), Friðrik Magnússon og Stefán Briem. 1991. *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.
- Katalog over den Arnamagnæanske håndskriftsamling* 1–2. 1889–94. [Kr. Kålund tók saman.] København: Kommissionen for det Arnamagnæanske Legat.
- [Kålund, Kr. (útg.).] 1906–11. *Sturlunga saga* 1–2. København.
- Kuryłowicz, Jerzy. 1945–49. La nature des proces dits „analogiques“. *Acta Linguistica* 5:15–37.

- Larsson, Ludvig. 1891. *Ordförrådet i de älsta isländska handskrifterna*. Leksikaliskt ock gramatiskt ordnat. Lund.
- de Leeuw van Weenen, Andrea (útg.). 1993. *The Icelandic Homily Book*. Perg. 15 4° in the Royal Library, Stockholm. Íslensk handrit, series in quarto 3. Reykjavík: Stofnun Árna Magnússonar á Íslandi.
- de Leeuw van Weenen, Andrea. 2004. *Lemmatized Index to the Icelandic Homily Book*. Perg. 15 4° in the Royal Library, Stockholm. Rit 61. Reykjavík: Stofnun Árna Magnússonar á Íslandi.
- Lexicon poeticum antiquæ linguæ septentrionalis — Ordbog over det norsk-islandske skjaldesprog*. 1931. Oprindeligt forfattet af Sveinbjörn Egilsson . . . 2. udgave ved Finnur Jónsson. København.
- Loth, Agnete (útg.). 1964. *Thomasskinna*. Early Icelandic Manuscripts in Facsimile 6. Copenhagen: Rosenkilde and Bagger.
- Loth, Agnete (útg.). 1977. *Fornaldarsagas and Late Medieval Romances*. Early Icelandic Manuscripts in Facsimile 11. Copenhagen: Rosenkilde and Bagger.
- Mańczak, Witold. 1958. Tendances générales des changements analogiques. *Lingua* 7:298–325, 387–420.
- Margrét Hermannsdóttir. 1987. Fornleifarannsóknir að Gásam og víðar í Eyjafirði árið 1986. *Súllur* 14 [nr. 27]:3–39.
- Matthías Jochumsson. 1956–58. *Ljóðmæli* 1–2. Árni Kristjánsson sá um útgáfuna. Reykjavík: Ísafoldarprentsmiðja.
- Nilsson, Jan. 1975. *Plurala ortnamn på Island*. Morfologiska iakttagelser. Umeå Studies in the Humanities 8. Umeå.
- Noreen, Adolf. 1923. *Altnordische Grammatik* 1. Altisländische und altnorwegische Grammatik (Laut- und Flexionslehre) unter Berücksichtigung des Urnordischen. Vierte vollständig umgearbeitete Auflage. Sammlung kurzer Grammatiken germanischer Dialekte 4. Halle (Saale): Verlag von Max Niemeyer.
- Nucleus Latinitatis*. 1994. Ný útgáfa. Guðrún Kvaran og Friðrik Magnússon sá um útgáfuna. Orðfræðirit fyrri alda 3. Reykjavík: Orðabók Háskólans.
- ONP Registre = Ordbog over det norrøne prosasprog. Registre*. A Dictionary of Old Norse Prose. Indices. 1989. København: Den arnamagnæanske kommission.
- Ólafur Halldórsson (útg.). 1904. *Jónsbók*. Kong Magnus Hakonssons Lovbog for Island vedtaget paa Altinget 1281. København.



- Ólafur Lárusson (útg.). 1934. *Lögbók Íslendinga. Jónsbók 1578*. Monumenta Typographica Islandica 3. Copenhagen: Levin & Munksgaard.
- Ólafur Lárusson (útg.). 1936. *Staðarhólsbók. The Ancient Lawbooks Grágás and Járnsíða*. Corpus Codicum Islandicorum Medii Aevi 9. Copenhagen: Levin & Munksgaard.
- Páll Eggert Ólason. 1944. *Sextánda öldin*. Höfuðþættir. Saga Íslendinga 4. Reykjavík: Menntamálaráð og Þjóðvinafélag.
- Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch* 1–2. Tübingen: Francke Verlag.
- Runólfur Jónsson [= Runolphus Jonas]. 1651. Recentissima antiqvissimæ linguæ septentrionalis incunabula, Id est *Grammaticæ Islandicæ rudimenta*. Nunc primum adornari cæpta & edita per Runolphum Jonam Islandum. Hafniae.
- Sigurgeir Steingrímsson, Ólafur Halldórsson og Peter Foote (útg.). 2003. *Biskupa sögur* 1. Íslensk fornrit 15. Reykjavík: Hið íslenska fornritafélag.
- Stefán Karlsson (útg.). 1983. *Guðmundar sögur biskups* 1. Editiones Arnarnagænanæ B6. Kaupmannahöfn: C.A. Reitzels Forlag.
- Storm, Gustav (útg.). 1888. *Islandske Annaler indtil 1578*. Christiania.
- Tiersma, Peter Meijes. 1982. Local and general markedness. *Language* 58:832–49.
- de Vries, Jan. 1962. *Altnordisches etymologisches Wörterbuch*. Zweite verbesserte Auflage. Leiden: E.J. Brill.
- Österholm, Nils (útg.). 1987. *Þorleifs þáttr jarlaskálds*. Handskriftsstudier i en islandsk kortsaga. Stockholm: Almqvist & Wiksell International.

## Abstract

In this paper the phonological and morphological development of Old Icelandic *gós*, plur. *géss* ‘goose’ down to Modern Icelandic is examined and compared with the development of the place name that appears as *at Gásum* (< *Gósum* dat. plur.) in the earliest Old Icelandic sources. It is commonly assumed that the place name is the plural of *gós* and that in Old Icelandic the plural of the common noun and the place name were identical. On their way down to Modern Icelandic, however, they have developed in two different directions.

Section 2 describes the phonological prehistory of Old Icelandic *gós*. The comparative evidence, including the cognate Old High German *gans*, suggests Old Icelandic *gós* contained a nasalized root vowel. Following the general merger of nasalized *ǫ* and *ó* in Old Icelandic, *gós* should therefore have become *gós*. The form *gós*, however, is not attested. Instead, *gós* became *gás*, which cannot be the phonological result of *gós*, but must have arisen through analogical restoration of the low back vowel *ǫ* under the influence of words with the non-nasal *ǫ*.

Old Icelandic *gós*, plur. *géss* 'goose', which originally belonged to the same inflectional class as *mús*, plur. *mýss* 'mouse' and *lús*, plur. *lýss* 'louse', has, in the course of its development down to Modern Icelandic, undergone two changes that distinguish it from *mús* and *lús*, as described in section 3. On the one hand, the nom./acc. plur. *géss*, phonologically becoming *gæss* and later *gæs*, acquired a new plural morpheme, becoming *gæsir*. On the other hand, the root vowel *æ* was generalized throughout the paradigm, replacing *á* (< *ǫ*). Both changes make their earliest appearance in 16th century written records, and the evidence suggests the addition of the plural marker *-ir* preceded the generalization of the root vowel *æ*, albeit not by much (§3.1).

The question is bound to arise why the morphologically similar *mús*, plur. *mýs* (< *mýss*) and *lús*, plur. *lýs* (< *lýss*) did not undergo the same development. In section 3.2 it is argued that the changes in the word for goose can be attributed to *local markedness*, as defined by Tiersma (1982): due to the fact that geese tend to appear in flocks, the plural forms of the word for goose are more frequently used than the singular forms; therefore, contrary to general markedness, the plural of the word for goose is unmarked and the singular is marked. The unmarked nom./acc. plur. *gæs* (< *gæss* < *géss*) therefore serves as a basic form in language acquisition (instead of the nom. sing. *gás* from earlier *gós*). When plural forms serve as basic forms they are, as demonstrated by Tiersma (1982:837–39), prone to under-analysis whereby the original plural marker is regarded as part of the stem, which, in turn, calls for the addition of a new plural marker. Thus the nom./acc. plur. *gæs*, which was formally distinguished from the nom./acc./dat. sing. *gás* only by the root vowel alternation, acquired a new plural marker, becoming *gæsir*. The paradigmatic generalization of the root vowel *æ* (at the expense of *á*) also follows from the fact that the nom./acc. plur. *gæs*, *gæsir* was the unmarked form. In the words for *mús* and *lús*, by

contrast, the singular is unmarked and the plural is marked (general markedness); therefore these words did not undergo the same development as the word for goose.

As discussed in section 4.2, the place name appears in nom./acc. plur. as *Gásir* already in the fifteenth century and in later sources the form *Gásar* is also in evidence, as well as the presumably masculine acc. plur. *Gási* and *Gæsi*. The development of the place name therefore differs from the development of the common noun on at least three points: (1) the root vowel *á* has been generalized, replacing *æ* in nom./acc. plur.; (2) the place name not only adds the new plural morpheme *-ir* but also *-ar*; and (3) the place name also can appear as masculine. These changes, it is argued, also are due to local markedness. Typically the nominative case is unmarked and the oblique cases marked. As discussed by Mańczak (1958:388–401) and Tiersma (1982:843), place names, however, frequently are unmarked in the locative case; in Icelandic this is the dative case. Accordingly, the dat. plur. *Gásum* is the basic form in the paradigm of the place name. This explains why, in the place name, the root vowel *á* has been extended (from the dative) throughout the paradigm. Moreover, the dative form is indeterminate with respect to gender and inflectional class, since in Icelandic the dat. plur. has the same morpheme, *-um*, in all three genders, across all inflectional classes. Therefore, the ambiguous dat. plur. form *Gásum* is prone to reinterpretation with respect to gender and inflectional class. This explains why it has sometimes been reanalyzed as masculine, as well as why it has shifted between inflectional classes, as is manifest in the addition of the nom./acc. plur. morpheme *-ar*, beside *-ir*.

**Keywords:**

historical morphology, analogy, markedness, place names

*Haraldur Bernharðsson*

*Hugvísindastofnun*

*Nýja-Garði*

*Háskóla Íslands*

*IS-101 Reykjavík, ÍSLAND*

*haraldr@hi.is*



Katrín Axelsdóttir

# Hvað er klukkan?

## 1 Inngangur

Að segja til um hvað klukkan er mun þykja hversdagsleg athöfn. En hún á sér að líkindum ekki ýkja langa sögu hér á landi meðal almennings. Um stundaklukkan er getið á biskupsstólunum, á Hólum á 16. öld og í Skálholti á 17. öld, og talið er að á sumum efnaheimilum hafi verið til stundaklukkan fyrir 1800. En klukkan fóru ekki að breiðast út fyrr en eftir miðja 19. öld. Svipaða sögu er að segja um vasaúr, en armbandsúr voru enn síðar á ferðinni. Þau voru varla til fyrr en eftir heimsstyrjöldina fyrri. Áður studdust menn við stöðu sólar og stjarna, eða þá sjávarföll, og sumir áttu stundaglös. Orðin sem notuð voru um tíma dagsins voru eyktamörkin (*óttu, miður morgunn, dagmál, hádegi, nón, miðaftann, náttmál* og *miðnætti*), enda skipti nákvæmur tími litlu máli í bændasamfélaginu.<sup>1</sup> Konráð Gíslason (1851:244) þýddi *Klokk-en to venter jeg ham* með 'jeg á von á honum einni stundu fyrir nón', en hugsanlega var hann hér að fyrna mál sitt, enda hreintungumaður. Kona sem fædd var í Húnavatnssýslu 1913 mundi aðeins eftir grófum tímaviðmiðum þegar hún var að alast upp.<sup>2</sup> Stefán Einarsson (1897–1973), sem var Austfirðingur, úr Breiðdal, segir (1949:204) að öll eyktaheiti nema *óttu* og e.t.v. *rismál* og *dagmál* séu algeng í sveitum á Ís-

---

<sup>1</sup>Um tímamæla og tímaviðmið, sjá Jónas Jónasson 1961:1–2 og Árna Björnsson 1990:82–91, og rit sem hann vitnar til.

<sup>2</sup>Herborg Gestsdóttir (1913–2005), samtal.

landi.<sup>3</sup> En vitanlega hafa þessi gömlu tímaviðmið ekki lifað svona lengi í þéttbýli.

Hér verða fyrst skoðuð nokkur gömul dæmi, einkum úr ritmálsafni Orðabókar Háskólans, um orðalag þar sem vísað er til klukkunnar og þau borin saman við nútímamál eftir því sem ástæða er til (2). Þá verður sjónum beint að tveimur orðalagskerfum í nútímamáli um nákvæman tíma (3) og giskað á ástæður þess að hið yngra kerfi kom upp og vann á (4). Að lokum eru helstu atriði dregin saman.

## 2 Gömul dæmi um að vísað sé til klukku

Þótt klukkur hafi lengstum verið lítt þekktar hér á landi eru til nokkur býsna gömul dæmi um orðalag þar sem vísað er til stunda á klukku. Í (1) er dæmi úr Reykjahólabók, safni þýddra heilagramannasagna, frá því á fyrri hluta 16. aldar (*Reykjahólabók* I 1969:282):

- (1) sem klvckan mvndi vera a millvm.x. xi. en þat er a vora thavlv nærre hadeige

Tvö önnur álíka dæmi er að finna í þessu riti (*Reykjahólabók* I 1969:330, 338) og alls staðar þykir þýðandanum jafnframt ástæða til að orða tím-ann að hætti Íslendinga.

Í ritmálssafni Orðabókar Háskólans er að finna nokkur talsvert gömul dæmi, sbr. dæmin í (2), en þau eru frá 16. og 17. öld:

- (2) a. A mornnana fra þuj kluckann slær sex (DI X, 217; 16. öld)  
 b. um það bil klukkan var 4 (JÓIInd I, 41; mið 17. öld)  
 c. þa er Kluckann 3 epter Middag (ÞPEnchir E VIIr; 1671)  
 d. þa Solen er j fullu Austre Kluckan 6 fyrer Middag (ÞPCal, 96; 1692)  
 e. Þegar Klukkann slær Eitt (BeerFerðam I, Ir; 1694)  
 f. einni stundu eftir dagmál, þegar klukkan er tíu (Alþb VIII, 494; 1695)

Dæmi (2a) er í íslenskri þýðingu á kirkjuordinansíu Kristjáns III. en þar er fjallað um skólahald.<sup>4</sup> Dæmi (2b) er í ævisögu Jóns Ólafssonar

<sup>3</sup>*Rismál* 'fótaferðatími, um kl. 6' eru vanalega ekki talin til eyktaheita þótt Stefán geri það.

<sup>4</sup>Mörg álíka dæmi koma í kjölfar þessa (DI X 1911–1921:217–218).

Indíafara og á eftir því kemur eins og skýring: *það er stund af nóni*.<sup>5</sup> Dæmi (2c–d) eru í ritum um tímataalsfræði og því ekki óvænt. Dæmi (2e) er í þýðingu og dæmi (2f) er í texta eftir Dana.<sup>6</sup> Öll þessi dæmi, og önnur dæmi þessum ritum (sbr. nmgr. 4 og 5) eiga væntanlega rætur í erlendri menningu.

Elstu dæmin í ritmálssafninu um að vísað sé til nákvæmari tíma en heillar stundar eru frá 17. og 18. öld. Í þeim er vísað til hálf tíma og kortera.

- (3) a. Kluckann 1 og 2 Qvarter (ÞÞCal, 97; 1692)
- b. játar þar, að klukkan sé þá eitt og 2 kvarter. (PVíd-Skýr, 73; fyrsti þriðjungur 18. aldar)
- c. er sú eyktin [...] mið á nóni, en byrjast klukkan hálf-gengin 2 (PVídSkýr, 57; fyrsti þriðjungur 18. aldar)
- d. Klukkan hjer um hálf-gengin eður rjettara þrjú kvarter til sjö (JÓIGrv. ÁMSkr. I 2, 32; mið 18. öld)<sup>7</sup>

Í (3a–b) er talað um að klukkan sé 1.30 og orðalagið er ólíkt því sem tíðkast í nútímamáli. Þetta minnir á hvernig tekið er til orða í ýms-um erlendum málum, t.d. í frönsku: *une heure et demie* ‘klukkan eitt og helmingi betur’. Í (3c) er talað um að klukkan sé *hálf-gengin 2*, án forsetningar, og í (3d) er svo forsetningin *til*. Þetta er ólíkt nútímamáli því að þar er yfirleitt forsetningin *í* höfð á eftir *gengin*, sbr. 3 hér á eftir.<sup>8</sup> Þar að auki er *gengin* aldrei notað með hálfá tímanum í nútímamáli,

<sup>5</sup>Fleiri dæmi eru um klukkuna í ævisögu Jóns Indíafara, t.d. *til þess að klukkan slær 10 fyrir miðdag; um þriðju stund eftir miðdag; nær þriðju stund dags eftir miðdegi* (Æfisaga Jóns Ólafssonar Indíafara 1908–1909:293, 302, 303).

<sup>6</sup>Dæmið er í reglum sem Christian Möller amtmaður, með aðstoð íslenskra lög-  
manna, tók saman. Athyglisvert er að oft er vísað til tíma dagsins í þessari málsgrein en aðeins einu sinni til klukkunnar: „Til lögréttu skal hringjast í fyrsta sinn *að miðjum morgni*, í annað sinn *á dagmálum*, ... Í þriðja sinn skal hringjast *einni stundu eftir dagmál, þegar klukkan er 10*, en þá skulu allir innnefndir lögréttumenn að lögréttu komnir vera. ... Síðan trakterist þau mál, sem þar koma, *inn til miðmunda*. Á *miðmunda* gangi menn að tjöldum til máltíðar. Á *nóni* sé hringt eitt sinn til lögréttu, og gangi menn þá þangað tafarlaust ... og siti þar málum að gegna *til náttmála*.“ (Alþingisbækur Íslands VIII 1949–1955:493–494).

<sup>7</sup>Dæmið er úr ævisögu Árna Magnússonar eftir Jón Ólafsson úr Grunnavík og þar er að finna mörg önnur dæmi. Í dagbókarskrifum Jóns og lýsingu hans á brunanum í Kaupmannahöfn 1728 er mjög oft vísað til klukku (sbr. Jón Ólafsson 2005) svo að það hefur verið Jóni tamt.

<sup>8</sup>Í nýrri útgáfu á skrifum Jóns Ólafssonar úr Grunnavík hefur á einum stað verið bætt við forsetningunni *í* á eftir *gengin* (Jón Ólafsson 2005:70), en það er óþarfi.

aðeins með korteri og mínútum. Nú er alltaf sagt *klukkan er hálf tvö, hálf þrjú* o.s.frv.

Dæmin í (4) eru líklega tvíræð í augum flestra nútímamanna, sem gætu væntanlega bæði skilið (4a) sem 7.15 og 7.45, og (4b) sem 12.15 og 12.45:

- (4) a. Kl. eitt korter til 8 kom litla Sigga inn (SkrifStap; mið 19. öld)
- b. kl. kortér í eitt (Fjallk 1890, auk 8)

En dæmið í (5), sem er frá sama tíma, er hafið yfir allan vafa; þar merkir *korter í eitt* greinilega 12.15:

- (5) Ég skildi við föðurbróður minn kl. kortér í eitt, enn kl. hálf-eitt fanst hann myrtr. (Fjallk 1890, auk 11)

Dæmin í (4) og (5) virðast vera styttingar úr *gengin til* og *gengin í*. Í málinu á þessum tíma hefur væntanlega verið venja að nota *vantar* . . . í á undan heila tímanum, þannig að þessar styttingar hafa ekki skapað neina tvíræðni.

Dæmið í (5) er elsta dæmið í ritmálssafninu um orðalagið *klukkan hálf<eitt>*; í eldri dæmum var haft annað orðalag í slíkum tilvikum, sbr. dæmin í (3). Þegar dæmi eru jafn fá og þau sem hér hafa verið skoðuð er auðvitað erfitt að fullyrða hvaða orðalag er eldra en annað. En ekki er ólíklegt að *klukkan er hálf<eitt>* sé stytting úr *klukkan er hálf gengin <eitt>*. Aldur dæma (3c) og (5) kemur líka heim við það.

Í (6) er elsta dæmið um orðalagið *yfir <eitt>*. Merkingin er sú sama og í nútímamáli og þetta er sama merking og sjá má í orðalaginu *klukkan korter til* í (4a) og *klukkan korter í* í (4b) og (5):

- (6) Hún [ɔ: klukkan] er eitt „kortjer“ yfir átta. [Neðanmáls:] Á hreinni íslenzku væri þetta: Það lifa þrír fjórðungar hinnar níundu stundar (HBriemEnsk, 135, 1875)

Eins og sjá má er höfundurinn ekki alls kostar sáttur við orðalagið *yfir <eitt>*. Hann nefnir þó hvorki *til* né *í* sem möguleika, en stingur upp á mun flóknara og formlegra orðalagi. Hugsanlega er *yfir* yngra en *til* og *í*; það gæti verið komið í málið vegna danskra áhrifa. Á dönsku er nú sagt *over <et>* og var vísast einnig á 19. öld.<sup>9</sup> Þetta gæti varpað ljósi

<sup>9</sup>Í *Ordbog over det danske sprog* 1928 X:645, 1929 XI:897, 1933 XIV:130 eru ýmis dæmi um *over <et>*.



á andúð höfundarins á orðalaginu, enda talar hann í næstu andrá um hreina íslensku.<sup>10</sup> Við þetta er því að bæta að annað orðalag um klukkuna er áreiðanlega einnig komið úr dönsku, s.s. *gengin (til), vantar... í*.<sup>11</sup> Þegar klukkur halda innreið sína á Íslandi eru dönsk áhrif allsráðandi og ekki að undra að orðalagið sé sótt í dönsku.

Að lokum er ástæða til að geta þess að í orðabók Jóns Árnasonar frá miðri 18. öld eru tvær þýðingar gefnar á *quota est hora*:

- (7) hvad er frammorded, hvad er Kluckann (JÁNucl, 1371; mið 18. öld)

Fyrri orðalagið er sjaldan eða aldrei notað lengur. Konráð Gíslason (1851:244) hefur aðeins ‘hvað er framorðið?’ sem þýðingu á *Hvad er Klokken* í orðabók sinni, en kannski var hann að reyna að sneiða hjá því orðalagi sem líktist dönsku mest.

### 3 Tvö orðalagskerfi um nákvæman tíma

Fyrir nokkrum áratugum voru samböndin í *töflu 1* jafnan notuð við að segja til um nákvæman tíma:

Tafla 1: Orðalag um nákvæman tíma fyrir nokkrum áratugum

Eftir heila tímann t.d. 12.05	klukkan er <fimm> mínútur gengin í <eitt> klukkan er <fimm> mínútur yfir <tólf>
Fyrir heila tímann t.d. 11.55	klukkuna vantar <fimm> mínútur í <tólf>

<sup>10</sup>En auðvitað er allt eins sennilegt að hann sé aðallega að amast við orðinu *korter*, ekki orðalaginu *yfir <eitt>*.

<sup>11</sup>Dæmi eru um orðalagið *halvgaen <et>*, *klokken gaar til <et>*, *klokken mangler <to> minutter i <et>*, (*Ordbog over det danske sprog* 1928 X:646). Jafnframt eru til dæmi sambærileg dæmunum í (4): „et kvarter til ell. paa eet ... et kvarter over tolv“ *Ordbog over det danske sprog* 1929 XI:897.

Þannig tíðkaðist tvenns konar orðalag eftir heila tímann, *gengin í og yfir*.<sup>12</sup> Hvort tveggja virðist hafa verið nokkuð jafngilt.<sup>13</sup> Fyrir heila tímann var notað orðalagið *vantar... í*.<sup>14</sup>

Sem kunnugt er hefur þessi notkun breyst á síðustu áratugum. Líklega eru þau börn fá sem nota samböndin *klukkan er <fimm> mínútur gengin í <tólf>* og *klukkuna vantar <fimm> mínútur í <tólf>*. Yngstu kynslóðirnar nota jafnan *klukkan er <fimm> mínútur yfir <tólf>* og *klukkan er <fimm> mínútur í <tólf>*. *Klukkan er <fimm> mínútur yfir <tólf>* skilja allir, enda ótvírætt. Hins vegar getur sambandið *klukkan er <fimm> mínútur í <tólf>* misskilist; roskið fólk áttar sig oft ekki á því hvort þetta merki 'gengin' eða 'vantar'. Orðalagið er varla viðurkennt og það er ekki að finna í uppflettiritum.<sup>15</sup> Samt má búast við því að *yfir* og *í* verði ofan á að lokum en hin eldri, *gengin í* og *vantar... í*, hverfi smám saman úr daglegu tali, þótt þau eigi vísast eftir að heyrast í þularkynningum Ríkisútvarpsins enn um sinn.<sup>16</sup> Í *töflu 2* eru þessi tvö kerfi sýnd og kölluð „gamla“ og „nýja“ kerfið.

<sup>12</sup>Nú tíðkast nær eingöngu afbrigðið *gengin í*. En *gengin til* og *gengin*, án forsetningar, þekktust einnig og kannski eru þau afbrigði eldri, sbr. dæmin í (3) hér að framan. Þessi afbrigði þekkjast þó enn á 20. öld, sbr. Stefán Einarsson (1949:204): „hún er langt gengin tvö, hún er kortér gengin (til) tvö, 10 mínútur gengin tvö, kortér gengin tvö.“ Og enn er sagt: *klukkan er farin að ganga <eitt>*, án forsetningar.

<sup>13</sup>Óformleg könnun meðal nokkurra Íslendinga fæddra fyrir miðja 20. öld leiddi eftirfarandi í ljós: Flestir nota *gengin í* og *yfir* jöfnum höndum. Sumir telja *gengin í* og *yfir* jafngilt, aðrir fullyrða að *gengin í* hafi þótt betra mál en *yfir*. Þetta getur vel staðist ef lítið er til þess að það sem er eldra er jafnan talið betra frá sjónarhóli málvöndunarmanna. Elsta dæmið um *yfir* hér að framan, dæmi (6), er frá 1875 en eldri dæmi voru um *gengin*, sbr. dæmi (3c–d) sem bæði eru frá 18. öld. Hvort tveggja orðalagið er þó að líkindum úr dönsku eins og fram hefur komið.

<sup>14</sup>Orðalagið *<tvær> mínútur fyrir <eitt>* tíðkaðist einnig, og tíðkast enn. Það virðist hins vegar dálítið sérhæfðara og ýmislegt óljóst um notkun þess. Hægt er að segja *hittumst klukkan tíu mínútur fyrir eitt* en varla *\*klukkan er tíu mínútur fyrir eitt*. Hugsanlega er á reiki hvort notað er þolfall eða þágufall (*<tvær/toeim> mínútur/mínútum fyrir*). Í ritmálssafninu er dæmi um þágufall: „stundarfjórðungi fyrir klukkan tólf lentum við í vörinni“ (GhagalMaríum, 227). En líklega segja menn frekar *hittumst korter fyrir eitt* en *\*hittumst korteri fyrir eitt*.

<sup>15</sup>Reyndar er fremur lítið um orðalag tengt klukkunni að finna í orðabókum. *Yfir <eitt>* er hvorki að finna í Orðabók Blöndals (1920–1924) né *Íslenskri orðabók* 2002. *Ganga (klukkan er að ganga (er langt gengin) 5; klukkan er farin að ganga sex)* er í báðum þessum bókum, en *vanta (klukkuna vantar fimm mínútur í tvö)* er aðeins í þeirri síðarnefndu. Hjá Jóni Hilmari Jónssyni 2001 og 2005 er *yfir, gengin í* og *vantar... í*.

<sup>16</sup>Rétt er að geta þess að margir nota orðalagið *klukkan er farin að ganga <eitt>* þótt þeir noti annars aldrei *gengin í*.

Tafla 2: Orðalag um nákvæman tíma um þessar mundir

	„Gamla“ kerfið	„Nýja“ kerfið
Eftir heila tímann t.d. 12.05	klukkan er <fimm> mínútur gengin í <eitt> klukkan er <fimm> mínútur yfir <tólf>	klukkan er <fimm> mínútur yfir <tólf>
Fyrir heila tímann t.d. 11.55	klukkuna vantar <fimm> mínútur í <tólf>	klukkan er <fimm> mínútur í <tólf>

Ekki er vitað nákvæmlega hvenær gamla kerfið tekur að þoka fyrir hinu nýja, en það hefur líklega verið um miðja síðustu öld.<sup>17</sup> Hér verður sjónum ekki beint að tímanum heldur ástæðum breytingarinnar.

## 4 Ástæður breytingarinnar

Ósennilegt er að gamla kerfið hafi hörfað að ástæðulausu. Hér verða rædd nokkur atriði sem kunna að hafa valdið því að nýja kerfið varð til.

Í íslensku er hægt að miða bæði við heila tímann og hálfa tímann. Við getum t.d. sagt *klukkan er fimm mínútur yfir hálf(eitt)* og *klukkuna vantar fimm mínútur í hálf(eitt)*.<sup>18</sup> Þetta er hins vegar aðeins hægt með *yfir*, *í* og *vantar* ... *í*; ekki er hægt að nota *gengin í* við hálfa tímann. 12.35 væri *klukkan er þrjátíu og fimm mínútur gengin í eitt* eða þá *klukkuna vantar tuttugu og fimm mínútur í eitt*. Þetta eru nokkuð langar segðir því tölurnar eru svo háar. Hugsanlega er þarna að finna eina ástæðu þess að breytingin varð. Þegar miðað er við hálfa tímann má nota lægri og meðfærilegri tölur: *klukkan er fimm mínútur í hálf(eitt)*, *klukkuna vantar fimm mínútur í hálf(eitt)*, *klukkan er fimm mínútur yfir hálf(eitt)*.<sup>19</sup> Orðalag sem ekki var hægt að nota við hálfa tímann, þ.e. *gengin í*, varð e.t.v. óvinsælt fyrir vikið.

<sup>17</sup>Kona fædd 1944 man eftir því að hafa verið skömmuð fyrir að segja *klukkan er <tvær> mínútur í <eitt>* í merkingunni 'klukkuna vantar' um miðjan 6. áratuginn.

<sup>18</sup>Misjafnt er í tungumálum hvort hægt er að miða við hálfa tímann. Það er t.d. ekki hægt í ensku: \**five minutes past half eleven*. Það er hins vegar hægt í Norðurlandamálum: *fem minutter over halv elleve*.

<sup>19</sup>Eins og svigarnir sýna getur verið nóg að segja *hálf* (í stað *hálf(eitt)*, *hálf(tvö)* o.s.frv.) og segðin verður þá enn styttri.

Samkvæmt gamla kerfinu tíðkaðist tvenns konar orðalag eftir heila tímanna, *gengin í* og *yfir*. Hið síðarnefnda mátti einnig nota við hálfan tímann og það var því fjölhæfara. Þetta kann að hafa valdið því að *yfir* var tekið fram yfir *gengin í*. Einfaldara var að hafa eina aðferð en tvær, og líklegra er að sú einfaldari eða fjölhæfari sigri.

Þegar *gengin í* tekur að víkja verður *vantar... í* óþarft. Ástæðan er sú að nú er nóg að segja *í*; það er ekki tvírætt í máli þeirra sem nota aldrei *gengin í*. Þannig kemur að lokum upp ný staða, þ.e. nýja kerfið *yfir* og *í*.

Það sem hér hefur verið sagt um hugsanlega þróun má taka saman:

- Gamla kerfið: *gengin í/yfir* og *vantar... í*
- Hægt er að segja <fimm> *mínútur yfir* hálfan tímann
- *gengin í* viku
- *vantar... í* verður óþarft
- Nýja kerfið: *yfir* og *í*

Einnig getur verið að *gengin í* og *vantar... í* hafi látið undan síga því að þar er stundum krafist mun lengri setninga. Hægt er að segja *ég kem (klukkan) fimm mínútur yfir þrjú* og *ég kem (klukkan) fimm mínútur í þrjú*. Ef notuð eru samböndin *gengin í* og *vantar... í* ætti hins vegar að segja *ég kem þegar klukkan er fimm mínútur gengin í fjögur* og *ég kem þegar klukkuna vantar fimm mínútur í þrjú*, sem eru mun lengri segðir.<sup>20</sup> Þá kann að skipta máli að í „parinu“ *gengin í* og *vantar... í* er sögnin ýmist persónuleg eða ópersónuleg. Með *yfir* og *í* er alltaf persónuleg sögn og það var ef til vill einfaldara. Einnig kann að hafa flækt málin í huga fólks að ekki er miðað við sömu klukkustundina í *gengin í* og *vantar... í: klukkuna vantar fimm mínútur í þrjú* (2.55), en *klukkan er fimm mínútur gengin í fjögur* (3.05). *Yfir* og *í* vísa hins vegar til sama heila tímans: *klukkan er fimm mínútur í þrjú* (2.55) og *klukkan er fimm mínútur*

<sup>20</sup>Líklega segja margir *ég kem (klukkan) fimm mínútur gengin í fjögur* en setningafræðilega fer kannski illa á því þar sem sögnina *vera* vantar. Þessi háttur er samt hafður á í gömlum dæmum, sbr. (3c–d) hér að framan. Hins vegar gengur þetta alls ekki með sögninni *vanta*: \**ég kem klukkan vantar fimm mínútur í þrjú*; \**ég kem klukkuna vantar fimm mínútur í þrjú*.

*yfir þrjú* (3.05). Síðast en ekki síst er vert að nefna að kerfi sambærilegt nýja kerfinu tíðkast í dönsku: *over* og *i*.<sup>21</sup> Engin ástæða er til að útiloka áhrif úr þeirri átt. Eins og kom fram í 2 hér að framan er annað orðalag um klukkuna, s.s. *gengin (til)* og *vantar . . . í*, vísast einnig úr dönsku. Íslenskan hefur því einfaldlega haldið áfram að elta dönskuna.

Ef til vill hefur því margt lagst á eitt og valdið breytingum og sumt kann vitanlega að hafa skipt meira máli en annað. Þau atriði sem hér hafa verið nefnd sem *hugsanlegar* ástæður breytinganna verða að lokum tekin saman:

- Hálfí tíminn veldur því að *yfir* verður vinsælla en *gengin í*, og *vantar... í* verður óþarft.
- Langar setningar valda því að *gengin í* og *vantar... í* láta undan síga.
- Misræmi í persónu sagna veldur því að *gengin í* og *vantar... í* láta undan síga.
- *Gengin í* og *vantar... í* vísa ekki til sömu klukkustundar og hafa því átt erfitt uppdráttar.
- Danska kerfið, *over* og *i*, hefur haft áhrif á íslensku.

## 5 Niðurlag

Ekki er langt síðan klukkan urðu almenningsseign hér á landi og orðalag þeim tengt á sér því fremur stutta sögu. Hér hafa verið skoðuð ýmis gömul dæmi um orðalag þar sem vísað er til klukku og eins og við er að búast eiga þau elstu rætur að rekja til útlanda og útlendinga. Sum gömul dæmi um orðalag tengt klukkunni eru nokkuð frábrugðin nútímamáli. Orðalag um klukku hlýtur að hafa borist í málið úr dönsku.

Á 20. öld hafa átt sér stað talsverðar breytingar varðandi klukkuna því að þá öttu kappi tvö orðalagskerfi, gamla kerfið *gengin í/yfir* og *vantar... í*, og nýja kerfið *yfir* og *í*. Hið nýja virðist stefna í að verða einrátt og hér hafa verið nefndar til ýmsar ástæður þess.

<sup>21</sup>Í sænsku er *över* og *i*, og í norsku *over* og *på*.

## Heimildir

- Alþingisbækur Íslands* VIII. 1949–1955. Reykjavík.
- Árni Björnsson. 1990. Tímatál. Í: Frosti F. Jóhannesson (ritstj.) *Íslensk þjóðmenning* VII, bls. 51–101. Reykjavík: Bókaútgáfan Þjóðsaga.
- DI X = *Diplomatarium Islandicum. Íslenzkt fornbréfasafn*. 1911–1921. Reykjavík.
- Íslensk orðabók*. 2002. Ritstjóri Mörður Árnason. Þriðja útgáfa, aukin og endurbætt. Reykjavík: Edda.
- Jón Hilmar Jónsson. 2001. *Orðastaður*. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2005. *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.
- Jón Ólafsson. 2005. *Relatio af Kaupinhafnarbrunnum sem skeði í október 1728. Dagbók 1725–1731 og fleiri skrif*. Sigurgeir Steingrímsson gaf út. Reykjavík: Góðvinir Grunnavíkur-Jóns.
- Jónas Jónasson. 1961. *Íslenzkir þjóðhættir*. Þriðja útgáfa. Einar Ól. Sveinsson bjó undir prentun. Reykjavík: Ísafoldarprentsmiðja hf.
- K[onráð] Gíslason. 1851. *Dönsk orðabók með íslenzkum þýðingum*. Kaupmannahöfn.
- Ordbog over det danske sprog*. 1918–1956. Kaupmannahöfn.
- Orðabók Blöndals, sjá Sigfús Blöndal.
- Reykjahólabók* I–II. 1969–1970. Útg. Agnete Loth. Kaupmannahöfn: Munksgaard.
- Sigfús Blöndal. 1920–1924. Reykjavík: *Íslensk-dönsk orðabók*.
- Stefán Einarsson. 1949. *Icelandic*. Baltimore: The Johns Hopkins Press.
- Æfisaga Jóns Ólafssonar Indíafara*. 1908–1909. Með athugasemdum eftir Sigfús Blöndal. Kaupmannahöfn.

## Heimildir dæma í Ritmálssafni Orðabókar Háskólans

- Alþb VIII = *Alþingisbækur Íslands*. 1912–1969. Reykjavík.
- BeerFerðam I = [Dominicus Beer.] 1694. *Sa Andlege Ferda-Madur*. [Þýðandi Þorsteinn Illugason.] Skálholti.
- DI X = *Diplomatarium Islandicum. Íslenzkt fornbréfasafn*. 1911–1959. Reykjavík.
- Fjallk 1890 = *Fjallkonan*. 1884–1911. Reykjavík.
- GhagalMaríum = Guðmundur Gíslason Hagalín. 1950. Við Maríumenn. Akureyri: Bókaútgáfa Pálma H. Jónssonar.

HBriemEnsk = Halldór Briem. 1875. *Kennslubók í enskri tungu eptir Halldór Briem cand. theol.* Reykjavík.

JÁNucl = Jón Árnason. 1738. *Nucleus Latinitatis*. Kaupmannahöfn.

JÓlGrv. ÁMSkr. I 2 = [Jón Ólafsson úr Grunnavík]. 1930. *Vita Arnæ Magnæi af J. Ól.* Kaupmannahöfn. [Prentað í ritinu *Árni Magnússons levned og skrifter.*]

Jóllnd I = Jón Ólafsson. 1946. *Reisubók Jóns Ólafssonar Indíafara samin af honum sjálfum 1661 I–II.* Guðbrandur Jónsson gaf út. Reykjavík: Bókfellsútgáfan.

PVídSkýr = Páll Vídalín. 1849–1854. *Skýringar yfir fornnyrði lögbókar þeirrar, er Jónsbók kallast.* Reykjavík.

SkrifStap = *Skrifarinna á Stapa. Sendibréf 1806–1877.* 1957. Finnur Sigmondsson bjó til prentunar. Reykjavík: Bókfellsútgáfan.

ÞÞCal = Þórður Þorláksson. 1692. *Calendarium Perpetuum Ævarande Tijmatal, Edur Rijm Iislendskt til ad vita hvad Arsins Tijdum lijdur.* Skálholti.

ÞÞEnchir E = Þórður Þorláksson. 1671. *Enchiridion.* Hólum.

## Abstract

### ‘What time is it?’

Some old examples (from the 16th century onwards) referring to the clock are discussed and compared to the modern language. In Modern Icelandic there are two competing systems denoting exact time. An attempt is made to explain the development of the younger system and why the older one is becoming obsolete.

### Keywords:

time denotations, language change, simplification

Katrín Axelsdóttir

Íslenskuskor

Háskóla Íslands

Nýja-Garði við Sæmundargötu

IS-101 Reykjavík

katax@hi.is





Margrét Jónsdóttir

## Um *ærsl*, *busl* og *usl* í orðasamböndum<sup>1</sup>

### 1 Inngangur

Í því sem hér fer á eftir verður sjónum einkum beint að tveimur íslenskum orðasamböndum, fyrst og fremst *ærsl* og *busl* en einnig *usl* og *busl* ásamt þeim fylgiorðum sem fara með samböndunum. Skoðaðar verða heimildir um þessi sambönd og jafnframt verður reynt að rekja sögu þeirra í málinu.

Efnisskipan greinarinnar verður sem hér segir. Í öðrum hluta verða raktar heimildir úr orðabókum. Í þeim þriðja verður gerð grein fyrir beinum notkunarheimildum. Í fjórða hluta er örstuttur kafli um orðasambandið *pêle-mêle*. Það er franskt að uppruna en var og er notað í dönsku. Eins og lesa má í öðrum kafla þýddi Gunnlaugur Oddsson það m.a. með *ærsl* og *busl*. Í fimmta hluta verður rætt um parasambönd og eðli þeirra. Í þeim síðasta verða örstutt lokaorð.

---

<sup>1</sup>Þakka ber ókunnum yfirlesara/yfirlesurum mjög gagnlegar ábendingar. Aðalsteini Eyþórssyni, Áslaugu J. Marinósdóttur, Gunnlaugi Ingólfssyni, Magnúsi Hauksyni, Oddnýju G. Sverrisdóttur og Peter Weiss skal líka þakkað fyrir veitta aðstoð af ýmsum toga. Þess skal jafnframt getið að Pétur Halldórsson hjá RÚVAK spurðist fyrir um orðasambandið í úttvarpsþætti nýlega en hafði ekki árangur sem erfiði. En honum ber líka að þakka.

## 2 Heimildir úr orðabókum

Í orðabók Gunnlaugs Oddssonar (1991:212 (1819)) segir m.a. undir flettunni *ærsl*:

- (1) **ærsl**... á *ærsl* og *busl* - > *pêle-mêle*<sup>2</sup>

Undir *pêle-mêle* segir svo (bls. 67):

- (2) *pêle-mêle*... upp og nidr, á *æsl* [svo] og *busl*.<sup>3</sup>

Orðasambandið með *ærsl* og *busl* er að finna hjá Sigfúsi Blöndal (1920–1924:121, undir *busl*). Þar segir:

- (3) **busl** ... 3. (hávaði), Støj, Tummel: *allt er komið í ærsl* og *b.[usl]*

Hér er merkingin ‘hávaði’ og því frábrugðin því sem lesa má út úr (1) og (2) þar sem vísað er til óreiðu. Í *ÍO* (= *Íslensk orðabók* 2002) er orðasambandið *ærsl* og *busl* ekki nefnt. Þar er hins vegar afbrigðið með *usl* og *busl* undir flettunni *usl* (bls. 1662) og merkingin vísar til óreiðu rétt eins og í (1) og (2):

- (4) **usl** HK · *kliður*, *busl*, *ys*, *þys* > *allt er komið á usl* og *busl* þ.e. á ringulreið, í glundroða

Þetta sama afbrigði er í *Íslenskri samheitaorðabók* (1985:513) undir flettunni *ringulreið*.

Sem fyrr segir er orðasambandið með *ærsl* og *busl* ekki að finna í *ÍO*, aðeins það með *usl* og *busl*. Og það kemur ekki á óvart að flest bendi til þess að sambandið með *ærsl* og *busl* sé flestum ókunnugt nú, t.d. eru engin dæmi um það í talmálssafni *OH* (= Orðabók Háskólans). Á Netinu eru þó nokkur dæmi en við nánari skoðun má rekja þau öll nema eitt til fyrirlestrar um sálmabók þá sem kölluð var *Leirgerður*. Í þeirri bók er vísa með þessu orðasambandi, sbr. (6) í þriðja hluta. Dæmið sem eftir er sýnir heldur ekki sjálfstæða notkun orðasambandsins; um það verður rætt í lok þriðja kafla. Engin dæmi fundust um sambandið

<sup>2</sup>Hjá Gunnlaugi er *pêle-mêle* skrifað með öfugum broddi. Eldri mynd sambandsins er *pesle-mesle* sem skv. *Dictionnaire historique de la langue française* (1998:2636) er sagt vera orðið til úr *mesle-mesle*. Af upprunaástæðum, þ.e. vegna *s*-ins, er því rétt að hafa hatt yfir *e*-inu enda er þannig farið að í þeim frönsku orðabókum sem koma við sögu hér á eftir. Í þessari grein verður stafsetningin samræmd og því ávallt ritað *pêle-mêle* án tillits til ritháttar hverju sinni.

<sup>3</sup>Í *Fransk-íslenskrri orðabók* (1995:799) er *pêle-mêle* þýtt á svipaðan hátt: *pêle-mêle*... ‘tvist og bast; í óreiðu/einni bendu, í hrúgu; á rúi og stúi’.

með *usl* og *busl*. Það er hins vegar að finna í *ÍO* og *Íslenskri samheita-orðabók* eins og fram hefur komið.

### 3 Notkunarheimildir

Nokkur dæmi eru um orðasambandið með *ærsl* og *busl* í söfnum *OH*.<sup>4</sup> Áður verður þó að geta eins dæmis en það er úr þýðingu Jón Ólafssonar úr Grunnavík á Nikulási Klím, sbr. *ROH* (= *Ritmálssafn Orðabókar Háskólans* undir *busl*). Þar er talað um *ærsl* og *busl* í ákveðnu samhengi.

- (5) ... því eg gjörðe hjer so mikinn óróa, *ærsl*, *busl* og um-  
brot, sem hinn hafði gjörðt í vorum efna heime.  
JÓlGrvKlím, 221 Aldur: 18m

Ekki er hægt að líta svo á að *ærsl* og *busl* sé hér raunverulegt orðasamband eins og sjá má í dæmunum hér á eftir enda standa sögn og/eða forsetning ekki með orðaparinu. Og það skiptir líka máli að hér er *ærsl* notað í stað *ærsl*. Þannig verður hljóðlíkingin við *busl* fjarlægari.<sup>5</sup> Á hinn bóginn gæti orðavalið bent til þess að höfundur hafi þekkt orðasambandið, a.m.k. er merkingin ekki alveg ókunn.

Elstu skráðu dæmin um orðasambandið eru frá um 1800. Það er úr vísu sem Finnur Magnússon orti árið 1802 og tilefnið var útkoma sálmabókarinnar 1801<sup>6</sup>; í daglegu tali er sú bók jafnan nefnd *Leirgerður*; á hana var minnst hér á undan. Textinn (vísuþrot), sem tekinn er úr handritinu Lbs. 1754 4to, innskotsblað á milli bl. 16 og 17 (eignað Finni), er svohljóðandi:

- (6) Allt er komið í *ærsl* og *busl*,  
Andleg sálma er orðin bók  
Andskotalaust rusl.

<sup>4</sup>Öll dæmin sem nefnd verða eru byggð á söfnum Orðabókarinnar. Það skal hins vegar tekið fram að þau hafa verið skoðuð sérstaklega og stundum aukið við það samhengi sem er á seðlinum hverju sinni.

<sup>5</sup>Í nútímamáli eru *ærsl* og *ærsl* til hlið við hlið, sbr. *ÍO*. Þess ber að geta að langflest hvorugkynsorð með *æ* sem stofnsérhljóð enda á *i*. Orð með *u* í stofni gera það hins vegar aldrei.

<sup>6</sup>Dæmi úr textasafni *OH* leiddi greinarhöfund að texta Finns Magnússonar (6). Dæmið er úr *Gesti* (2. bindi, bls. 119), ritsafni sem Gils Guðmundsson gaf út (1985).

Um þetta segir Finnur (á sama stað):

- (7) Eg kvað þessar vísur 1802... Talshátturinn „Allt er komið í *ærsl* og *busl*“ tíðkaðist síðan, um þær mundir, á Suðurlandi; hvört enn, veit eg ekki.

Í (8) eru dæmi úr *ROH* (undir *busl*).<sup>7</sup>

- (8) a. Tíundarstatútunni sé í svo mörgum greinum ekki fylgt nú, að allt komizt á *ærsl* og *busl*, ef menn vildu fara að fylgja henni stranglega.  
Ldsyrd VI, 367 (1852) Aldur: 19m
- b. ... enn ef alt kemst á *ærsl* og *busl*, geta menn trúað, að enn fleiri Bakka bræðra nótar villist inn á þingið enn nú eru þar.  
Fjallk 1897, 153 Aldur: 19s
- c. Hið náttúrlega og yfirnáttúrlega var í óendanlegum ruglingi hvað innan um annað; það var eins og „allt væri komið í *ærsl* og *busl*“.  
Þjóð 1889, 208 Aldur: 19s

Hér á undan í (4) var minnst á sambandið með *usl* og *busl*. Um það eru eftirtalin dæmi í söfnum *OH*:<sup>8</sup>

- (9) a. allt er komid á *usl* og *busl*.  
Lbs. 220 8vo, bls. 61, undir *busl*<sup>9</sup>
- b. Við sjáum þá hvað setur, enda verður allt á *usli* og *busli*, einlæggar fylkja-kosningar.  
SGStBR I, 283 Aldur: 19s20f (skrifað 1911)

<sup>7</sup>Úr þessu dæmasafni úr *ROH* er sleppt einu dæmi úr *Austra* (1884) þar sem ekki hefur tekist að sannreyna það (vegna rangs blaðsíðutals). Auk þess er sleppt tveimur dæmum sem eru samhljóða (6). Jafnframt verður að geta hér eins dæmis úr *ROH* (undir *busl*):

- a. með þýzku orðatiltæki „Thun und treiben“ (*ærsl* og *busl*).  
Nf II, 33, 19m

Í orðabók Jóns Ófeigssonar (1953:635–636) er *Treiben* undir flettunni *treiben*; *sein Tun und Treiben* er þýtt sem ‘athafnir hans og hegðun’. Af orðasambandabók *Duden* (1992:739–740, undir *tun*) má ráða að sambandið sé notað í formlegu máli. Nafnorðið *Tun* er sagnanafnorð af sögninni *tun* ‘gera’ og *Treiben* er sömuleiðis sagnanafnorð, af *treiben* ‘drífa, reka, elta; gera’. Merking nafnorðanna er í þessu samhengi sú sama. Sú óreiðu- eða ringulreiðarmerking sem einkennir *pêle-mêle* er því hér víðs fjarri.

<sup>8</sup>Hér verður líka að geta tveggja dæma með *usl* og *busl* þar sem merkingin er ljóslega ‘hávaði’, sbr. einnig Sigfús Blöndal (1920–1924: 896, undir *usl*). Notkunin er því nokkuð ólík því sem er í (8) og (9).

Tvennt vekur einkum athygli. Séu dæmin *um ærsl/usl og busl* skoðuð í heild sést að sögnin *koma* (með *vera*) kemur langoftast við sögu en líka *komast*; í (9b) er sögnin *verða* í stað *vera*; þar er hins vegar hvorki *koma* né *komast*. Sögninni fylgir ávallt forsetning, ýmist *í eða á*. Í annan stað kemur í ljós að dæmin eru öll frá 19. öld nema b.-dæmið í (9). Því eru ummæli Finns Magnússonar, sbr. (7), frá árinu 1802 forvitnileg. Hann tengir notkunina á orðasambandinu Suðurlandi. En hann segir svo: „hvört enn, veit eg ekki“. Hvort Finnur er með þessum orðum að tengja sambandið við Suðurland eða vísa til notkunarinnar skal ósagt látið. Dæmin úr *ROH* sýna á hinn bóginn dæmi sem ná fram á 20. öld og engin sérstök ástæða er til að tengja þau Suðurlandi.

## 4 Um *pêle-mêle*

Eins og fram kom í upphafi annars hluta þýddi Gunnlaugur Oddsson orðasambandið *pêle-mêle* m.a. með *ærsl og busl*, sbr. (1) og (2). En hvað er um *pêle-mêle* að segja? Er hægt að rekja sögu þess á einhvern hátt?

Í *Nudansk Ordbog* (1982:715) er orðasambandið *pêle-mêle*, sem þýtt er með ‘hulter til bulter’, sagt komið úr frönsku.<sup>10</sup> Í *Dictionnaire historique de la langue française* (1998:2636) má sjá að það er gamalt og jafnframt að dæmi eru um þá óreiðumerkingu sem rætt var um í öðrum hluta frá 1596. En hvenær skyldi *pêle-mêle* fyrst koma fyrir í dönsku? Um það er ekki gott að segja. Þó er sambandið að finna í orðabók frá 1807 um erlend orð í dönsku eftir Carl Friedrich Primon; þar er það þýtt á dönsku með almennum orðum en ekki orðasambandi: ‘iblandt

- 
- a. ... þeir [hverirnir] gjósa báðir undir eins, og heyrast *uslið og buslið* í þeim langar leiðir, ...  
ÞThFerð II, 200, 19s20f
- b. ...að menn urðu varir við mikið *usl og busl*, ræskíngar, andvörp og stunnur frammí bíslaginu, líkt og þar væri skipshöfn að farast, en það var prófasturinn.  
HKLSalka 1959, 102, 20m

Þess má geta að Sigfús Blöndal tilgreinir einmitt dæmið frá Þorvaldi í bók sinni. Hvort Halldór kynntist *usl og busl* hjá Sigfúsi skal ósagt látið. En í þessu sambandi er við hæfi að rifja upp orð Guðrúnar Kvaran (2002:220) um Halldór: „Meðal annars þaulþekkti hann orðabók Sigfúsar Blöndals og nýtti sér hana til hins ítrasta.“

<sup>9</sup>Handritið Lbs. 220 8vo er íslensk orðabók frá fyrri hluta 19. aldar. Margt bendir til þess að Hallgrímur Scheving hafi verið höfundurinn. Sjá nánar Guðrúnu Kvaran (1988:54). Gunnlaugur Ingólfsson benti á þetta dæmi; það kemur úr ‘gulu seðlunum’.

<sup>10</sup>Í *Dansk-íslenskri orðabók* (2004:592) segir að *pêle-mêle* þýði ‘óreiða, ringulreið’.

hverandre, broget, i hverandre'. En eins og fram kemur hjá Jóni Hilmari Jónssyni (1991:xv) studdist Gunnlaugur Oddsson m.a. við þá bók við samningu bókar sinnar.<sup>11</sup> Hvort Gunnlaugur kynntist *pêle-mêle* í gegnum bókina er hins vegar engin leið að fullyrða.

Hér áður hefur oftlega verið minnst á óreiðumerkinguna í *pêle-mêle*. Því má bæta hér við að í þýsku er *pêle-mêle* í óreiðumerkingu af einhverjum toga að finna í framandorðabók *Duden* (2005:777). Í ensku merkir sambandið *pell-mell* eitthvað svipað (Onions 1967:663) og eru dæmi um það frá 17. öld í þessari mynd en frá 16. öld sem *pelly melly*.

## 5 Um parasambönd og eðli þeirra

Í *Nudansk Ordbog* (1982:715) er *pêle-mêle* þýtt með 'hulter til bulter' eins og áður hefur komið fram. Í fransk-enska hlutanum af Collins-Robert... (1987:510) er *pêle-mêle* m.a. þýtt með 'higgledy-piggledy'. Í báðum þýðingunum er því spilað á hljómræma eiginleika.

Eins og lesa má í (2) þýddi Gunnlaugur Oddsson orðasambandið *pêle-mêle* á tvennan hátt. Annars vegar með *upp* og *niður*, hins vegar með sambandi með *ærsl* og *busl*. Á þessum samböndum er nokkur munur. Í sambandinu *upp* og *niður*<sup>12</sup> liggur áherslan á sérhljóðunum. Sambandið með *ærsl* og *busl* er hins vegar af svoltið öðrum toga. Það er fyrst og fremst vegna endurtekningar samhljóðanna þar sem röðin *-sl* er endurtekin og í sambandinu með *usl* og *busl* bætist svo sérhljóðið við. Það er því innrím sem felst í endurtekningu sem skiptir höfuðmáli. Jón G. Friðjónsson (1997:ix) segir innrím eitt það sem einkennir orðatiltæki. Hann nefnir ýmis dæmi máli sínu til stuðnings. Þar má t.d. nefna *nótt sem/og nýtan dag*, þar sem er stuðlasetning á *n* í framstöðu auk þess sem röðin *Vt-* skiptir máli í innstöðu. Þetta samband er þekkt frá fyrri hluta 19. aldar (Jón G. Friðjónsson 1997:235) en líka má t.d. nefna *réttur og sléttur* og *vinna hvorki né spinna*. Bæði samböndin eru af sama toga og sambönd með *ærsl* og *busl* og *usl* og *busl*. Jón (bls. 206) nefnir líka þá athyglisverðu staðreynd að í sambandinu *eftir hlátur kemur grátur* er þolfallið *grátur* notað innrímisins vegna en ekki *grát*.

<sup>11</sup>Það skal tekið fram að skrif Jóns Hilmars um áhrif Primons á verk Gunnlaugs urðu til þess að greinarhöfundur fékk Magnús Hauksson, lektor í Kaupmannhöfn, til að skoða bókina. Og þar fann hann dæmið.

<sup>12</sup>Jón Friðjónsson (1993:460) segir að orðasambandið *upp* og *niður* sé þekkt frá upphafi 17. aldar

Þetta samband er úr Guðbrandsbiblíu. Mörg fleiri innrímssambönd er þar að finna, t.d. samböndin *vinna hvorki né spinna* (bls. 367), *fríðurinn nærir, ófríðurinn fortærir* (bls. 205) og *hneigja sig og beygja* (bls. 309). En Jón nefnir líka önnur sambönd eins og t.d. *réttur og sléttur* sem er frá 16. og 17. öld (bls. 148) og *með ráð(um) og dáð(um)* frá 15. öld (bls. 193). Í ROH má lesa að elstu dæmi um sambandið *á rúi og st(r)úi* eru frá 17. öld.

En það er ekki aðeins hljómræna endurtekningu að finna í samböndunum með *ærsl* og *busl* eða *usl* og *busl*. Endurtekningin er líka merkingarleg þar sem nánast sama merkingin kemur fyrir í tvígang. Í íslensku eru önnur dæmi af sama toga, t.d. *arga og garga, veina og kveina, væla og skæla, úa og grúa og á rúi og st(r)úi*. Í þessum samböndum er raunar meiri endurtekning en í sambandinu með *ærsl* og *busl* þar sem sama stofnsérhljóðið auk niðurlagsins er í hverju pari fyrir sig. Þau líkjast því meira sambandinu með *usl* og *busl*. Í samböndunum *tangur* og *tetur* eða *sorg* og *sút* er aðeins merkingin endurtekin.

Einkenni parasambanda geta verið fleiri. Ýmsir, m.a. Cooper og Ross (1975:71 o.áfr.), hafa fjallað um eðli slíkra sambanda af ýmsum gerðum í ensku. Þeir gera ráð fyrir að flokka megi samböndin í sjö meginflokkka. Sé sambandið með *ærsl* og *busl* skoðað í því ljósi sést að það getur auðveldlega fallið undir tvær viðmiðanir þeirra. Þar ber fyrst að nefna að það þykir dæmigert að hindrunarhljóð/hljómleysingi (e. obstruent) skuli vera í framstöðu orðsins *busl* þar sem samhjóðið er aðeins eitt. Það telst líka dæmigert að sérhljóðið í fyrra orðinu (eða a.m.k. seinni hluti þessa tvíhljóðs), þ.e. *ærsl*, skuli vera fram-mælt en uppmælt í *busl*, því síðara. Á hinn bóginn skiptir það ekki máli um sambandið með *usl* og *busl* enda sérhljóðið eitt og hið sama.<sup>13</sup> Ein viðmiðana Cooper og Ross er sú að í seinna orðinu séu fleiri samhjóð í framstöðu en í því fyrra. Orðið *ærsl* hefst ekki á samhjóði. En sé viðmiðunin túlkuð svo að það geti talist eðlilegt gangvart framstöðusamhljóðinu í *busl* gæti sambandið *ærsl* og *busl* fallið undir þrjár viðmiðanir.

Nú vaknar sú spurning hvers vegna Gunnlaugur Oddsson kaus að þýða erlent orðasamband með íslenskum orðasamböndum, hvort

<sup>13</sup>Sé orðasambandið *upp* og *niður* skoðað út frá hugmyndum Cooper og Ross þá er helsta einkenni þess að fyrra orðið er styttra en það síðara. Þetta er í raun lögmál kennt við þýska málfræðinginn Otto Behagel. Skv. því kemur lengsta orðið síðast í röð þeirra orða sem hafa sama vægið. Sjá nánar Collinge (1985:241).

sem það er *upp og niður* eða *ærsl og busl* með fylgiorðum. Fyrirmyndina að því að nota íslensku gæti hann hafa sótt til Primons sem notaði dönsku í skýringum sínum. Það skilur hins vegar á milli Gunnlaugs og Primons að Gunnlaugur þýðir erlent orðasamband með íslenskum orðasamböndum. Primon notaðist hins vegar ekki við orðasamband heldur einungis almennt orðafar, þ.e. 'iblandt hverandre, broget, i hverandre', sbr. líka fjórða hluta hér að framan. Þriðji möguleiki Gunnlaugs hefði verið að þýða erlent orðasamband með orðasambandi úr erlendu máli. Skv. Ásgeiri Blöndal Magnússyni (1989: 360) kom orðasambandið *holt og bolt* inn í íslensku úr dönsku á 18. öld. Það hefði t.d. getað þjónað vel sem þýðing á *pêle-mêle* rétt eins og í nútímadönsku. En þá leið valdi Gunnlaugur ekki enda ekki í samræmi við afstöðu hans til málsins. Í því sambandi er vert að vísa til orða Jóns Hilmars Jónssonar (1991:xxxii) sem hefur skoðað orðafarið í orðabók Gunnlaugs, m.a. í ljósi hreintungustefnunnar. Hann segir m.a.:

- (10) Sú mynd sem birtist í orðabók Gunnlaugs Oddssonar og öðrum ritum hans að þessu leyti er býsna margbrotin og raunar harla sundurleit. Sú meðvitaða hreintungustefna og málhreinsun sem síðar verður ríkjandi er lítt komin til sögunnar því ýmissa danskra áhrifa gætir í orðafari og setningaskipan sem síðar var lögð mikil áhersla á að útrýma. En engu að síður er orðfærið mótað af eindreginni viðleitni til að tjá sig á skýru og meitluðu íslensku máli um nýstárleg fyrirbæri.

Í ljósi þessara orða Jóns Hilmars er sá kostur sem Gunnlaugur velur eðlilegur og í sjálfu sér til fyrirmyndar frá sjónarhóli nútímamálvöndunar. En hvort sem það var hann sem fyrstur þýddi *pêle-mêle* með *ærsl og busl* eða ekki (enda þótt hann væri ekki sá fyrsti til að nota orðasambandið sem slíkt) breytir það ekki þeirri staðreynd að þýðingin nær ekki þeim hljómrænu eiginleikum sem *pêle-mêle* hefur; *upp og niður*, sbr. (2), breytir þar heldur engu. En það eru einmitt þessir hljómrænu eiginleikar sem hafa orðið fyrirmyndir í öðrum málum, sbr. dönskuna *hulter til bulter* og enskuna *higgledy-piggledy*. Á hinn bóginn hefði Gunnlaugi tekist betur upp hefði hann notað sambandið *usl og busl* enda endurtekningin meiri.



## 6 Lokaorð

Hér á undan hefur verið dreginn saman örlítill fróðleikur um orðasambönd sem flestum eru ókunnug nú. Úr litlu er að móða enda heimildirnar litlar. Þær eru allar ritmálsbundnar. Engin leið er að skýra hvers vegna orðasamböndin með *ærsl* og *busl* og *usl* og *busl* í viðeigandi umhverfi virðast nánast hafa horfið úr málinu, a.m.k. ef marka má ritmálsheimildir.

### Heimildir

- Ásgeir Blöndal Magnússon. 1989. *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Collinge, N. E. 1985. *The Laws of Indo-European*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Collins·Robert French~English English~French Dictionary. 1987. Second Edition. London, Glasgow & Toronto: Collins. New York: Harper & Row.
- Cooper, William E. og Ross, John Robert. 1975. World Order. *Papers on the Parasession on Functionalism* April 17. 1975, bls. 63–111. Chicago: Chicago Linguistic Society.
- Dictionnaire historique de la langue française*. 1998. Ritstjórar: Alain Rey [aðalritstjóri], Marianne Tomi, Tristan Hordé, Chantal Tanet. Paris: Dictionnaires Le Robert.
- Duden*. 1992. Redewendungen und sprichwörtliche Redensarten. Wörterbuch der deutschen Idiomatik. Bearbeitet von Günther Drosdowski und Werner Scholze-Stubenrecht. Band 11. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Duden*. 2005. Fremdwörterbuch. 8., neu bearbeitete und erweiterte Auflage. Herausgegeben von der Dudenredaktion. Band 5. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Dönsk-íslensk orðabók*. 2004. 2. útgáfa. Ritstjóri Halldóra Jónsdóttir. Reykjavík: Mál og menning.
- Frönsk-íslensk orðabók*. 1995. Ritstjóri: Þór Stefánsson. Örn og Örlygur hf. Reykjavík, Paris: Dictionnaires Le Robert.
- Gils Guðmundsson. *Gestur. Íslenskur fróðleikur gamall og nýr*. 2. Gils Guðmundsson safnaði efninu. Reykjavík. 1985.

- Guðrún Kvaran. 1988. Sérsofn orðabókar Háskólans. *Orð og tunga* 1:51–64.
- Guðrún Kvaran. 2002. Úr fórum Halldórs Laxness. *Íslenskt mál og almenn málfræði* 24:219–235.
- Gunnlaugur Oddsson. 1991 (1819). *Orðabók sem inniheldur flest fágæt, framandi og vandskilin orð er verða fyrir í dönskum bókum*. Ný útgáfa með íslenskri orðaskrá. Jón Hilmar Jónsson sá um útgáfuna ásamt Þórdísi Úlfarsdóttur. Reykjavík: Orðabók Háskólans.
- ÍO = *Íslensk orðabók*. 2002. Þriðja útgáfa, aukin og endurbætt. Ritstjóri: Mörður Árnason. Reykjavík: Edda.
- Íslensk samheitaorðabók*. 1985. Ritstjóri: Svavar Sigmundsson. Reykjavík: Styrktarsjóður Þórbergs Þórðarsonar og Margrétar Jónsdóttur Háskóla Íslands.
- Jón G. Friðjónsson. 1993. *Mergur málsins. Íslensk orðatiltæki, uppruni, saga og notkun*. Reykjavík: Örn og Örlygur, Bókaklúbbur hf.
- Jón G. Friðjónsson. 1997. *Rætur málsins. Föst orðasambönd, orðatiltæki og málshættir í íslensku biblíumáli*. Reykjavík: Íslenska bókaútgáfan.
- Jón Hilmar Jónsson. 1991. Um orðabók Gunnlaugs Oddssonar. Í: Gunnlaugur Oddsson. 1991 (1819). *Orðabók sem inniheldur flest fágæt, framandi og vandskilin orð er verða fyrir í dönskum bókum*, bls. ix–xxxvii. Jón Hilmar Jónsson sá um útgáfuna ásamt Þórdísi Úlfarsdóttur. Reykjavík: Orðabók Háskólans.
- Jón Ófeigsson. 1953. *Þýzk-íslensk orðabók. Deutsch-isländisch Wörterbuch*. Önnur útgáfa. Reykjavík: Ísafoldarprentsmiðja h.f.
- Nudansk Ordbog*. 1982. 11. reviderede og forøgede udgave ved Erik Oxenvad. København: Politikens Forlag.
- Onions, C. T. 1966. *The Oxford Dictionary of English Etymology*. Edited by C. T. Onions. Oxford: The Clarendon Press.
- ROH = *Ritmálssafn Orðabókar Háskólans*. Vefsíða OH: [www.lexis.hi.is](http://www.lexis.hi.is)
- Sigfús Blöndal. 1920–1924. *Íslensk – dönsk orðabók*. Reykjavík.

### Vísanir til dæma úr söfnum Orðabókar Háskólans

- Austri = *Austri*. 1. árgangur. Seyðisfirði. 1884.
- Fjallk = *Fjallkonan*. 14. árgangur. Reykjavík. 1897.
- HKLSalka = Halldór Kiljan Laxness. *Salka Valka*. Reykjavík. 1959.

JÓlGrvKlím = Ludvig Holberg: Nikulás Klím. Íslenzk þýðing eftir Jón Ólafsson úr Grunnavík. Jón Helgason bjó til prentunar. Kaupmannahöfn. 1948. *Íslenzk rit síðari alda* 3.

Ldsyrd = *Landsyfjerréttardómar og hæstaréttardómar í íslenzkum málum 1802–1873*. VI. 1845–1852. Reykjavík 1946–1950.

Nf = *Norðanfari*. 2. árgangur. Akureyri. 1863.

SGStBR = Stephan G. Stephansson: *Bréf og ritgerðir* 1. Reykjavík. 1938.

Þjóð = *Þjóðólfur*. 41. árgangur. Reykjavík. 1889.

ÞthFerð = Þorvaldur Thoroddsen. *Ferðabók*. Skýrslur um rannsóknir á Íslandi 1882-1898. Annað bindi. Kaupmannahöfn. 1914.

Handritin *Lbs. 1754 4to* og *Lbs. 220 8vo* eru á handritadeild Landsbókasafns Íslands – Háskólabókasafns.

## Abstract

### On ærsl, busl and usl in Icelandic idioms

In Icelandic literary sources, idioms containing the word pairs *ærsl og busl* or *usl og busl* are found (always with different determiners). Examples almost exclusively stem from the 19th century, the oldest (*ærsl og busl*) dating from 1802.

At present, these idioms seem to have been almost universally forgotten, even though *usl og busl* is found in the latest edition of *Íslensk orðabók* (2002). In the paper, the story of these idioms is traced and remarks on the nature of such idioms and some their properties are included.

### Lykilorð:

Icelandic idioms

Margrét Jónsdóttir

Hugvísindadeild Háskóla Íslands

Árnagarði við Suðurgötu

IS-101 Reykjavík.

mjons@hi.is



Þórdís Úlfarsdóttir

# Málfræðileg mörkun orðasambanda

## 1 Inngangur

Hér verður lýst tilraun til að marka orðasambönd eins og þau sem eru í orðasambandaskrá Orðabókar Háskólans.<sup>1</sup> Að *marka* er það kallað þegar tölvutækur texti er greindur málfræðilega á vélrænan hátt og greiningarstrengur látinn fylgja hverju orði. Forritið sem notað er til þess nefnist *markari* og verknaðurinn heitir *mörkun*. Orðasamböndin sem um ræðir eru af margvíslegu tagi, orðasamband getur til dæmis verið lýsingarorð sem stendur með nafnorði: *feitt embætti*, sögn með sagnfyllingu: *vera glaður og svipléttur*, eða orðtak: *aka seglum eftir vindi*. Þessi þrjú orðasambönd eru meðal þeirra sem prófað var að marka.

Málfræðileg greining orðasambanda er í því fólgin að tekin eru valin sambönd og þau undirbúin fyrir vélræna greiningu, m.a. með því að setja aðgreinandi tákni á milli þeirra. Síðan er markarinn látinn greina málfræðilega einstök orð í samböndunum. Orðasambönd hafa ekki áður verið greind á þennan hátt í íslensku svo vitað sé.

Byrjað var að marka orðasamböndin í mars 2005 og enda þótt ekki væri farið af stað með ákveðið markmið í huga varð fljótlega ljóst að mörkunin mundi gefa forvitnilegar niðurstöður. Skömmu áður hafði Sigrún Helgadóttir skilað lokaskýrslu um verkefnið **Markarann** sem var nýtt tungutækniól fyrir íslensku, en hún hafði tekið að sér tæknilega umsjón forritsins og þjálfun þess fyrir íslenska texta hjá Orðabók

<sup>1</sup>Jón Hilmar Jónsson hefur haft umsjón með orðasambandaskránni og hugmyndin að því að marka orðasamböndin er frá honum komin.

Háskólans. Með þetta verkfæri í höndum opnuðust nýir möguleikar til málrannsóknna því nú var hægt að greina ýmsar tegundir texta í málfræðilegar einingar á skjótan hátt með aðstoð tölvu. Áhugavert var að sjá hvað kæmi út úr mörkun orðasambanda og athuga hvort og hvernig hægt væri að hagnýta þær niðurstöður. Einn ávinningur af þessari tilraun var að mögulegt var að raða orðasamböndunum á nýjan hátt, ekki eftir stafrófsröð heldur eftir greiningarstrengjunum, og komu þá í ljós ýmis forvitnileg setningarleg mynstur sem orðasamböndin mynduðu.

Efnisskipan greinarinnar er þannig að í 2. kafla er fjallað um orðasamböndin og í 3. kafla um markarann og mörkunina. Í 4. kafla er ferli mörkunar lýst og í 5. kafla segir frá nákvæmni mörkunar og helstu takmörkunum hennar. Í 6. kafla eru listar með mörkuðum orðasamböndum og þar er jafnframt greint frá niðurstöðum mörkunarinnar. Í 7. kafla er lýst hvernig hægt er að raða orðasamböndum eftir málfræðigreiningunni, og 8. kafli fjallar um það þegar markaðar eru nokkrar línur af skýringartexta orðabókar. Niðurstöður eru síðan dregnar fram í lokaorðum í 9. kafla.

## 2 Orðasamböndin

Orðasambönd í orðasambandaskrá Orðabókar Háskólans eru nú um 135 þúsund talsins. Þau eiga upptök sín í Ritmálsskrá Orðabókar Háskólans og eru úrvinnsla úr dæmasafni hennar. Jón Hilmar Jónsson hefur byggt upp orðasambandaskrána sem er aðgengileg á vefnum. Hægt er að fletta upp í skránni á [www.lexis.hi.is/osamb/osamb.pl](http://www.lexis.hi.is/osamb/osamb.pl) og þar er einnig að finna upplýsingar um hana.

Orðasamböndin hafa að geyma mikinn og fjölbreytilegan orðaforða, í þeim 135 þúsund orðasamböndum sem nú eru tiltæk koma fyrir rúmlega 57.500 mismunandi lykilorð (þ.e. orð sem mynda merkingarkjarna sambandsins). Orðasambandaskránni er ætlað að veita yfirsýn um föst orðasambönd og tilbrigði þeirra og endurspeglar dæmigerða notkun orðanna auk þess sem hún veitir innsýn í sögu og samhengi íslensks orðaforða á síðari öldum. Í skránni er að finna orðasambönd af margvíslegu tagi, allt frá föstum samböndum í óbreytanlegri mynd (*gera sér gláðan dag*) til lauslegra orðastæðna (*aka á hundasleða*).

Á dæmunum fyrir neðan má sjá að orðasamböndin eru ekki "setningar" þar sem í aðeins sumum þeirra er sögn í persónuhætti, og oft

er engin sögn í sambandinu. Til að lýsa nánar því efni sem orðasambandaskráin hefur að geyma eru sýnd hér nokkur sambönd sem tengjast ákveðnum lykilorðum, fyrst koma dæmigerð sambönd með sögn en svo með lýsingarorði. Þessi orðasambönd voru meðal þeirra sem mörkunartilraunin náði yfir.

### 16 orðasambönd með sögninni aka:

aka angri á bug  
 aka á hundasleða  
 aka á tún  
 aka á völl  
 aka á þvottabretti  
 aka barnakerru  
 aka (< bílunum >) um vegleysur  
 aka dráttarvél  
 aka/geta ekið (< nokkurn >) bug á < keisarann, Frakka >  
 aka greitt  
 aka/hafa (< lengi >) ekið höllu  
 aka heilum vagni heim  
 aka hjólasleða  
 aka hjólbörum  
 aka < honum, henni > á bug  
 aka höllu fyrir < honum, henni >

### 10 orðasambönd með lýsingarorðinu feitur:

feit staða  
 feit steik  
 feitt embætti  
 feitt ket  
 feitt kjöt  
 feitt letur  
 feitt og holdasamt < fé >  
 hafa ekki um feitt að sleikja  
 vera feitur og líkamaþungur  
 vera feitur og pattaralegur

Orðasamböndin hafa nokkur séreinkenni sem sjá má á dæmunum að ofan. Meðal þeirra má nefna að svigaliðir koma mjög oft fyrir í þeim. Innihald sviganna er þá valfrjáls viðbót við orðasambandið sem þó er dæmigerð fyrir viðkomandi orðanotkun. Dæmi um slíkan svigalið er í sambandinu *aka (< bílunum >) um vegleysur*, þar sem hægt er að sleppa sviganum og hafa styttra form á sambandinu: *aka um vegleysur*.

Annað einkenni á orðasambandaskránni er að tilbrigði í tíðum sagna eru sýnd með skástriki á milli, eins og til dæmis í sambandinu *aka/hafa (< lengi >) ekið höllu*. Þetta er gert til að sögnin komi fram í

flettimynd sinni, nafnhætti, og til að notkun hennar birtist einnig á sem eðlilegastan hátt í orðasambandinu.

Í fyrstu tilraun til að marka orðasambönd voru notuð 2296 sambönd úr orðasambandaskránni sem öll byrjuðu á bókstafnum *a-*, án tillits til einstakra lykilorða innan sambandanna og orðflokks þeirra. Með **lykilorði** er átt við orð eins og *aka* og *feitur* í dæmunum að ofan, orð sem eru þungamiðja sambandanna eða í það minnsta mikilvæg orð innan þeirra. Það varð þó fljótlega ljóst að niðurstöður yrðu markvissari með því að miða mörkunina við ákveðin lykilorð fremur en stafrófsröðina sem orðasamböndin röðuðust eftir.

Í tilrauninni sem hér er til umræðu voru mörkuð 54 orðasambönd með sögninni *aka*, 29 orðasambönd með sögninni *afla*, 75 orðasambönd með lýsingarorðinu *feitur* og 46 orðasambönd með lýsingarorðinu *gláður*. Í 6. kafla verður litið á þessi orðasambönd og afdrif þeirra eftir mörkunarferlið.

### 3 Markari og mörkun

#### 3.1 Markarinn

Þjálfunmarkara fyrir íslensku vartungutækniverkefni á vegum menntamálaráðuneytisins sem var unnið á Orðabók Háskólans frá haustinu 2002 fram í febrúar 2004 (sjá Eirík Rögnvaldsson o.fl. 2002). Til er greinargóð lýsing á verkefninu hjá Sigrúnu Helgadóttur (2004a og 2005) en hún sá um að þjálfar og prófa slík forrit fyrir íslensku. Fimm mismunandi markarar voru þjálfaðir fyrir íslenskan texta og reyndist svonefndur **TnT-markari** gefa besta raun. Markaranum var frá upphafi ætlað að vera fyrri hluti stærra verkefnis, og þegar þetta er ritað er verið að vinna að framhaldsverkefni hans sem ber nafnið **Mörkuð íslensk málheild** (sjá Sigrúnu Helgadóttur 2004b). Það felst í því að búa til markað textasafn (nefnt *tagged corpus* á ensku) sem hefur að geyma fullmarkaða íslenska nútímatexta, en slík textasöfn eru til víða erlendis.

Við mörkun orðasambandanna lá beinast við að nota það markaraforrit sem hafði gefið bestan árangur þegar því var beitt á venjulegan, samfelldan texta, það er að segja TnT-markarann. En þar sem orðasamböndin í þessari tilraun eru ekki samfelldur texti þótti ástæða til að prófa einnig aðra tegund markara því ekki var hægt að vita fyrir-



fram hver árangurinn af mörkun orðasambanda yrði og hvort hann lyti sömu lögmálum og þegar venjulegur texti var markaður.

Markarinn **fnTBL** var því einnig látinn spreyta sig á orðasamböndunum. Í stuttu máli sagt gaf **fnTBL** ekki sérlega góða raun. Mikill munur var á hæfni þessara tveggja markara til að greina orðasamböndin rétt, miklu meiri en þegar þeir voru látnir marka samfelldan texta.<sup>2</sup>

Markarinn **fnTBL** var aðeins notaður fyrst í stað til reynslu. Við greiningu 55 orðasambanda sem byrja á bókstafnum *a*- gerði **fnTBL** villur í 28 orðasamböndum þar sem **TnT** gerði aftur á móti villur í 6 orðasamböndum. Það samsvarar því að á móti hverju rangt greindu orðasambandi hjá **TnT** voru 4.67 rangt greind sambönd hjá **fnTBL**.

Ástæðan fyrir þessum mikla árangursmun er væntanlega sú að **fnTBL**-markarinn þarf talsvert meira textasamhengi en **TnT** til að geta markað rétt og í orðasamböndunum er samhengið oft mjög lítið (orðasamböndin eru oft mjög stutt, stundum eru þau aðeins tvö orð eins og til dæmis *glæða sólskin* og *afla matfanga*). Vegna þessa slaka árangurs **fnTBL**-markarans var horfið frá því að nota hann og ákveðið að nota eingöngu **TnT**-markarann við mörkun orðasambandanna.

### 3.2 Mörkun

Þegar texti er markaður fær hvert orð greiningarstreng sem fylgir því. Við þjálfun markaranna fyrir íslenskan texta var tekið mið af þeirri greiningu sem var notuð í *Íslenskri orðtíðnibók* sem kom út árið 1991 (Jörgen Pind o.fl. 1991). Markaskráin var einnig sú sama og í Orðtíðnibókinni, en með **markaskrá** er átt við lista yfir alla greiningarstrengi (mörk) sem koma fyrir í tilteknu mörkuðu textasafni (sbr. Sigrúnu Helgadóttur 2004a:60). Greiningarstrengurinn samanstendur af skammstöfunum þar sem fyrsti bókstafurinn táknar orðflokk og næstu stafir á eftir tákna málfræðipætti sem eru breytilegir eftir orðflokki. Taka má dæmi til að lýsa þessu. Þegar orðasambandið *aka heilum vagni heim* er markað fæst eftirfarandi greining orðanna:

aka sng heilum lkepsf vagni nkeþ heim aa

Sé lesið úr greiningarstrengjum orðanna stendur þetta:

<sup>2</sup>Hjá Sigrúnu Helgadóttur (2004a:61) kemur fram að **TnT**-markarinn náði 98.14% nákvæmni en **fnTBL** 97.25% nákvæmni þegar markararnir voru látnir greina orðin í textum Orðtíðnibókarinnar, og var þá miðað við einfaldasta mögulega greiningarstreng, þ.e. eingöngu orðflokkinn.

aka sng	sögn í nafnhætti, germynd
heilum lkeþsf	lýsingarorð í karlkyni, eintölu, þágufalli, sterkri beygingu, frumstigi
vagni nkeþ	nafnorð í karlkyni, eintölu, þágufalli
heim aa	atviksorð

Við mörkunina slitnar textinn í sundur og verður að lista þar sem eitt orð er í hverri línu. Mörkunarferlinu er lýst nánar í 4. kafla hér á eftir.

TnT-markarinn var þjálfaður á textasafni Orðtíðnibókarinnar. Hægt er að mæla nákvæmni mörkunar með fjölda réttra greiningarstrengja miðað við heildarfjölda þeirra. Sé aðeins tekið mið af orðflokki en ekki öðrum hlutum greiningarstrengsins, t.d. kyni, tölu og falli nafnorða, náði TnT-markarinn að meðaltali 98,14% nákvæmni við mörkun venjulegra texta (Sigrún Helgadóttir 2004a:61). Til samanburðar mörkuðust orðasambönd með sögninni *aka* með 95,07% nákvæmni (miðað við orðflokk eingöngu en þó með fallstjórnarmerkingu forsetninga og falli persónufornafna og afturbeygða fornafnsins), og orðasambönd með lýsingarorðinu *gladur* mörkuðust með 96,96% nákvæmni.

## 4 Ferli mörkunar

Mörkun texta fer þannig fram að gefin er skipun sem setur af stað mörkunarferlið og skilar það síðan niðurstöðuskrám. Sökum þess að orðasambönd eru ekki venjulegur texti þarf að gera sérstakar ráðstafanir til að mörkunin skili nothæfum niðurstöðum. Fyrst þarf að undirbúa textann, síðan er markaranum beitt á hann og loks er markaði textinn lagaður til svo að unnt sé að lesa úr honum niðurstöðurnar með hægu móti. Mörkun orðasambanda þarf því að fara fram í fimm til sex skrefum.

1. Orðasamböndin eru snyrt og einfölduð.
2. Orðasamböndin eru aðgreind hvert frá öðru með tákningu #.
3. Orðasamböndin eru mörkuð með TnT-markaranum sem skilar þeim sem lóðréttum lista.
4. Orðasamböndin eru færð aftur til fyrra horfs svo að hvert orðasamband sé í sérlínu.
5. Mörkin eru skilin frá orðunum og flutt aftast í línurnar.
6. Mörkin eru einfölduð eftir því sem ástæða þykir til.

1. Orðasamböndin er hægt að einfalda á tvennan hátt. Fyrsta skrefið er að taka út alla svigaliði, enda tákna þeir að innihald þeirra sé valfrjáls viðbót við orðasambandið.

Næsta skref er að huga að tilbrigðum í tíðum sagna (með skástriki á milli tilbrigða) eins og þau koma fram t.d. í orðasambandinu *aka/hafa* (< lengi >) *ekið höllu*. Þegar slík sambönd eru mörkuð kemur vel til greina að einfalda þau svo að aðeins fyrra tilbrigðið standi eftir. Ef sambandið *aka/hafa* (< lengi >) *ekið höllu* er einfaldað til fulls, er svigaliðurinn fyrst fjarlægður svo eftir stendur *aka/hafa ekið höllu*; því næst er sagnmyndin einfölduð í *aka höllu*. Við mörkun orðasambanda var bæði prófað að marka þau óbreytt sem og fullstytt, og hafði einföldun þeirra ekki sýnileg áhrif á nákvæmni mörkunarinnar. Í þessari tilraun var ákveðið að stytta orðasamböndin, en slíkt hlýtur að fara eftir aðstæðum og tilgangi hverju sinni.

2. Skilatóknið # er sett fremst og aftast í orðasambandið til að greina það frá öðrum samböndum. Það er nauðsynlegt vegna þess að hér er ekki um að ræða venjulegar setningar sem byrja á stórum staf og enda á greinarmerki — í orðasamböndunum koma greinarmerki önnur en komma mjög óvíða fyrir, og punktur kemur hvergi fyrir í þeim.

3. Nú er hægt að marka textann með TnT-markaranum og skilar hann niðurstöðum í sérstakar skrár. Ein þeirra hefur að geyma markaðan texta þar sem eitt orð er í línu ásamt markinu:

aka	sng
á	ao
hundasleða	nkeo

4. Til að fullvinna efnið þarf nokkur skref til viðbótar. Færa þarf orðasamböndin aftur til fyrra horfs svo að hvert samband sé í sérlínu. Á þessu stigi er hentugt að setja táknið \_ á milli orðs og marks vegna síðari vinnslu:<sup>3</sup>

```
# aka_sng á_ao hundasleða_nkeo #
# aka_sng < honum_fpkeþ, henni_fpveþ > úr_aþ sporun-
um_nkfþg #
```

Þarna eru þá orðasamböndin aftur komin í lárétta stöðu að viðbættum greiningarstrengjunum sem er skeytt aftan við hvert orð sambandsins.

5. Stundum getur verið hentugt að flytja alla greiningarstrengi orðasambandanna aftast í hverja línu, þ.e. á eftir orðasambandinu. Við það verður orðasambandið læsilegra og einnig auðveldar það áframhaldandi vinnslu á mörkuðu samböndunum, til dæmis ef ætlunin er að

<sup>3</sup>Þakkir fær Auður Rögnvaldsdóttir fyrir að búa til forrit sem gerir þetta tvennt.

láta tölvuna raða lista af orðasamböndum eftir greiningarstrengjunum. Þetta er búið að gera í eftirfarandi dæmum:

```
aka á hundasleða # _sng _ao _nkeo
aka < honum , henni > úr sporunum # _sng < _fpkeþ , _fpveþ > _ap
_nkfbg
```

Ólæsilegasti hluti textans, greiningarstrengirnir, eru nú allir komnir aftast í línurnar í rétttri röð miðað við orðin í sambandinu.

6. Loks eru greiningarstrengirnir einfaldaðir með sérstakri skriftu. Þá er upplagt að nota tækifærið og gera skammstafanirnar gegnsærri og læsilegri. Í töflu 1 má sjá skammstafanir markanna eftir að þeim hefur verið breytt.

### Tafla 1

*Skammstafanir sem eru notaðar í einfaldaðri gerð markanna*

no	nafnorð
lo	lýsingarorð
so	sögn
ao	atviksorð
st	samtenging, nafnháttarmerki
to	töluorð
fn	fornafn (annað en afturbeygt fornafn og persónufornafn)
fn-n	persónufornafn í nefnifalli
fn-a	afturbeygt eða persónufornafn í þolfalli
fn-d	afturbeygt eða persónufornafn í þágufalli
fn-g	afturbeygt eða persónufornafn í eignarfalli
fs-a	forsetning með þolfalli
fs-d	forsetning með þágufalli
fs-g	forsetning með eignarfalli

Orðasamböndin líta svona út eftir að búið er að einfalda greiningarstrengina. Þau eru orðin töluvert læsilegri:

```
# aka á hundasleða # so fs-a no
# aka < honum , henni > úr sporunum # so < fn-d , fn-d > fs-d no
```

## 5 Nákvæmni mörkunar og helstu takmarkanir

Vert er að hafa í huga að mörkun orðasambanda er krefjandi verkefni fyrir markann og er það fyrst og fremst vegna þess hve textasambandið er stutt. TnT-markarinn vinnur á þann hátt að hann skoðar orðið sem hann er að greina ásamt tveimur orðum á undan því. Oft eru orðasamböndin aðeins tvö eða þrjú orð og er árangur hans við

mörkun þeirra þess vegna heldur verri en þegar hann er látinn marka samfelldari texta, eins og búast má við.

Það eru þrír þættir sem hafa mest áhrif á það hvernig tekst til með mörkun orðasambandanna: 1) markaraforritið sem er notað, 2) sú nákvæmni í greiningunni sem krafist er og 3) mörkun með eða án viðbótarorðasafns.

### *Markaraforritið*

Eins og fram kom í kafla 3.1 var mikill munur á nákvæmni í greiningu milli þeirra tveggja markara sem voru prófaðir. Við mörkun orðasambanda gaf TnT-markarinn miklu betri niðurstöður en fnTBL-markarinn.

### *Nákvæmni greiningarstrengja*

Því flóknari sem greiningarstrengir eru þeim mun meiri líkur eru á villum í þeim. Í *Íslenskri orðtíðnibók* er notuð stór markaskrá með 639 mismunandi greiningarstrengjum (sbr. Sigrúnu Helgadóttur 2005:258), en það er sama safn greiningarstrengja og það sem TnT-markarinn notar. Það er þó mjög bundið aðstæðum hversu nákvæma greiningu nauðsynlegt er að hafa og oft er alveg nóg að fá orðflokkinn greindan rétt. Heill greiningarstrengur lýsingarorðs er 6 stafir, greiningarstrengur sagnar í persónuhætti er sömuleiðis 6 stafir og nafnorðs 4–5 stafir. Nóg er að einn stafur í strengnum sé rangur til að markið teljist rangt.

Við mörkun orðasambandanna var miðað við að orðflokkurinn væri almennt nægilegur. Vegna séreinkenna orðasambandaskrárinnar var þó tekið með fall persónufornafna og afturbeygða fornafnsins, og fallstjórn forsetninga var einnig höfð með.

### *Viðbótarorðasafn*

Það skiptir miklu máli hvort markað er með eða án viðbótarorðasafns (lexíkons). Þegar TnT-markarinn var þjálfður var notast við orðasafn *Íslenskrar orðtíðnibókar*, en í því eru 31.876 orð í flettimynd sem koma fyrir í 59.358 orðmyndum (sjá Jörgen Pind o.fl. 1991:2 og Sigrúnu Helgadóttur 2004a:59). Það telst ekki vera stórt orðasafn. Orðasamböndin í skrá Orðabókar Háskólans eru eins og fyrr segir unnin upp úr ritmálssafni þeirrar stofnunar, og þótt þau séu mörg hver afar stutt er þar að finna auðugan orðaforða, eða 57.545 mismunandi lykilorð í flettimynd sinni samkvæmt talningu greinarhöfundar. Með því að nota aðeins orðasafn *Íslenskrar orðtíðnibókar* við mörkun orðasambandanna urðu niðurstöðurnar ekki sérlega góðar þar sem

mikið greindist af óþekktum orðum, sem markarinn gaf til kynna með sérstöku tákni. Lýsingarorð sem markarinn þekkti ekki greindi hann oft sem nafnorð og öfugt, auk þess sem hann gerði fleiri mistök í greiningu orðflokks.

Svo heppilega vill til að stórt viðbótarorðasafn er aðgengilegt. Það eru orðin sem beygð voru fyrir verkefnið *Beygingarlýsing íslensks nútímamáls*, sem var eitt af tungutækniverkefnum menntamálaráðuneytisins, unnið á Orðabók Háskólans. Í útgáfu 2.0 frá 2004 eru beygð 176.030 orð sem alls hafa 4.854.094 mismunandi beygingarmyndir, eins og kemur fram hjá Kristínu Bjarnadóttur, verkefnisstjóra Beygingarlýsingarinnar (2004:23–24).

Þegar markarinn var látinn nota viðbótarorðasafnið auk grunnorðasafnsins batnaði árangur hans stórlega. Allar niðurstöður úr mörkun sem birtar eru á þessum síðum voru fengnar með því að nota viðbótarorðasafnið.

### *Algengar villur*

Algeng villa í greiningu orðasambandanna er sú að nafnháttur sagnar fremst í sambandinu greinist sem nafnorð eða lýsingarorð eins og sjá má á eftirfarandi dæmi. Hér eru greiningarstrengirnir hafðir eins og markarinn skilar þeim:

```
afla_nkeþ soðfisks_nkee
```

Þetta er röng greining, sögnin er greind sem nafnorð í karlkyni, eintölu, þágufalli.

```
afla_sng sér_fpkeþ fjár_nhee
```

Hér er greiningin hins vegar rétt, sögnin er greind sem sögn í nafnhætti.

Annar galli á markaranum er sá að hann á erfitt með að tengja forsetningu við fallorðið sem hún stýrir. Þegar greiningarstrengurinn er einfaldaður á þann hátt að aðeins orðflokkurinn er eftir kemur það ekki svo mjög að sök, en þegar það þarf nákvæmari greiningu er þetta óheppilegt. Dæmi um það þegar markarinn skilur ekki tengslin milli forsetningar og fallorðs má sjá í eftirfarandi orðasambandi:

```
þar_aa fór_sfg3eþ feitur_lkensf biti_nken í_að <Þjóðverja_nkfe-s>  
(þar fór feitur biti í <Þjóðverja>)
```

Í orðasambandinu er *í* greint sem forsetning sem stýrir þágufalli (greining **ap**). *Þjóðverja* er greint sem nafnorð í karlkyni, fleirtölu, eign-

arfalli, sérnafn (greining **nkfe-s**). Þótt ekki sé neinn vafi á því að forsetningin í stýri fallinu á nafnorðinu *Þjóðverja* virðist markarinn ekki geta tengt þetta tvennt saman. Ástæðan er að hluta til sú að orðið *Þjóðverja* er haft innan 'breytiliða' sem trufla samhengið frá sjónarmiði markarans, en þó kemur stundum fyrir að mörkuð fallstjórn forsetningar samsvarar ekki falli orðsins sem kemur beint á eftir, þótt það sé breytliðalaust.

Ruglingur með fallstjórn forsetninga er algeng villa sem öllum mörkorum hættir til að gera, samkvæmt Sigrúnu Helgadóttur (2004a:59). Kafli 4 endaði á því að sýnd voru tvö orðasambönd þar sem búið var að einfalda greininguna. Þau voru þessi:

- # aka á hundasleða # so fs-a no
- # aka < honum , henni > úr sporunum # so < fn-d , fn-d > fs-d no

Fyrri dæmið að ofan, *aka á hundasleða*, fékk greininguna **so fs-a no**. Það er: sögn, forsetning með þolfalli, nafnorð. Þetta er ekki rétt greining, fallstjórn forsetningarinnar hefði átt að greinast þágufall en ekki þolfall þar sem orðið *hundasleða* er í þágufalli. Ljóst er að markarinn getur ekki vitað um hvort fallið er að ræða þar sem *hundasleða* er eins í þolfalli og þágufalli, og þar að auki getur forsetningin á stýrt báðum þessum föllum. Þarna hefði lengra samhengi ekki getað komið markaranum til hjálpar þar sem hann tekur ekki mið af mismunandi merkingu í samböndum eins og *aka á hundasleða* (þágufall) og *aka á ljósastaur* (þolfall), og orðin *hundasleða* og *ljósastaur* eru eins í þolfalli og þágufalli.

Þessi ruglingur á þolfalli og þágufalli er dæmigerð villa sem kemur stundum fyrir í mörkun orðasambandanna.

## 6 Mörkuð orðasambönd

Í þessum kafla eru fjórir listar með orðasamböndum, fyrst með sögunum *aka* og *afla* og svo með lýsingarorðunum *feitur* og *gláður*. Búið er að einfalda samböndin á þann hátt sem lýst er í kafla 4: svigaliðir hafa verið teknir út og tilbrigði í sagnmyndum hafa verið einfölduð. Þar sem villa er í greiningunni er hún höfð með **feitu letri** og jafnframt er merkið \*\* sett á eftir sambandinu.

Nákvæmni mörkunar í orðasamböndunum er mæld á tvennan hátt. Í fyrsta lagi er miðað við orðasambandið sem heild og ef ein eða

fleiri villur koma fyrir í heildarmörkum orðanna er það talið rangt greint. Í öðru lagi er nákvæmnin metin eftir fjölda einstakra orða sem eru rétt greind, en það er sú aðferð sem Sigrún Helgadóttir (2004a og 2005) notar til að mæla nákvæmni mörkunar.

#### 54 orðasambönd með sögninni aka:

- 1 aka léttan # so lo
- 2 aka á hundasleða # so **fs-a** no \*\*
- 3 aka sér úr hlصاصstöðunum # so fn-d fs-d no
- 4 aka of þungu hlصاصi # so ao lo no
- 5 aka < vörunum > á hjólsleðum # so < no > fs-d no
- 6 aka hjólbörum # so no
- 7 aka hjólasleða # so no
- 8 aka < vatninu > á hestkerru # so < no > **fs-a** no \*\*
- 9 aka < steypuefninu > á hestakerru # so < no > fs-d no
- 10 aka höllu fyrir < honum, henni > # so lo fs-d < fn-d, fn-d >
- 11 aka höllu # so lo
- 12 aka í hágír # so fs-d no
- 13 aka **greitt** # so **so** \*\*
- 14 aka í fereykisvagni # so fs-d no
- 15 aka á þvottabretti # so **fs-a** no \*\*
- 16 aka á völlum # so fs-a no
- 17 að aka á völlum # st so fs-a no
- 18 aka um vegleysur # so fs-a no
- 19 eiga heilum vagni heim að aka # so lo no ao st so
- 20 aka heilum vagni heim # so lo no ao
- 21 aka á tún # so fs-a no
- 22 < honum, henni > ekst < illa > í tauma # < fn-d, fn-d > so < ao > fs-a no
- 23 < allt > ekst í tauma # < fn > so fs-a no
- 24 < þetta > ekst < honum, henni > í tauma # < fn > so < fn-d, fn-d > **fs-d** no \*\*
- 25 < sá uggur > ekst < honum, henni > í tauma # < fn no > so < fn-d, fn-d > **fs-d** no \*\*
- 26 aka spölkorn # so no
- 27 aka < honum, henni > úr sporunum # so < fn-d, fn-d > fs-d no
- 28 akast úr sporunum # so fs-d no
- 29 aka sleðanum # so no
- 30 aka í sleða # so fs-d no
- 31 aka skrykkjótt # so lo
- 32 aka í skrautvagni # so fs-d no
- 33 aka seglum # so no
- 34 aka seglum eftir veðri # so no fs-d no



- 35 aka seglunum eftir veðri # so no fs-d no  
 36 aka seglunum eftir veðrinu # so no fs-d no  
 37 aka seglunum eftir vindinum # so no fs-d no  
 38 aka seglum eftir vindi # so no fs-d no  
 39 < honum, henni > er nærri ekið # < fn-d, fn-d > so ao so  
 40 aka mykju # so no  
 41 aka í móinn # so fs-a no  
 42 aka í léttivagni # so fs-d no  
 43 aka léttivagni # so no  
 44 < þetta > ekur < öllu > á kaldan klaka # < fn > so < fn > fs-a lo no  
 45 < þeir, þær, þau > akast á erindum um < þetta > < þangað til-S > # < fn, fn-n, fn-n > so fs-d no fs-a < fn > < ao tp > \*\*  
 46 aka í drosju # so fs-d lo  
 47 aka dráttarvél # so no  
 48 aka í dagvagni # so fs-d no  
 49 láta ekki aka sér á bug # so ao so fn-d fs-a no  
 50 aka bug á < keisarann, Frakka > # so no fs-a < no, no >  
 51 aka < honum, henni > á bug # so < fn-d, fn-d > fs-a no  
 52 aka angri á bug # so no fs-a no  
 53 aka í bifreið # so fs-d no  
 54 aka barnakerru # so no

Orðasamböndin með *aka* markast þannig að af 54 samböndum greinast 47 rétt en villa kemur fyrir í 7 þeirra. Það samsvarar 87,04% nákvæmni sé miðað við heil orðasambönd en 95,07% nákvæmni sé miðað við einstök orð innan sambandanna.

## 29 orðasambönd með sögninni *afla*:

- 1 **afla** fiskjar # **no** no \*\*  
 2 afla sér fæðslu # so fn-d no  
 3 afla fóðurforða # so no  
 4 afla sér fjár # so fn-d no  
 5 afla sér þekkingar # so fn-d no  
 6 afla sér viðar # so fn-d so  
 7 **afla** viðskiptasambanda # **no** no \*\*  
 8 **afla** vetrarforða # **no** no \*\*  
 9 afla sér veltufjár # so fn-d no  
 10 < þetta > **aflar** vanheilsu # < fn > **no** no \*\*\*  
 11 afla sér vanblessunar # so fn-d no  
 12 afla sér tiltrúar # so fn-d no  
 13 **afla** sölva # **no** so \*\*\*  
 14 afla soðningar # so no  
 15 **afla** soðmatar # **no** no \*\*  
 16 **afla** soðfisks # **no** no \*\*  
 17 **afla** til soðs # **no** fs-g no \*\*

- 18 það aflast til soðs # fn-n so fs-g no
- 19 **afla** næringar sinnar # **no** no fn \*\*
- 20 **afla** nýmetis # **no** no \*\*
- 21 afla sér menntunar # so fn-d no
- 22 **afla** matfanga # **no** no \*\*
- 23 afla sér fjár og mannvirðingar # so fn-d no st no
- 24 afla sér lífsuppeldis # so fn-d no
- 25 afla sér kristindómsþekkingar # so fn-d no
- 26 **afla** til kola # **no** fs-g no \*\*
- 27 afla sér kennaramenntunar # so fn-d no
- 28 afla sér doktorsnafnbótar # so fn-d no
- 29 afla sér **bjargar** # so fn-d **so** \*\*\*

Af 29 orðasamböndum með *afla* markast 13 vitlaust en 16 rétt. Þetta er ekki góður árangur, aðeins 55,17% nákvæmni miðað við orðasambandið sem heild en 82,5% miðað við einstök orð. Það stafar einkum af því að markarinn ruglar saman sögninni *afla* og nafnorðsmyndinni *afla*, en þrjár mismunandi beygingarmyndir af karlkynsorðinu *afla* eru *afla* (ef orðið er ekki notað í fleirtölu). Hér þarf að grípa til eftirvinnslu sem felst í því að breyta marki orðs sem endar á *-a* í upphafi orðasambands úr nafnorði (no) í sögn (so).

Eftir þessa leiðréttingaraðgerð fremst í línu verða aðeins 3 villur eftir í mörkuninni og nákvæmnin mælist þá 89,65% miðað við orðasamböndin sem heild en 96,25% miðað við einstök orð.

Villurnar þrjár sem eftir eru eru þá þessar (merktar með 3 stjörnum): 1) í sambandinu < þetta > *aflar vanheilsu* er miðorðið greint sem nafnorð í stað sagnar. 2) Sambandið *afla sölvu* er rangt greint því *sölvu* er sagt vera sögn en ekki nafnorð, og 3) í sambandinu *afla sér bjargar* er síðasta orðið greint sem sögn í stað nafnorðs.

### 75 orðasambönd með lýsingarorðinu *feitur*:

- 1 < éta yfir sig þá sjaldan > hnífur < þeirra > kemur í feitt ket # < so fs-a fn-a ao ao > no < fn-e > so fs-a lo no
- 2 **feiti** kláði # **so** no \*\*
- 3 feit jörð # lo no
- 4 feit leirjörð # lo no
- 5 feit **mjól**k # lo **lo** \*\*
- 6 feit mold # lo no
- 7 feit moldarjörð # lo no
- 8 feit moldjörð # lo no
- 9 feit olía # lo no
- 10 feit sauðarföll # lo no

- 11 feit staða # lo no  
 12 feit steik # lo no  
 13 feitt embætti # lo no  
 14 feitt ket # lo no  
 15 feitt kjöt # lo no  
 16 feitt letur # lo no  
 17 feitt og holdasamt < fé > # lo st lo < no >  
 18 **feitt** og holdsamt < fé > # **so** st lo < no > \*\*  
 19 feitt prestakall # lo no  
 20 feitt prófastsdæmi # lo no  
 21 feitt slátur # lo no  
 22 feitu níurnar # lo no  
 23 feitur biti # lo no  
 24 feitur dráttur # lo no  
 25 feitur eldishestur # lo no  
 26 feitur jarðvegur # lo no  
 27 feitur moldarjarðvegur # lo no  
 28 feitur naflastrengur # lo no  
 29 feitur og holdlaginn < uxi > # lo st lo < no >  
 30 finna ekki feitan gölt að flá < í íslenskum heimildum > # so ao lo no st lo  
 < fs-d lo no >  
 31 finna hvar er feitt á stykkinu # so ao so lo fs-d no  
 32 finna < lítið > feitt á stykki # so < lo > lo fs-d no  
 33 < finna, sjá > það sem feitt er á stykkinu # < so, so > fn st lo so fs-d no  
 34 finna **það** sem **feitt** er á stykkinu # so **fn-n** st so so fs-d no \*\*  
 35 flá ekki feitan kött # so ao lo no  
 36 flá feitan grís # so lo no  
 37 flá feitan gölt # so lo no  
 38 flá feitan gölt af < síldar kaupunum > # so lo no fs-d < no >  
 39 < fuglinn > er feitur og bústinn # < no > so lo st lo  
 40 **góna á** < hann, hana > eins og hundur á feitt ket # **no fs-d** < fn-a, fn-a >  
 ao st no fs-a lo no \*\*  
 41 **góna á** < hann, hana > eins og hundur á feitt kjöt # **no fs-d** < fn-a, fn-a >  
 ao st no fs-a lo no \*\*  
 42 hafa ekki feitan gölt að flá # so ao lo no st so  
 43 hafa ekki um feitt að sleikja # so ao fs-a lo st so  
 44 hafa < varla lengi > um feitt að sleikja # so < ao ao > fs-a lo st so  
 45 < hesturinn > er feitur og strykinn # < no > so lo st lo  
 46 < hér > er ekki feitan gölt að flá # < ao > so ao lo no st so  
 47 < hér > er ekki um feitt að sleikja # < ao > so ao fs-a lo st so  
 48 < hér > er **feitt** á stykkjunum # < ao > so **so** fs-d no \*\*  
 49 hnífur < hans, hennar > kemst í feitt # no < fn-e, fn-e > so fs-a lo

- 50 hnífur < hans, hennar > kemur í feitt # no < fn-e, fn-e > so fs-a lo  
 51 < honum, henni > fellur **feitt** flesk í ketil # < fn-d, fn-d > so **so** no fs-a no  
 \*\*  
 52 kjósa **heldur** magra sætt en feitan prósess # so **so** lo no st lo no \*\*  
 53 ríða feitum hesti < þaðan, frá þeim viðskiptum > # so lo no < ao, fs-d fn  
 no >  
 54 < síldin > er feit á fiskinn # < no > so lo fs-a no  
 55 sjá < þar > feitan slag á borði # so < ao > lo no fs-d no  
 56 < ufsinn > er feitur á fisk # < no > so lo fs-a no  
 57 vera feitur og árlegur # so lo st lo  
 58 vera feitur og hjassalegur # so lo st lo  
 59 vera feitur og ístrumikill # so lo st lo  
 60 vera feitur og líkamapungur # so lo st lo  
 61 vera feitur og pattaralegur # so lo st lo  
 62 vera feitur og sællegur # so lo st lo  
 63 vera feitur og þriflegur # so lo st lo  
 64 vera feitur svoli # so lo no  
 65 það er ekki feitan gölt að flá # fn-n so ao lo no st so  
 66 það er ekki feitt að sleikja # fn-n so ao lo st so  
 67 það er ekki um feitt að sleikja # fn-n so ao fs-a lo st so  
 68 það er < sjaldnast > feitum hesti heim að ríða # fn-n so < ao > lo no ao st  
 so  
 69 < þar > er af feitum bita að klípa # < ao > so fs-d lo no st so  
 70 < þar > er ekki feitan gölt að flá # < ao > so ao lo no st so  
 71 < þar > er feitt í búi # < ao > so lo fs-d no  
 72 þar fór feitur biti í < Þjóðverja > # ao so lo no **fs-d** < no-s > \*\*  
 73 < þar, þarna > er **feitt** á stykkinu # < ao, ao > so **so fs-d** no \*\*  
 74 < þessir atburðir > eru feitir til frásagna # < fn no > so lo fs-g no  
 75 þrýstinn og feitur < sauður > # lo st lo < no >

Orðasamböndin með *feitur* markast þannig að af 75 orðasamböndum voru 11 rangt greind. Það samsvarar 85,33% nákvæmni í mörkun sé miðað við sambandið í heild. Alls fengu 14 orð ranga greiningu og sé miðað við einstök orð í textanum er nákvæmnin 95,88%.

#### 46 orðasambönd með lýsingarorðinu glaður:

- 1 gera sér einn glaðan dag # so fn-d fn lo no
- 2 gera sér glaðan dag # so fn-d lo no
- 3 glaða sólskin # lo no
- 4 glaða tungsljós # lo no
- 5 glaður eldur # lo no
- 6 glatt sólskin # lo no
- 7 glatt tunglskin # lo no
- 8 glatt veður # lo no
- 9 gleðjast með glöðum # so fs-d lo

- 10 glöð birta # lo no  
 11 glöð stjarna # lo no  
 12 góðan daginn, glaðan haginn # lo no, lo no  
 13 grípa glaðan á < öllu er getur glætt vonir þeirra um betri umskipti > #  
 so lo **fs-a** < fn so so so no fn-e fs-a lo no > \*\*  
 14 < hér > **brennur** < ekki lengur > glatt ljós í kolu # < ao > **no** < ao ao > lo  
 no fs-d no \*\*  
 15 með < **þeim** > brennur ekki glatt ljós í kolu < á síðustu árum > # fs-d <  
**fn** > so ao lo no fs-d no < fs-d lo no > \*\*  
 16 < mér > er **glatt** í geði # < fn-d > so **so** fs-d no \*\*  
 17 sjá < varla nokkurn > glaðan dag < upp frá þeim tíma > # so < ao fn > lo  
 no < ao fs-d fn no >  
 18 < taka **þeim** kostum > með glöðu geði # < so **fn-d** no > fs-d lo no \*\*  
 19 vera brosmildur og glaður # so lo st lo  
 20 vera glaður af víni # so lo fs-d no  
 21 vera glaður í anda # so lo fs-d no  
 22 vera glaður og fagnaðarfullur # so lo st lo  
 23 vera glaður og glámmikill # so lo st lo  
 24 vera glaður og góðfældur # so lo st lo  
 25 vera glaður og góðkátur # so lo st lo  
 26 vera glaður og góðsinnaður # so lo st lo  
 27 vera glaður og hreifur # so lo st lo  
 28 vera glaður og kátur # so lo st lo  
 29 vera glaður og léttlyndur # so lo st lo  
 30 vera glaður og líflegur # so lo st lo  
 31 vera glaður og rósamur # so lo st lo  
 32 vera glaður og skemmtinn # so lo st lo  
 33 vera glaður og skrafhreifinn # so lo st lo  
 34 vera glaður og spaugsamur # so lo st lo  
 35 vera glaður og svipléttur # so lo st lo  
 36 vera glaður og viðmótsgóður # so lo st lo  
 37 vera glaður og viðmótsþýður # so lo st lo  
 38 vera glaður yfir < þessu > # so lo fs-d < fn >  
 39 vera himinljómandi glaður # so ao lo  
 40 vera hjartanlega glaður # so ao lo  
 41 vera hýrlyndur og glaður # so lo st lo  
 42 vera innilega glaður # so ao lo  
 43 vera óumræðilega glaður # so ao lo  
 44 verða glaður við að < þessum bágingum létti af > # so lo fs-a st < fn no  
 so fs-d >  
 45 það er glaða sólskin # fn-n so lo no  
 46 < þar > **brennur** glatt ljós í kolu # < ao > **no** lo no fs-d no \*\*

Orðasamböndin með *gláður* markast þannig að af 46 orðasamböndum voru 6 rangt greind. Það samsvarar 86,94% nákvæmni í mörkun sé miðað við sambandið í heild, en sé miðað við einstök orð er nákvæmnin 96,96%.

## 7 Röðun sambanda og kortlagning orða

Eftir að orðasamböndin hafa verið mörkuð má fara að huga að því að nota mörkunina á einhvern hátt. Það má til dæmis prófa að raða orðasamböndunum, ekki í hefðbundna stafrófsröð heldur eftir greiningarstrengjunum, þ.e. raða þeim eftir því sem kemur á eftir tákningu #. Þetta er gert til þess að reyna að fá fram eins konar kortlagningu af helstu setningargerðum með ákveðnu orði.

### 7.1 Orðasambönd með 'feitur'

Fyrir neðan eru 60 orðasambönd með lýsingarorðinu *feitur*. Búið er að leiðrétta villur í mörkuninni og raða samböndunum eftir greiningarstrengjunum.

<þar, þarna> er feitt á stykkinu	# <ao, ao> so lo fs-d no
<hér> er ekki um feitt að sleikja	# <ao> so ao fs-a lo st so
<hér> er ekki feitan gölt að flá	# <ao> so ao lo no st so
<þar> er ekki feitan gölt að flá	# <ao> so ao lo no st so
<þar> er af feittum bita að klípa	# <ao> so fs-d lo no st so
<hér> er feitt á stykkjunum	# <ao> so lo fs-d no
<þar> er feitt í búi	# <ao> so lo fs-d no
þar fór feitur biti í <þjóðverja>	# ao so lo no fs-a <no-s>
<honum, henni> fellur feitt flekk í ketil	# <fn-d, fn-d> so lo no fs-a no
<þessir atburðir> eru feitir til frásagna	# <fn no> so lo fs-g no
það er <sjaldnast> feittum hesti heim að ríða	# fn-n so <ao> lo no ao st so
það er ekki um feitt að sleikja	# fn-n so ao fs-a lo st so
það er ekki feitan gölt að flá	# fn-n so ao lo no st so
það er ekki feitt að sleikja	# fn-n so ao lo st so
feit moldarjörð	# lo no
feit moldjörð	# lo no
feit mold	# lo no
feit olía	# lo no

feit staða	# lo no
feit steik	# lo no
feitt embætti	# lo no
feitt ket	# lo no
feitt kjöt	# lo no
feitt letur	# lo no
feitt prestakall	# lo no
feitt prófastsdæmi	# lo no
feitur eldishedstur	# lo no
feitur jarðvegur	# lo no
feitt og holdasamt <fé>	# lo st lo <no>
feitt og holdsamt <fé>	# lo st lo <no>
feitur og holdlaginn <uxi>	# lo st lo <no>
þrýstinn og feitur <sauður>	# lo st lo <no>
hnífur <hans, hennar> kemst í feitt	# no <fn-e, fn-e> so fs-a lo
hnífur <hans, hennar> kemur í feitt	# no <fn-e, fn-e> so fs-a lo
<síldin> er feit á fiskinn	# <no> so lo fs-a no
<ufsinn> er feitur á fisk	# <no> so lo fs-a no
<fuglinn> er feitur og bústinn	# <no> so lo st lo
hafa <varla lengi> um feitt að sleikja	# so <ao ao> fs-a lo st so
hafa ekki um feitt að sleikja	# so ao fs-a lo st so
flá ekki feitan kött	# so ao lo no
sjá <þar> feitan slag á borði	# so <ao> lo no fs-d no
finna ekki feitan gölt að flá <í íslenskum heimildum>	# so ao lo no st lo <fs-d lo no>
kjósa heldur magra sætt en feitan prósess	# so ao lo no st lo no
hafa ekki feitan gölt að flá	# so ao lo no st so
finna hvar er feitt á stykkinu	# so ao so lo fs-d no
finna það sem feitt er á stykkinu	# so fn-a st lo so fs-d no
sjá það sem feitt er á stykkinu	# so fn-a st lo so fs-d no
góna á <hann, hana> eins og hundur á feitt ket	# so fs-a <fn-a, fn-a> ao st no fs-a lo no
góna á <hann, hana> eins og hundur á feitt kjöt	# so fs-a <fn-a, fn-a> ao st no fs-a lo no
finna <lítið> feitt á stykki	# so <lo> lo fs-d no
flá feitan grís	# so lo no
flá feitan gölt	# so lo no
vera feitur svoli	# so lo no
ríða feittum hesti <þaðan, frá þeim viðskiptum>	# so lo no <ao, fs-d fn no>

flá feitan gölt af <síldarkaupunum>	# so lo no fs-d <no>
vera feitur og árlegur	# so lo st lo
vera feitur og hjassalegur	# so lo st lo
vera feitur og pattaralegur	# so lo st lo
vera feitur og sællegur	# so lo st lo
vera feitur og þriflegur	# so lo st lo

Næsta skref er að sleppa orðasambandinu svo greiningin ein standi eftir. Línur sem hafa eins greiningu eru aðeins sýndar einu sinni, og talan á undan markastrengnum táknar hér tíðni. Við þetta hefur línunum fækkað úr 60 í 32:

- 1 < ao, ao > so lo fs-d no
- 1 < ao > so ao fs-a lo st so
- 2 < ao > so ao lo no st so
- 1 < ao > so fs-d lo no st so
- 2 < ao > so lo fs-d no
- 1 ao so lo no fs-a < no-s >
- 1 < fn-d, fn-d > so lo no fs-a no
- 1 < fn no > so lo fs-g no
- 1 fn-n so < aoe > lo no ao st so
- 1 fn-n so ao fs-a lo st so
- 1 fn-n so ao lo no st so
- 1 fn-n so ao lo st so
- 14 lo no
- 4 lo st lo < no >
- 2 no < fn-e, fn-e > so fs-a lo
- 2 < no > so lo fs-a no
- 1 < no > so lo st lo
- 1 so < ao ao > fs-a lo st so
- 1 so ao fs-a lo st so
- 1 so ao lo no
- 1 so < ao > lo no fs-d no
- 1 so ao lo no st lo < fs-d lo no >
- 1 so ao lo no st lo no
- 1 so ao lo no st so
- 1 so ao so lo fs-d no
- 2 so fn-a st lo so fs-d no
- 2 so fs-a < fn-a, fn-a > ao st no fs-a lo no
- 1 so < lo > lo fs-d no
- 3 so lo no
- 1 so lo no < ao, fs-d fn no >
- 1 so lo no fs-d < no >
- 5 so lo st lo



Í listanum að ofan má sjá á tíðnitölunni hvaða setningargerð með lýsingarorðunu *feitur* er algengust í orðasambandaskránni, þ.e. **lo no** (lýsingarorð nafnorð). Fjórtán orðasambönd af þeim sextíu sem hér var raðað hafa þessa setningargerð, en fulltrúi hennar er til dæmis sambandð *feitt prófastsdæmi*.

Næstalgengasta setningargerðin er **so lo st lo** (sögn lýsingarorð samtenging lýsingarorð) sem kemur fimm sinnum fyrir, og er það til dæmis orðasambandið *vera feitur og sællegur*.

Þriðja algengasta mynstrið er **lo st lo < no >** sem kemur fjórum sinnum fyrir. Um er að ræða setningargerðina **lýsingarorð samtenging lýsingarorð < nafnorð >** þar sem nafnorðið er innan breytliða. Dæmi um slíkt orðasamband er *þrýstinn og feitur < sauður >*.

## 7.2 Orðasambönd með 'aka'

Í 6. kafla voru birt 54 orðasambönd með sögninni *aka*. Nú verður litið aftur á þessi sömu sambönd eftir að búið er að leiðrétta villurnar í mörkuninni og raða þeim á sama hátt og samböndunum með *feitur*, þ.e. eftir greiningarstrengnum. Til að einfalda samböndin enn meira er breytliðurinn *< honum, henni >* stytur í *< honum >*.

- 1 < honum > ekst < illa > í tauma # < fn-d > so < ao > fs-a no
- 2 < honum > er nærri ekið # < fn-d > so ao so
- 3 < sá uggur > ekst < honum > í tauma # < fn no > so < fn-d > fs-a no
- 4 < þeir > akast á erindum um < þetta > < þangað til-S > # < fn-n > so fs-d no fs-a < fn > < ao \_tp >
- 5 < allt > ekst í tauma # < fn > so fs-a no
- 6 < þetta > ekst < honum > í tauma # < fn > so < fn-d > fs-a no
- 7 < þetta > ekur < öllu > á kaldan klaka # < fn > so < fn > fs-a lo no
- 8 aka of þungu hlasi # so ao lo no
- 9 láta ekki aka sér á bug # so ao so fn-d fs-a no
- 10 aka < honum > á bug # so < fn-d > fs-a no
- 11 aka < honum > úr sporunum # so < fn-d > fs-d no
- 12 aka sér úr hlصاصtöðunum # so fn-d fs-d no
- 13 aka á tún # so fs-a no
- 14 aka á völl # so fs-a no
- 15 aka í móinn # so fs-a no
- 16 aka um vegleysur # so fs-a no
- 17 aka í drosju # so fs-d lo

- 18 aka á hundasleða # so fs-d no  
 19 aka á þvottabretti # so fs-d no  
 20 aka í bifreið # so fs-d no  
 21 aka í dagvagni # so fs-d no  
 22 aka í fereykisvagni # so fs-d no  
 23 aka í hágír # so fs-d no  
 24 aka í léttivagni # so fs-d no  
 25 aka í skrautvagni # so fs-d no  
 26 aka í sleða # so fs-d no  
 27 akast úr sporunum # so fs-d no  
 28 aka greitt # so lo  
 29 aka höllu # so lo  
 30 aka léttan # so lo  
 31 aka skrykkjótt # so lo  
 32 aka höllu fyrir < honum > # so lo fs-d < fn-d >  
 33 aka heilum vagni heim # so lo no ao  
 34 eiga heilum vagni heim að aka # so lo no ao st so  
 35 aka barnakerru # so no  
 36 aka dráttarvél # so no  
 37 aka hjólasleða # so no  
 38 aka hjólbörum # so no  
 39 aka léttivagni # so no  
 40 aka mykju # so no  
 41 aka seglum # so no  
 42 aka sleðanum # so no  
 43 aka spölkorn # so no  
 44 aka angri á bug # so no fs-a no  
 45 aka bug á < keisarann , Frakka > # so no fs-a < no , no >  
 46 aka < steypuefninu > á hestakerru # so < no > fs-d no  
 47 aka < vatninu > á hestkerru # so < no > fs-d no  
 48 aka < vörunum > á hjólsleðum # so < no > fs-d no  
 49 aka seglum eftir veðri # so no fs-d no  
 50 aka seglum eftir vindi # so no fs-d no  
 51 aka seglunum eftir veðrinu # so no fs-d no  
 52 aka seglunum eftir veðri # so no fs-d no  
 53 aka seglunum eftir vindinum # so no fs-d no  
 54 að aka á völl # st so fs-a no

Þegar orðasamböndunum með sögninni *aka* að ofan er raðað eftir greiningarstrengnum sést vel kerfið sem þau mynda. Sem dæmi um það hvernig orðasamböndin hópa sig saman fá sambönd 49–53 öll sömu greininguna, en þessi fimm orðasambönd eru mismunandi tilbrigði af sama orðtakinu, *aka seglum eftir veðri*. Í línunum 13–16 eru sambönd með forsetningarlið í þolfalli, og í línunum 17–26 er sögnin með forsetningarlið í þágufalli. Þetta er einmitt aðalástæðan fyrir því að álitíð var nauðsynlegt að taka með fallstjórn forsetninga í greiningar-

strengnum þótt hann væri einfaldaður niður í eintóman orðflokk alls staðar annars staðar — nema í sumum fornöfnum af þessari sömu ástæðu.

## 8 Mörkun orðabókartexta

Orðasambönd af því tagi sem hér eru til umræðu eru ekki venjulegur texti sem verður á vegi manna daglega. Þó minnir hann í mikilvægum atriðum á annan texta sem margir þekkja vel, þ.e. skýringartexta orðabóka. Þessar tvær textagerðir eiga ýmislegt sameiginlegt og má nefna eftirfarandi atriði:

- Textinn er ekki alltaf (og raunar sjaldnast) heilar setningar.
- Fremsta orðið stendur yfirleitt í flettimynd sinni, þ.e. sögn fremst stendur í nafnhætti (t.d. *aka á tún* í orðasambandaskránni) og nafnorð stendur í nefnifalli eintölu (t.d. *aðdráttarkraftur jarðar* í orðasambandaskránni).

Til að skýra þetta nánar má taka sögnina *aka* eins og hún er skýrð í *Íslenskri orðabók*, 3. útgáfu 2002.

Hér er aðeins birtur skýringartextinn við *aka* ásamt skýringartexta við orðasambönd og dæmi innan sagnarinnar. Það sem verið er að skýra er ekki sýnt, þ.e. sögnin sjálf og feitlettruð og skálettruð sambönd og dæmi.

- 1 keyra, fara (flytja) á vagni, sleða e.þ.h.
- 2 nota hest sem dráttardýr og stjórna honum með aktaumum
- 3 flytja mykju (áður í kláfum á hesti) til áburðar á tún
- 4 þ.e. voru svo síðir á hestinum að þeir sópuðu snjónum með sér
- 5 hreyfa hægt, mjaka til
- 6 skaka sig örlítið, kiða sér
- 7 láta í minni pokann (fyrir e-m)
- 8 færa seglin (eftir vindstöðu og stefnu skipsins)
- 9 hegða sér eftir aðstæðum, leika tveim skjöldum
- 10 e-r er í klípu
- 11 hreyfast hægt, mjakast til
- 12 hrekja e-n
- 13 metast á, deila
- 14 malda í móinn
- 15 koma e-m í klípu
- 16 e-ð stöðvast hjá e-m

- 17 hreyfa sig smávegis
- 18 nudda e-m til verks
- 19 hörfa
- 20 sleppa við e-ð
- 21 e-r er alltaf jafn óheppinn
- 22 losna úr klípu

Tölurnar í fyrri dálknum eru settar hér til hagræðis en eru vitanlega ekki hluti af skýringartextanum. Þegar textinn var markaður voru sömu reglur notaðar og við mörkun orðasambanda, þ.e. orðflokkurinn var látinn nægja en þó tekið með fall persónuformafna og afturbeygða fornafnsins sem og fallstjórn forsetninga.

Hér að neðan má sjá hvernig mörkin líta út eftir að þau hafa verið einfölduð. Röng mörk eru með feitu lettri. Línur þar sem röng mörk koma fyrir enda á \*\*.

- 1 keyra, fara (flytja) á vagni, sleða e.þ.h. # so, so (so) fs-d no, no x
- 2 nota hest sem dráttardýr og stjórna honum með aktaumum # so no st no st so fn-d fs-d no
- 3 flytja mykju (áður í kláfum á hesti) til áburðar á tún # so no (ao fs-d no fs-d no) fs-g no fs-a no
- 4 þ.e. voru svo síðir á hestinum að þeir sópuðu snjónum með sér # **no** so ao lo fs-d no st fn-n so no fs-d fn-d \*\*
- 5 **hreyfa hægt**, mjaka til # **no lo**, so fs-g \*\*
- 6 skaka sig örlítið, kiða sér # so fn-a ao, so fn-d
- 7 láta í minni pokann (fyrir **e-m**) # so fs-a lo no (fs-d **no**) \*\*
- 8 færa seglin (eftir vindstöðu og stefnu skipsins) # so no (fs-d no st no no)
- 9 hegða sér eftir aðstæðum, leika tveim skjöldum # so fn-d fs-d no, so to no
- 10 **e-r** er í klípu # **no** so fs-d no \*\*
- 11 hreyfast hægt, mjakast til # so ao, so fs-g
- 12 hrekja **e-n** # so **lo** \*\*
- 13 metast á, deila # so ao, so
- 14 malda í móinn # so **fs-d** no \*\*
- 15 koma **e-m** í klípu # so **no fs-d** no \*\*
- 16 **e-ð** stöðvast hjá **e-m** # **no** so fs-d **lo** \*\*
- 17 hreyfa sig smávegis # so fn-a ao
- 18 nudda **e-m** til verks # so **no** fs-g no \*\*
- 19 hörfa # so
- 20 sleppa við **e-ð** # so fs-a **no** \*\*
- 21 **e-r** er alltaf jafn óheppinn # so **ao ao lo** \*\*
- 22 losna úr klípu # so fs-d **ne**>

Það blasir við eftir þessa mörkun að sá háttur sem hefur tíðkast í orðabókum að nota skammstafanir á óákveðna fornafninu *einhver*: *e-r*, *e-n* o.s.frv., er ekki skiljanlegur markaranum. Í þessum 22 línun kemur skammstöfun á *einhver* fyrir í ýmsum föllum níu sinnum og er mark-ið í öllum tilvikum rangt. *Einhver* er 7 sinnum greint sem nafnorð og tvisvar sem lýsingarorð.

Auðvitað væri hægt að bæta árangur markarans með því að lengja þessar skammstafanir fyrir mörkun þótt slíkt fæli í sér óæskilegt og tafsamt inngríp í textann. Önnur og betri leið væri að bæta skammstöfuninni *e-r* í öllum myndum inn í viðbótarorðasafnið.

Alls koma fyrir 14 villur í greiningu þessa orðabókartexta, þar af eru fimm villur af öðru tagi en skammstafanir á *einhver*. Það verður að teljast ágætur árangur því eins og í orðasamböndunum er í þessum stuttu textum lítið setningarlegt samhengi fyrir markarann að styðjast við.

## 9 Lokaorð

Þessi tilraun til að greina orðasambönd málfræðilega náði til um 200 orðasambanda sem höfðu eitt þessara fjögurra lykilorða: *afla*, *aka*, *feitur* og *gladur*. Tilgangurinn var upphaflega sá að prófa nýtt tól, TnT-markarann, og athuga hvernig gengi að marka orðasambönd með honum því ekki hafði áður verið reynt að greina þetta sérstaka textaform á vélrænan hátt í íslensku.

Árangur mörkunarinnar varð ekki fjarri þeim árangri sem hefur fengist þegar valinn samfelldur texti er markaður þótt hann væri að vísu heldur verri eins og gert hafði verið ráð fyrir. Nefna má tölur til samanburðar: TnT-markarinn náði að meðaltali 98,14% nákvæmni við mörkun venjulegra texta (miðað við orðflokk eingöngu) en orðasambönd með sögninni *aka* mörkuðust með 95,07% nákvæmni (miðað við orðflokk eingöngu en þó með fallstjórnarmerkingu forsetninga og falli persónufornafna og afturbeygða fornafnsins). Orðasambönd með lýsingarorðinu *gladur* mörkuðust með 96,96% nákvæmni.

Tveir þættir hafa mest áhrif á árangur mörkunarinnar: markarinn sem er notaður og mörkun með eða án viðbótarorðasafns. Tveir markarar voru prófaðir á orðasamböndunum og gaf TnT-markarinn mun

betri niðurstöður en fnTBL-markarinn. Notað var stórt viðbótarorðasafn þegar orðasamböndin voru mörkuð því þegar minna orðasafn var notað (orðasafn Orðtíðnibókarinnar) varð árangur mörkunarinnar með öllu óviðunandi.

Villur sem koma oftast fyrir við mörkun orðasambanda eru í fyrsta lagi röng fallstjórn forsetninga og í öðru lagi ruglingur með orðflokk sagnorðs sem er fremst í sambandinu. Dæmi um seinna atriðið er *afla til soðs* þar sem *afla* er greint sem nafnorð í stað sagnar.

Prófað var að nota niðurstöður úr greiningunni til að hópa saman líkum orðasamböndum. Það var gert með því að raða orðasamböndunum með *feitur* og *aka* eftir greiningarstrengnum. Þetta leiddi í ljós ákveðin mynstur í setningargerðum viðkomandi orða, t.d. sást að algengasta mynstrið með *feitur* í orðasambandaskránni er **lo no** (lýsingarorð nafnorð) sem er m.a. að finna í orðasambandinu *feitit prófastsdæmi*.

Tilraunin náði einnig að dálitlu leyti til skýringartexta *Íslenskrar orðabókar* og var árangur þeirrar mörkunar svipaður og í orðasamböndunum þótt ekki væru gerðar nákvæmar mælingar á niðurstöðunum.

Ljóst er að mörkun orðasambanda gefur áhugaverðar niðurstöður. Það má líta á hana sem leið til að fá yfirlit um helstu setningargerðir orða, og eru þá sagnir sérstaklega athyglisverðar þar sem setningargerðir þeirra eru oft mjög margbreytilegar. Slík aðkoma að sagnakerfinu væri ólík þeirri sem fæst við það að skoða flóknar sagnir eins og þær eru settar fram í orðabókum, og vísast þar einkum til meðferðar sagna í *Íslenskri orðabók* (2002) og *Orðastað* Jóns Hilmars Jónssonar (2001). En efnið í orðasamböndunum getur einnig nýst við greiningu á setningargerðum orða af öðrum orðflokkum, hægt er að kortleggja orð af öllum orðflokkum með þessari aðferð, sbr. lýsingarorðið *feitur* sem fjallað hefur verið um.

## Heimildir

Eiríkur Rögnvaldsson, Auður Þ. Rögnvaldsdóttir, Kristín Bjarnadóttir og Sigrún Helgadóttir. 2002. Vélræn málfræðigreining með námfúsum markara. *Orð og tunga* 6:1–9.

*Íslensk orðabók*. 2002 (3. útgáfa). Ritstj. Mörður Árnason. Reykjavík: Edda.

- Jón Hilmar Jónsson. 2001. *Orðastaður* (2. útgáfa). Reykjavík: JPV.
- Jörgen Pind (ritstj.), Friðrik Magnússon og Stefán Briem. 1991. *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.
- Kristín Bjarnadóttir. 2004. Beygingarlýsing íslensks nútímamáls. Í: *Samspil tungu og tækni. Afrakstur tungutækni-verkefnis menntamálaráðuneytisins*. Bls. 23–25.
- Sigrún Helgadóttir. 2004a. Markari fyrir íslenska texta. Í: *Samspil tungu og tækni. Afrakstur tungutækni-verkefnis menntamálaráðuneytisins*. Bls. 55–64.
- Sigrún Helgadóttir. 2004b. Mörkuð íslensk málheild. Í: *Samspil tungu og tækni. Afrakstur tungutækni-verkefnis menntamálaráðuneytisins*. Bls. 65–71.
- Sigrún Helgadóttir. 2005. Testing Data-Driven Learning Algorithms for PoS Tagging of Icelandic. Í: *Nordisk Sprogteknologi 2004 – Nordic Language Technology*. Bls. 257–265.

## Heimildir á vefnum

- Orðasambandaskrá Orðabókar Háskólans: [www.lexis.hi.is/osamb/osamb.pl](http://www.lexis.hi.is/osamb/osamb.pl)
- Upplýsingar um Orðasambandaskrá Orðabókar Háskólans: [www.lexis.hi.is/osamb/info.pl](http://www.lexis.hi.is/osamb/info.pl)

## Abstract

The topic of this article is the electronic tagging of phrases, fixed expressions and idioms as found in the collection of Orðabók Háskólans (Institute of Lexicography). The tool used for this, called *TnT-tagger*, grammatically analyses the words contained within the phrases. For this experiment, ca. 200 phrases were used, centred on two verbs (*afla* and *aka*) and two adjectives (*feitur* and *gláður*). The tagging process is described as well as the outcome of the tagging and the errors which occurred. The results are measured. The experiment also includes the tagging of several definitions from a dictionary (*Íslensk orðabók*). It is shown how the tags produced can be used for sorting the phrases so that they fall into groups of syntactic patterns. It is argued that this method is useful for finding patterns in the syntax of the keywords in question and for evaluating the frequency of various constructions.

**Keywords:**

phrases, idioms, grammatical analysis, PoS tagging, tagger

*Þórdís Úlfarsdóttir*

*Orðabók Háskólans*

*Neshaga 16*

*107 Reykjavík*

*dísa@lexis.hi.is*



# Orðabókar- og rannsóknarverkefni

## Beygingarlýsing íslensks nútímamáls

Markmið verkefnisins *Beygingarlýsing íslensks nútímamáls* er að koma upp beygingarlýsingu á tölvutæku formi til birtingar á vefsíðu Orðabókar Háskólans og til nota í ýmiss konar tungutækni-verkefni.

Verkefnið á rætur að rekja til þess að Orðabókin og Edda hf. fengu styrk frá verkefnisstjórn menntamálaráðuneytisins í tungutækni árið 2002. Gengið var til samninga um verkið í ágúst 2002. Samkvæmt samningi þessum skyldu vera 100 þúsund beygingardæmi í Beygingarlýsingunni. Orðabókin sá að öllu leyti um vinnuna og lagði til húsnæði, alla aðstöðu og aðgang að gögnum. Edda hf. lagði til beygingarlýsingu þá sem unnin var fyrir tölvuútgáfu *Íslenskrar orðabókar* (2000). Þessum áfanga verksins lauk í febrúar 2004 þegar menntamálaráðuneytinu var afhentur geisladiskur með útgáfu 1.0 af Beygingarlýsingunni, alls 173.389 beygingardæmi á formi XML-skráa.

Beygingardæmunum var komið fyrir á vefsíðu Orðabókarinnar jafnhliða því að unnið var að útgáfunni sem ætluð var til tungutækni-nota. Slóðin er <http://www.lexis.hi.is/beygingarlysing>.

Á vefsíðunni eru nú tæplega 200 þúsund beygingardæmi með öllum beygingarmyndum hvers orðs. Í safninu eru beygingar ósamsettra og samsettra orða úr almennu nútímamáli, auk mannanafna. Takmarkið með birtingu beygingardæmanna er að einskorða efnið við raunverulegar myndir hvers orðs, þ.e. að sýna afbrigði þar sem það á við og eyður þar sem beygingarmyndir eru ekki til. Þannig eru t.d. gefin tvö afbrigði af þágufalli eintölu af orðinu *spölur*, þ.e. *speli/spöl* en fleirtölu af eintöluorðum er sleppt (t.d. af *sykur* og *kaffi*) og germynd er

ekki sýnd af miðmyndarsögnum (t.d. af *óttast*). Án afbrigða eru beygingarmyndir nafnorðs 16, þ.e. fjögur föll í eintölu og fleirtölu, án greinis og með greini. Beygingarmyndir sagnar í persónuhætti eru 48, auk boðháttar, lýsingarháttar og viðskeyttra spurnarmynda (t.d. *ferðu* og *fórstu* af sögninni *fara*). Alls geta beygingarmyndir sagnar orðið 109, án afbrigða. Beygingarmyndir lýsingarorðs sem stigbreytist eru 120, án afbrigða, ef engar eyður eru í beygingardæminu.

Orðaforðinn í Beygingarlýsingunni var upprunalega fenginn úr 3. útgáfu *Íslenskrar orðabókar* og úr söfnum Orðabókarinnar en nú er verið að vinna við viðbótarefni úr öðrum heimildum. Helstu heimildir við rannsóknir á einstökum orðum og beygingarflokkum eru Ritmálskrá og Textasafn Orðabókarinnar, auk handbóka, greina og ritgerða um íslenska málfræði.

Ítarleg beygingarlýsing er grundvöllur að vélrænni greiningu á íslenskum textum, nauðsynlegur undanfari orðflokkgreiningar og setningagreiningar. Beygingarlýsingin nýtist t.d. við mörkun texta og við gerð leitarvéla, auk þess að vera forsenda skilvirkrar orðabókargerðar og heimildasöfnunar. Þá er Beygingarlýsingin einnig nauðsynlegur efniviður við gerð leiðréttingar- og þýðingarforrita.

Beygingarlýsingin er nú notuð í ýmsum verkefnum Orðabókar Háskólans, t.d. við vinnu við ÍSLEX, við Markaða íslenska málheild og við leit í Textasafninu. Þá hefur tekist gott samstarf við hugbúnaðarfyriertækið Spurl ehf. sem notar Beygingarlýsinguna í verk sín, t.d. í leitarvélina Emblu sem er á vefsíðu Morgunblaðsins og við leit í símaskránni, á ja.is.

Nú er unnið að nýjum gagnagrunni fyrir Beygingarlýsinguna í samvinnu við Spurl ehf., en styrkur til verkefnisins fékkst úr Tækniþróunarsjóði haustið 2005. Nýi gagnagrunnurinn mun auðvelda alla umsýslu við Beygingarlýsinguna og bæta leitaradgang á vefsíðu Orðabókarinnar til muna.

Menntamálaráðuneytið hefur með samningi falið Orðabókinni umsjón með og ráðstöfun á Beygingarlýsingunni til tungutækninota. Gerður er samningur um afnot af verkinu við hvern leyfishafa þar sem kveðið er á um skilmála.

Verkefnisstjóri er Kristín Bjarnadóttir. Faglega verkefnisstjórn skipa Eiríkur Rögnvaldsson, Sigrún Helgadóttir og Þórdís Úlfarsdóttir.

*Kristín Bjarnadóttir*

## ISLEX – Íslensk-norræn veforðabók

ISLEX er orðabókarverk sem er unnið á vegum Orðabókar Háskólans í samstarfi við stofnanir á Norðurlöndum. Áætlað er að verkið taki fimm ár í vinnslu og er stefnt að því að ljúka því árið 2011.

Um er að ræða rafræna orðabók með um 50.000 íslenskum uppflettiorðum ásamt þýðingum á sænsku, norsku og dönsku. Orðabókin á að endurspeglja íslenska málnotkun samtímans og er markmiðið að gera hana öllum aðgengilega á vefnum. Við ritstjórnina er notaður veftengdur gagnagrunnur sem er hannaður sérstaklega fyrir þetta verkefni. Gagnagrunnurinn gerir mögulegt að ritstjórnarvinnan fari fram samhliða í mörgum löndum.

Verkið hefur margvíslegt gildi í norrænu samstarfi, jafnt opinberu sem almennu. Það á að nýtast öllum þeim sem þurfa á íslensk-norrænum orðabókum að halda en sérstaklega verður leitast við að sinna þörfum sænskra, norskra og danskra notenda, ekki síst þýðenda úr íslensku.

ISLEX er samstarfsverkefni Orðabókar Háskólans og þriggja stofnana á Norðurlöndum, *Institutionen för svenska språket* við Gautaborgarháskóla, *Nordisk Institutt* við háskólann í Bergen og *Det Danske Sprog- og Litteraturselskab* í Kaupmannahöfn.

ISLEX-orðabókin á að birtast á vefnum og er leitast við að láta margvíslega möguleika rafrænnar miðlunar njóta sín. Þar verður því að finna ýmis nýmæli í íslenskri orðabókargerð, til dæmis litmyndir, hreyfimyndir, framburð orða og önnur hljóð. ISLEX-gagnagrunnurinn er enn í mótun og eftir er að ljúka frekari forritunar- og tæknivinnu, m.a. við þá hluta verksins sem snúa að endanlegri birtingu efnisins.

Orðaforðinn í verkinu, um 50.000 flettur, er að mestu kominn inn í gagnagrunninn. Auk flettiorðanna eru í ISLEX-orðabókinni tilfærð um 150.000 samsett orð, svokallaðar virkar samsetningar, sem ekki eru flettiorð en er skipað undir viðkomandi grunnorð. Þessi samsettu orð sýna m.a. hversu virk grunnorðin eru í orðmyndun og gefa vísbendingar um merkingartilbrigði þeirra.

Ætlunin er að sýna fullt beygingardæmi allra beygjanlegra fletti-orða ISLEX-orðabókarinnar og nýta í því skyni tungutækniverkefnið *Beygingarlýsing íslensks nútímamáls*. Er þetta eina íslenska orðabókin þar sem beygingar eru sýndar í heild sinni og án skammstafana.

Orðasamböndum er ætlað að skipa veigamikinn sess í orðabók-

arverkinu og er nú unnið að því að móta fyrirkomulagið á þeim. Framsetning á setningarlegum upplýsingum er einnig í mótun. Þetta varðar þætti eins og fallstjórn sagna og dæmigerð fylgiorð þeirra, svo dæmi séu nefnd.

Greining og vinnsla efnisins er komin nokkuð á veg og í fyrsta áfanga er lögð áhersla á frágang nafnorðanna.

Nú er hafin vinna við að bæta inn þýðingum á norsku, sænsku og dönsku. Verið er að leggja á ráðin um skipulag þýðingarvinnunnar og endanlega framsetningu markmálanna.

*Þórdís Úlfarsdóttir*

## Mörkuð íslensk málheild

Á vegum Orðabókar Háskólans er nú unnið að gerð markaðrar íslenskrar málheildar. Verkið er styrkt af tungutækniverkefni menntamálaráðuneytisins. Verkið hófst um mitt ár 2004 og skal því lokið árið 2007.

Með **markaðri málheild** (e. *tagged corpus*) er átt við safn fjölbreyttra textabúta sem hafa verið greindir á málfræðilegan hátt. Málheildin er geymd í rafrænu formi, venjulega í stöðluðu sniði. Hverjum textabút fylgja upplýsingar um textann sem búturinn er úr og hverri orðmynd fylgir **nefnimynd** (e. *lemma*) og greiningarstrengur, sem kallast **mark** (e. *tag*) og sýnir orðflokk og málfræðilega greiningu orðsins. Nefnimynd nafnorða er nefnifall eintölu, nefnimynd fornafna er nefnifall eintölu í karlkyni og nafnháttur er nefnimynd sagna. Taka má sem dæmi setningarbrotið *ég sagði*. Nefnimynd persónufornafnsins *ég* er *ég* og markið verður *f<sub>p</sub>1en*, þar sem *f* táknar fornafn, *p* táknar persónufornafn, *1* táknar fyrstu persónu, *e* táknar eintölu og *n* táknar nefnifall. Nefnimynd sagnarinnar *sagði* er *segja* og markið verður *sfg1eþ* þar sem *s* táknar sagnorð, *f* táknar framsöguhátt, *g* táknar germynd, *1* táknar fyrstu persónu, *e* táknar eintölu og *þ* táknar þátíð.

Valdir verða textar úr ritum sem gefin hafa verið út frá árinu 2000. Stefnt er að því að um 60% textanna komi úr bókum, 25% úr blöðum og tímaritum, 5–10% verði úr öðru útgefnu efni, 5–10% verði óútgefið efni og minna en 5% verði efni sem er skrifað til upplestrar. Enn fremur er stefnt að því að um 25% af textunum séu skáldverk og um 75%

verði nytjatexti sem skiptist milli texta um hagnýtt vísindi, náttúrufræði, þjóðfélagsfræði, heimsmál, viðskipti, listir, trúarbrögð, heimspeki og tómstundir. Stefnt er að því að í málheildinni verði í fyrstu um 25.000.000 lesmálsorð sem skiptast í um 900 textabúta. Hámarksstærð hvers textabúts verður 40.000 orð. Aldrei er tekinn heill texti. Ef texti er styttri en 40.000 orð er 10% af textanum sleppt.

Pegar vinna við málheildina var skipulögð var ekki gert ráð fyrir að safnað yrði talmáli og var það aðallega vegna þess hversu tímafrekt og dýrt það er. Nú hefur hins vegar komið í ljós að málheildin getur fengið talmálstexta úr öðrum verkefnum. Í fyrsta lagi er þar um að ræða texta sem hefur verið safnað á vegum verkefnisins *ÍSTAL – Íslenskur talmálsbanki* sem unnið var fyrir styrk frá Tæknisjóði á árunum 1999-2001. Í öðru lagi má nefna umræður á Alþingi sem var safnað á vegum verkefnisins *Tilbrigði í setningagerð* sem hlaut öndvegisstyrk frá RANNÍS 2005. Í þriðja lagi má nefna hópviðtöl um tökuorð og erlend áhrif sem voru hljóðrituð vegna norrænnar rannsóknar um viðhorf til tökuorða en verða fullskráð og frágengin á vegum verkefnisins *Tilbrigði í setningagerð*.

Stofn málheildarinnar er textasafn sem var útbúið vegna vinnu við *Íslenska orðtíðnibók* sem kom út á vegum Orðabókar Háskólans 1991. Í því safni eru um 500.000 lesmálsorð og fylgir hverri orðmynd nefnimynd og mark og hefur greining orða í textasafninu verið leiðrétt handvirkt. Textasafn Orðtíðnibókarinnar verður því notað sem fyrsti vísir að málheildinni. Árið 2002 veitti menntamálaráðuneytið styrk til verkefnis sem fólst í því að gera tilraunir til að marka íslenskan texta á vélrænan hátt. Vinna við verkið hófst síðla árs 2002 og var lokið í upphafi árs 2004. Niðurstöður verkefnisins verða nýttar við mörkun texta í málheildinni. Einnig hafa verið gerðar tilraunir við að finna nefnimyndir orða á vélrænan hátt. Stefnt er að því að lesmálsorð verði greind á vélrænan hátt með um 90% nákvæmni.

Við mörkunina þarf einnig að nota ýmsar hjálparskrár og orðasöfn. Stærst þessara hjálparskráa er orðasafn sem gert hefur verið úr Beygingarlýsingu íslensks nútímamáls. Beygingarlýsingin var upphaflega gerð fyrir styrk frá tungutækniverkefni menntamálaráðuneytisins en hefur síðan verið aukin verulega á vegum Orðabókar Háskólans. Einnig hefur verið aflað skráa yfir mannanöfn, örnefni, heiti fyrirtækja og skammstafanir.

Málheildir eru venjulega skráðar með stöðluðu sniði til þess að

tryggja að sem flestir geti nýtt efnið þrátt fyrir að menn noti ólíkar tölvur og hugbúnað. Notuð verður XML-útgáfa af sniði fyrir málheildir sem TEI-samtökin (TEI: *Text Encoding Initiative*) hafa skilgreint. Í þessu sniði er gert ráð fyrir að hverjum textabút fylgi haus þar sem skráðar eru margvíslegar upplýsingar um textann, höfund hans o.fl.

Notendur málheildarinnar eru einstaklingar, fyrirtæki og stofnanir sem vinna að orðabókargerð, margvíslegum tungutækniverkefnum og rannsóknum á íslensku nútímamáli. Úr málheildinni má lesa ýmiss konar gagnlegan fróðleik, t.d. upplýsingar um tíðni orðflokka, orða og beygingarmynda, orðasambönd, setningargerð og merkingu. Málheildir gefa einnig upplýsingar um hvernig tiltekið tungumál er notað á tilteknum tíma. Þær gefa vísbendingar um orðaforðann og einnig um málfræðilega og setningarfræðilega þætti.

Mörkuð málheild er því undirstaða fyrir þróun þýðingarforrita og mikilvæg fyrir nútíma orðabókargerð. Margir útgefendur orðabóka byggja nú gerð orðabóka á stórum mörkuðum málheildum. Upplýsingar sem fást úr markaðri málheild má einnig nota við gerð ýmissa tungutæknitóla, t.d. fyrir talgreiningu og talgervingu. Einnig eru slíkar upplýsingar nauðsynlegar við þróun hjálparforrita með ritvinnslu, t.d. forrita sem leiðbeina um stafsetningu og málfræði. Mörg tungutæknitól af þessu tagi nýtast sérstaklega fyrir blinda, heyrnarskerta og hreyfihamlaða og einnig þá sem glíma við skriftar- og lestrarörðugleika.

Gerður hefur verið samningur við menntamálaráðuneytið um að Orðabók Háskólans visti málheildina og veiti aðgang að henni. Ráðgert er að málheildin verð til ráðstöfunar til rannsókna í tungutækni og til þróunar tungutæknitóla. Einnig er stefnt að því að veita aðgang að málheildinni á vefsetri Orðabókar Háskólans með sérstökum leit-  
arhugbúnaði.

Til þess að unnt sé að hafa opinn aðgang að málheildinni er nauðsynlegt að semja við réttthafa texta um hvernig birtingu skuli háttáð. Í því sambandi skiptir höfuðmáli að engir textar verða birtir í heild í málheildinni þannig að útilokað er að endurgera verk með textum sem þar eru geymdir.

Ráðgert er að textabútar sem mynda málheildina verði sóttir í textasafn Orðabókar Háskólans. Í textasafninu eru textar af ýmsu tagi og frá ýmsum tímum en tækifærið verður notað til þess að auka það.

Verkefnisstjóri í verkinu er Sigrún Helgadóttir. Verkefnisstjórn,

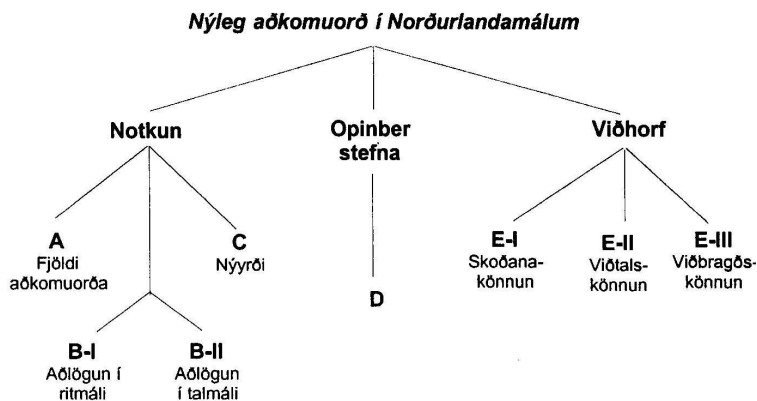
skipuð Ástu Svavarsdóttur, Eiríki Rögnvaldssyni og Kristínu Bjarnadóttur starfar með verkefnisstjóra.

Sigrún Helgadóttir

## Rannsókn á aðkomuorðum í Norðurlandamálum

Um nokkurra ára skeið hefur verið í gangi samanburðarrannsókn á nýlegum aðkomuorðum í Norðurlandamálum (*Moderne importord i språka i Norden*), sem nær til íslensku, dönsku, norsku, sænsku, finnsku, finnlandssænsku og færeysku. Meginmarkmið hennar er að skoða tiltekna þætti sem varða aðkomuorð og bera þá saman í málunum sjö. Með *aðkomuorðum* er átt við öll orð sem eiga uppruna í öðrum málum og koma fyrir í norrænu textasamhengi, allt frá fullaðlöguðum tökuorðum til erlendra orða sem bregður fyrir í textunum; *nýleg* merkir í rannsókninni orð sem hafa borist í Norðurlandamálin eftir seinni heimsstyrjöld (ca. 1945 eða síðar). Á þessu tímabili er fyrst og fremst um ensk áhrif að ræða og gildir það jafnt um öll málin.

Rannsóknin skiptist í þrjá meginhluta: rannsókn á *notkun* aðkomuorða, umfjöllun um *opinbera málstefnu* gagnvart aðkomuorðum og rannsókn á *viðhorfi almennings* til aðkomuorða. Þeir skiptast síðan flestir í rannsóknarþætti eins og sýnt er á eftirfarandi skýringarmynd:



Notkunarhlutinn skiptist í þrennt. Í fyrsta lagi (A) var kannaður hlutfallslegur fjöldi aðkomuorða í dagblaðatextum á tungumálunum sjö (blöð frá 1975 og 2000). Í öðru lagi (B) var rannsökuð aðlögun aðkomu-

orða í töluðu og rituðu máli, m.t.t. framburðar, stafsetningar, beyginga og orðmyndunar. Rannsóknin fólst annars vegar í spurningakönnun (talmál) og hins vegar í frekari athugun og úrvinnslu á dagblaðæfningu úr A-hluta verkefnisins (ritmál). Í þriðja lagi (C) voru athuguð nýyrði sem ætlað er að leysa aðkomuorð af hólmi og útbreiðslu þeirra. Rannsóknin beinist að ritmáli og styðst m.a. við dagblaðatextana sem notaðir voru í A. Annar hluti verkefnisins (D) felst í sögulegri lýsingu á opinberri málstefnu hvers lands á tímabilinu 1850–2000 og sáu málnefndir þátttökulandanna um þann þátt. Viðhorfsrannsóknirnar skiptast í þrjá rannsóknarþætti. Sá fyrsti fólst í skoðanakönnun þar sem fjöldi fólks (500-1000 eftir löndum) var spurður um ákveðin atriði sem snerta aðkomuorð í móðurmáli þátttakenda (E-I). Viðtalskönnun (E-II) gegnir því hlutverki að varpa skýrara ljósi á niðurstöður skoðanakönnunarinnar og gefa færi á því að kafa dýpra í ákveðin atriði með viðtölum við tiltölulega lítinn hóp fólks. Markmið viðbragðskönnunarinnar (E-III), sem nær til nokkurra stórra hópa, er að kalla fram viðbrögð fólks við notkun aðkomuorða. Tilgangurinn með því að skipta rannsókninni niður í marga smærri þætti er ekki síst sá að skoða viðfangsefnið frá ýmsum hliðum og með mismunandi rannsóknaraðferðum þannig að bæði sé hægt að bera saman niðurstöðurnar innbyrðis og á milli mála.

Að rannsókninni stendur hópur norræna fræðimanna og háskóla-stúdentna undir stjórn Helge Sandøy, prófessors við háskólann í Bergen. Íslensku þátttakendurnir eru Guðrún Kvaran (C-hluti) og Ásta Svavarsdóttir (B-I og B-II) á Orðabók Háskólans, Kristján Árnason (E-I), Hanna Óladóttir (E-II) og Halldóra Björt Ewen (E-III og a.n.l. E-II) við Háskóla Íslands. Auk þeirra hafa aðrir komið að einstökum rannsóknarþáttum, einkum stúdentar við HÍ, og Ari Páll Kristinsson skrifaði skýrslu um opinbera stefnu gagnvart aðkomuorðum (D-hluti) fyrir Íslenska málnefnd. Rannsóknirnar hafa verið styrktar af Norrænu málráði (Nordisk sprøgråd), NOS-H, Norræna menningarsjóðnum (Nordisk kulturfond), Rannsóknasjóði Háskóla Íslands og Nordplus sprog.

Niðurstöður úr öllum þáttum rannsóknarinnar eru birtar í *ritröðinni Moderne importord i språka i Norden* hjá Novus í Noregi. Þrjú rit eru þegar komin út. Í því fyrsta eru fyrirlestrar frá ráðstefnu sem haldin var í Bergen haustið 2002 (*Med 'bil' i Norden i 100 år. Ordlagning og tilpassing av utalandske ord*. Oslo 2003; ritstjóri Helge Sandøy), annað



er greinasafn um opinbera málstefnu á Norðurlöndunum (*Normering av importord i Norden. Historikk*. Oslo 2004; ritstjórar Helge Sandøy og Jan-Ola Östman) og það þriðja er skýrsla um sænsku viðtalsrannsóknina á viðhorfum málnotenda til aðkomuorða og erlendra máláhrifa (*Teamwork? Man kan lika gärna samarbeta. Svenska åsikter om importord*. Oslo 2005; höfundur Catharina Nyström Höög). Rit með niðurstöðum úr öðrum rannsóknarþáttum eru væntanleg á árinu 2006. Auk þess voru niðurstöður rannsóknarinnar í heild kynntar á ráðstefnu í Kaupmannahöfn í árslok 2005 og einnig hefur verið fjallað um verkefnið eða hluta þess í greinum og fyrirlestrum einstakra fræðimanna, bæði héraendis og erlendis.

Sjá nánar um verkefnið: <http://www.hf.uib.no/moderne/>.

Ásta Svavarsdóttir

## Spænsk-íslensk, íslensk-spænsk orðabók

Fyrsti vísir að orðabók á Íslandi var baskneskt-íslenskt orðasafn sem fæddist af samskiptum Íslendinga og baskneskra sjómanna á Vestfjörðum síðla 16. aldar og fram á þá 17.<sup>2</sup> Síðan þá hafa samskipti við Spán farið sívaxandi og nú er Spánn það land sem flestir Íslendingar sækja heim. Að sama skapi hefur Rómanska Ameríka laðað unnendur spænskrar tungu til sín og hefur vægi spænsku sem viðskiptamáls aldrei verið meira. Hefur það leitt til þess að spænskunemum hefur fjölgað þannig að flestir framhaldsskólar landsins bjóða upp á spænsku sem þriðja mál. Þar af leiðandi hefur skortur á nútímalegri spænsk-íslenskri orðabók aldrei verið jafn aðkallandi.

Fram til þessa hafa verið gefnar út tvær spænsk-íslenskar orðabækur. Sú fyrri er *Spænsk-íslensk vasaorðabók* eftir Elisabeth Hangartner Ásbjörnsson og Elvira Herrera Ólafsson (1978). Hún hefur nýst framhaldsskólanemum í fyrsta og öðrum áfanga spænskunáms en er barn síns tíma og gefur ekki ítarlegri upplýsingar um orð en kyn þeirra. Málnotkunardæmi eru engin. Síðari bókin, *Spænsk-íslensk orðabók*, er

---

<sup>2</sup>Basknesk íslensku orðasöfnin er að finna í þremur óprentuðum handritum. Eitt þeirra er á Árnastofnun (AM 987 4to) og hin er að finna á handritadeild Landsbókasafns (JS 284 8vo og JS 401 4to).

stærri.<sup>3</sup> Hún er eftir Sigurð Sigurmundsson frá Hvítárholti og kom hún út árið 1973 og aftur árið 1995. Bókin ber þess ýmis merki að höfundur hafði ekki menntun í spænsku en eftir stendur engu að síður aðdáun á frumkvöðlaverki sjálfmenntaðs manns.

Þrátt fyrir augljósa þörf á spænsk-íslenskri orðabók tók þó nokkurn tíma að afla fjár til að hægt væri að hefjast handa við gerð orðabókarinnar. Verkið er kostnaðarsamt og enginn fjárhagslegur ábati fyrir bókaforlög af því að leggja í slíkt verk. Því var nauðsynlegt að fá öfluga liðsmenn til að hleyppa verkinu af stað. Höfðinglegur styrkur frá Menntamálaráðuneyti og Minningarsjóði Margrétar Björgólfsdóttur gerði Háskólanum í Reykjavík og Eddu útgáfu kleift að setjast að samningaborði og hefja samvinnu um verkið.

Til verksins voru ráðnir 3 sérfræðingar í spænsku í rúmlega tvö stöðugildi, þær Guðrún H. Tulinius, Ragnheiður Kristinsdóttir og Sigrún Á. Eiríksdóttir. Ritstjóri verksins er Margrét Jónsdóttir og Laufey Leifsdóttir kemur að verkinu fyrir hönd Eddu útgáfu.

Til að spara tíma og fé var ákveðið að fá aðkeyptan orðabókargrunn frá hinni virtu útgáfu Harper-Collins. Um er að ræða rúmlega 20.000 flettur í hvora átt og er ætlunin að auka við þann fjölda sérvöldum orðaforða, þannig að spænsk-íslenski hlutinn innihaldi um 25.000 flettur. Íslensk-spænski hlutinn verður svipaður að umfangi.

Við val á gagnagrunni var haft í huga að bókin henti byrjendum í spænsku og nemendum sem enn eru ekki komnir það langt í að til-einka sér málið að þeir geti notast við spænsk-spænskar orðabækur. Jafnframt er miðað við að bókin nýtist öðrum, s.s. ferðamönnum, húseigendum á Spáni og þeim sem stunda viðskipti við hinn spænskumælandi heim. Þó íslensk-spænski hlutinn miðist aðallega við Íslendinga í spænskunámi, er einnig tekið mið af spænskumælandi fólki á Íslandi sem ekki hefur íslensku að móðurmáli.

Vinnan við orðabókina hófst í október 2005. Notast er við orðabókarforritið *Lexa* sem skrifað var af sérfræðingum á orðabókardeild Eddu. Ætlunin var að flytja Harper-Collins grunninn inn í Lexuna en þar sem samningar um kaup á orðagrunninum drögust á langinn hefur danskur grunnur frá GADE útgáfunni verið lagður til grundvallar og viðbætur og breytingar færðar inn í hann eftir prentuðu orðabókinni frá Harper-Collins. Vinnan hefur því aðallega verið fólgin í því að þýða ýmist úr dönsku eða ensku. Verkið hefur gengið vel og

<sup>3</sup>Sigurður Sigurmundsson, *Spænsk – íslensk orðabók*. Reykjavík 1995.

er innslætti á fyrri hluta bókarinnar lokið og yfirlestur hafinn. Vonir standa til að sá hluti verði aðgengilegur á veraldarvefnum á haustmánuðum 2006.

Í orðabókinni verður algengasti orðaforði spænskrar tungu undanfarinna ára, þ.á m. orð frá spænskumælandi löndum Ameríku. Ýmis orð eru sérmerkt annaðhvort Spáni (ESP) eða Ameríku (AM) og fyrir kemur að orð sé sérmerkt einstöku landi þar. Öll algengustu orð og hugtök eru uppflettiorð í bókinni. Gerð er ítarleg grein fyrir merkingu þeirra og notkun og fjölmörg notkunardæmi eru birt í bókinni. Eitt sérkenni bókarinnar eru upplýsingar af menningarlegum toga, sem ætlunin er að bæta verulega við. Einnig eru upplýsingar um fjölda gagnlegra skammstafana og málfraeðilegar upplýsingar, þ.á m. um beygingu sagna. Starfsmenn orðabókarinnar hafa bætt við talsverðum sérhæfðum orðaforða t.d. varðandi atvinnugreinar hér á landi, dýralíf, lögfræði, hagfræði, viðskipti, listir, heilbrigðismál, frístundir, ferðalög, jarðfræði og margt fleira. Er ætlunin að hafa í íslenska-spænska hlutanum samskonar menningarlegar upplýsingar og finna má í þeim spænsk-íslensk, en eitt af markmiðum ritstjórnar er að auka menningararlæsi þeirra sem nýta sér bókina, hvort sem um spænskumælandi lönd er að ræða eða Ísland.

Þegar fyrri hluta er lokið verður orðunum varpað yfir í hinn hlutann þannig að öll þau orð sem hafa einfalda skýringu fara sjálfvirkt inn í íslenska-spænska hlutann. Dæmi: *orgullo*: nm, stolt, mont. Bæði þýðingarorðin, *stolt* og *mont*, fara inn sem íslensk uppflettiorð og munu starfsmenn greina orðin og finna dæmi um notkun þeirra. Einnig verður grunnurinn borinn saman við aðrar tvímála orðabækur sem Edda útgáfa gefur út sem og íslenskan orðagrunn útgáfunnar.

Ef áætlanir standast kemur spænsk-íslensk, íslensk-spænsk orðabók út árið 2008. Þar sem verkinu hefur miðað framur vonum er líklegt að svo verði. Er það von undirritaðra að hér sé á ferðinni orðabók sem stenst tíma- og kostnaðaráætlanir og verði lifandi vitnisburður um gott samstarf háskóla og atvinnulífs.

Guðrún H. Tulinius og Margrét Jónsdóttir

## Tilbrigði í setningagerð

Rannsóknarverkefnið *Tilbrigði í setningagerð* hefur það meginmarkmið að gera grein fyrir ýmiss konar tilbrigðum í íslenskri setningagerð, athuga útbreiðslu þeirra og leitast við að skýra eðli þeirra og einkenni. Tilbrigðin eru skoðuð út frá landfræðilegri dreifingu, félagslegum breytum (t.d. kyni og aldri málnotenda) og mismunandi málaðstæðum (talmál/ritmál, formlegt/óformlegt mál o.s.frv.). Enn fremur felur verkefnið í sér rannsóknir á tilteknum atriðum í færeyskri setningagerð og samanburði þeirra við íslensku. Íslenska verkefnið er hluti af norræna verkefninu *Scandinavian Dialect Syntax* (ScanDia-Syn), sem miðar að því að rannsaka og bera saman mállýskubundin setningatilbrigði á öllu norræna málsvæðinu, þvert á mörk þjóðmálanna, og að koma upp aðgengilegum gagnagrunni með efniviði til setningafræðilegra rannsókna. Einnig hefur verið stofnað til tengsla og samvinnu við sambærileg verkefni annars staðar í Evrópu, einkum í Hollandi og á Norður-Ítalíu.

Í rannsóknunum er annars vegar byggt á spurningalistum sem lagðir verða fyrir fjölda málnotenda víða um land, bæði munnlega og skriflega. Með þeim verða rannsökuð valin atriði sem talin eru áhuga-verð frá setningarlegu sjónarmiði. Áhersla verður lögð á atriði þar sem einhvers konar tilbrigði koma fram í málnotkun og getur munurinn þá ýmist verið á milli einstaklinga (t.d. eftir uppruna þeirra eða aldri) eða í máli sama einstaklings (t.d. eftir aðstæðum). Hins vegar byggjast rannsóknirnar á samfelldum textum úr rituðu og þó einkum töluðu máli og verkefnið felur m.a. í sér umfangsmikla efnissöfnun úr talmáli. Þar er einkum um það að ræða að safna saman efni sem þegar er til, ljúka nauðsynlegri skráningu þess og ganga frá því í aðgengilegu formi. Þannig hefur verið lokið frágangi á efni sem safnað var í verkefninu ÍSTAL — *Íslenskur talmálsbanki* á árunum 1999–2000 (sjálfsprottin, persónuleg samtöl) og skráningu lokið á efni sem safnað var fyrir rannsókn á aðkomuorðum í íslensku árið 2002 (*MIN-verkefnið*; viðtöl) og á umræðum frá Alþingi, sem teknar voru upp og frumskráðar á vegum þingsins. Alls eru þetta nálægt 50 klst. af hljóðrituðu og umrituðu efni. Auk þess hefur verið unnið að skráningu og undirbúningi að frekari úrvinnslu hljóðritana í vörslu Árnastofnunar sem geyma viðtöl við Vestur-Íslendinga og einnig hefur verið rætt við aðstandendur fleiri rannsóknarverkefna um afnot af efni frá þeim.

Tekist hefur samvinna við verkefnið *Mörkuð íslensk málheild*, sem unnið er að við Orðabók Háskólans (sjá verkefnislýsingu Sigrúnar Helgadóttur í þessu hefti), um úrvinnslu og greiningu talmálfenins. Munu bæði verkefni njóta góðs af samstarfinu. Lokið verður við umritun efnisins innan setningafræðiverkefnisins og gengið frá því með samræmdu sniði. Efnið verður síðan lagt inn í fyrirhugaða málheild, þar sem ekki hefði verið kostur á að hafa talmálsefni ef slík samvinna hefði ekki hefði komið til. Á móti kemur að málheildarverkefnið skilar vélrænni mörkun og greiningu efniviðarins, sem gerir hann miklum mun aðgengilegri til leitar og rannsókna.

Verkefnið *Tilbrigði í setningagerð* fékk veglegan öndvegisstyrk frá RANNÍS 2005 til þriggja ára. Verkefnisstjóri er Höskuldur Þráinsson, prófessor við HÍ, og aðrir í stjórn verkefnisins eru Eiríkur Rögnvaldsson, Jóhannes Gísli Jónsson og Sigríður Sigurjónsdóttir frá HÍ, Þórunn Blöndal frá KHÍ og Ásta Svavarsdóttir frá OH. Auk þeirra koma fleiri fræðimenn að verkefninu auk fjölmargra háskólastúdenta í íslensku.

Frekari upplýsingar um verkefni sem vísað er til má finna á eftirtöldum vefsíðum:

*Tilbrigði í setningagerð* (verkefnislýsing á ensku):

<http://uit.no/scandiasyn/island/>

*ScanDiaSyn*: <http://uit.no/scandiasyn/scandiasyn/>

ÍSTAL — *Íslenskur talmálsbanki*: <http://www.hi.is/~eirikur/istal/>

MIN (Morderne importord i språka i Norden):

<http://www.hf.uib.no/moderne/>

*Mörkuð íslensk málheild*: <http://www.lexis.hi.is/malheild.htm>

*Ásta Svavarsdóttir*

## Tungutækni-verkefni sem Orðabók Háskólans tekur þátt í

Frá árinu 2001 hefur Orðabók Háskólans rekið *Íslenskt upplýsingasetur um tungutækni*, sem kostað var af norrænu tungutækniáætluninni fram á mitt ár 2005. Setrið hefur tekið þátt í samstarfsneti norrænna upplýsingasetra, *NorDokNet*, [www.nordoknet.org](http://www.nordoknet.org), en starfstíma þess lauk einnig á árinu. Upplýsingasetrið beitti sér fyrir samningu íslensks íð-orðasafns í tungutækni, en fé fékkst til þess úr norrænu tungutækni-

áætluninni. Valdís Ólafsdóttir, meistaraneimi í tungutækni, var ráðin til að leggja drög að safninu og vann hún við það um þriggja mánaða skeið (í 75% starfi). Verkið var unnið í samvinnu við Orðabanka Íslenskrar málstöðvar, og var iðorðasafnið opnað sem hluti bankans 1. september.

Þegar tungutækniáætlun menntamálaráðuneytisins lauk og verkefnisstjórn í tungutækni var lögð niður um áramótin 2004-5 yfirtók upplýsingasetrið vef hennar, [www.tungutækni.is](http://www.tungutækni.is). Til að halda utan um og ýta undir starfsemi á sviði íslenskrar tungutækni var ákveðið að stofna *Tungutækni-setur*, sem er samstarfsverkefni Orðabókarinnar, Málvísindastofnunar Háskóla Íslands og tækni- og verkfræðideildar Háskólans í Reykjavík. Í stjórninni situr einn fulltrúi frá hverjum samstarfsaðila, og er Sigrún Helgadóttir fulltrúi Orðabókarinnar. Meðal markmiða setursins er að halda árlega ráðstefnu um íslenska tungutækni, og verður sú fyrsta haldin nú í maí.

Norræna nýsköpunarmiðstöðin (Nordisk Innovationscenter) og Norræna rannsóknamiðstöðin (NordForsk) veittu norrænum upplýsingasetrum um tungutækni og samtökum iðnaðarins á Norðurlöndum styrk til forverkefnis sem nefnt var *NLTNet*. Styrknum var varið til að halda sameiginlegan fund í Kaupmannahöfn þar sem ræddar voru forsendur fyrir samvinnu um kynningu og eflingu tungutækni á Norðurlöndum. Þar var samþykkt að stefna að áframhaldandi samstarfi, og var gefin út skýrsla um niðurstöður fundarins – sjá [www.nltnet.org](http://www.nltnet.org). Snemma árs 2006 sendi sami hópur svo umsókn til Norrænu nýsköpunarmiðstöðvarinnar um nýtt tveggja ára verkefni, *Nordic ICT and Language*, þar sem ætlunin er að fylgja eftir þeirri stefnu sem mörkuð var í forverkefninu. Þegar þetta er ritað er ekki komið í ljós hvort styrkur fæst til þessa verkefnis.

Orðabókin og Háskóli Íslands tóku þátt í *Nordisk Netordbog*, samstarfsneti um rannsóknir og þróun í tungutækni, einkum margmála leit á netinu og í gagnabönkum. Aðrir þátttakendur í verkefninu eru Háskólinn í Bergen, Kungliga tekniska högskolan (KTH) í Stokkhólmi, Háskólinn í Helsinki og Center for sprogteknologi (CST) í Kaupmannahöfn, sem stýrir verkefninu (verkefnisstjóri Bente Maegaard). Norræna ráðherranefndin átti frumkvæði að þessu verki og hefur veitt fé til þess, ásamt Málráði Norðurlanda. Vinnan í verkefninu á árinu fólst einkum í því að afla ýmissa tví- eða margmála orðalista til að nýta í margmála leit. Þrír fundir hafa verið haldnir í verkefninu og hefur

Eiríkur Rögnvaldsson sótt þá. Framhald þessa verkefnis, *Tværsproglig søgning på tekster og ordbøger*, fékk svo styrk frá Nordplus Sprog í lok ársins. Þessi verkefni halda því áfram á árinu 2006.

Auk þessa átti Orðabókin aðild að þremur fjölþjóðlegum umsókn-um um styrki til tungutækniverkefna á árinu 2005. Þar var í fyrsta lagi um að ræða verkefnið *Nordic Multilingual Technologies*, umsókn til NOS-HS um norrænt öndvegissetur í margmála upplýsingatækni. Forsvarsmaður umsóknarinnar var Koenraad de Smedt prófessor í Bergen. Í öðru lagi var verkefnið *NorPar*, umsókn til NORA, Norrænu Atlantshafsnefndarinnar, um gerð samskipaðrar málheildar (parallel korpus) fyrir dönsku, norsku, sænsku, færeysku og íslensku, með áherslu á tvö síðastnefndu málin. Forsvarsmaður umsóknarinnar var Janne Bondi Johannessen prófessor í Osló. Í þriðja lagi var verkefnið *EuroDocNet*, umsókn í 6. rammaáætlun Evrópusambandsins. Hans Uzokoreit prófessor í Saarbrücken var forsvarsmaður umsóknarinnar, en markmið verkefnisins var að koma á evrópsku samstarfsneti upplýsingasetra í tungutækni, koma upp upplýsingasetrum í Eystrasaltlöndunum, koma upp sameiginlegum gagnagrunni með upplýsingum um tungutækni, þróa aðferðir við margmála leit, þróa og samhæfa iðorðaförða greinarinnar á mörgum tungumálum o.fl.

Engin þessara þriggja umsókna hlaut brautargengi, en þátttaka í þeim var samt mjög gagnleg fyrir Orðabókina. Með henni fékkst dýr- mæt reynsla af samningu umsókna af þessu tagi, auk þess sem Orðabókin komst í tengsl við ýmsar erlendar stofnanir og einstaklinga sem starfa á sviði tungutækni. Síðastnefnda umsóknin hefur nú verið endurskoðuð og send aftur til Evrópusambandsins en niðurstöðu er ekki að vænta fyrir en í haust.

*Eiríkur Rögnvaldsson*

## Bókafregnir

Jón Hilmar Jónsson. *Stóra orðabókin um íslenska málnotkun*. JPV útgáfa, Reykjavík 2005. ISBN 9979-791-26-8. xxx, 1562 bls. Rafræn útgáfa á geisladiski fylgir.

*Stóra orðabókin* er afar yfirgripsmikið verk sem gefur skýra mynd af íslenskum orðaforða og innra samhengi hans. Saga hennar er löng og nokkuð óvenjuleg. Hún byggir á tveimur bókum sem Jón Hilmar samdi á árunum 1989 til 2002. Hin fyrri er *Orðastaður* sem kom út 1994 og í endurskoðaðri útgáfu 2001. Síðara verkið er hugtakaorðabókin *Orðaheimur* sem gefin var út 2002. Þeim var báðum mjög vel tekið. Fljótlega varð ljóst að að efni beggja bókanna gat myndað samtæða heild sem sjálfstæð orðabók og höfundur réðst í það mikla verk að semja orðabók sem annars vegar byggði á fyrri bókunum tveimur en hins vegar á talsverðu nýju efni.

Bókin sameinar alhliða lýsingu á íslenskri málnotkun. Hún skiptist í tvo meginhluta. Annars vegar er um að ræða orðabókina sjálfa með um 13.000 flettiorðum, en hins vegar skrá um öll orð og orðasambönd í orðabókartextanum, um 85.000 talsins. Aftast er síðan lykilorðaskrá með dönskum, enskum og þýskum samsvörnum íslensku hugtakaheitanna.

Bókinni fylgir rafræn útgáfa á geisladiski sem veitir margvíslega leitarmöguleika sem ekki eru mögulegir í prentari bók.

Jón Hilmar Magnússon. *Íslensk-færeysk orðabók*. Hið íslenska bókmenntafélag, Reykjavík 2005. ISBN 9979-66-179-8. 877 bls.

Íslensk-færeysk orðabók er hin fyrsta sinnar tegundar. Hún hefur að geyma um 51.000 uppflettiorð og eru flettur þannig settar fram að



hverju orði fylgja helstu málfræðilegar upplýsingar en síðan eru birtar færeyskar skýringar, oftast í formi jafnheita. Notkunardæmi eru birt í sumum flettunum en höfundur gerir sér far um að sýna föst orðasambönd og orðatiltæki. Höfundurinn réðst einn í þetta stóra verk og vann að því í yfir tíu ár. Hann fékk færeyska orðabókamenn til að lesa merkingarskýringarnar yfir og gefur það henni aukið gildi. Í lok formála boðar höfundur að færeysk-íslensk orðabók sé í undirbúningi.

Halldóra Jónsdóttir [ritstj.] *Íslensk-dönsk, dönsk-íslensk vasaorðabók*. Mál og menning, Reykjavík 2005. ISBN 9979-3-2630-1. 782, iii bls.

Íslensk-dönsk orðabók hafði ekki komið út í áratugi þegar þessi vasa-orðabók kom á markað. Hún tekur mið af íslensku nútímamáli en sérstaklega var safnað orðum sem tengjast ferðamennsku. Dansk-íslenski hlutinn var sniðinn eftir nýjum orðagrunni frá Gyldendal í Danmörku, en í formála kemur fram að uppbygging flettna sé miðuð við nýendurskoðaða Dansk-íslenska skólaorðabók frá Eddu.

Í bókinni eru alls um 37.000 uppflöttiorð með um 12.000 dæmum um málnotkun. Notkunarhópurinn miðast við Íslendinga sem ferðast til Danmerkur, námsmenn og þá sem nota tungumál við dagle störf.

*Tölvuorðasafn*. Íslenskt-enskt, enskt-íslenskt. 4. útgáfa, aukin og endurbætt. Orðanefnd Skýrslutæknifélags Íslands tók saman. Ritstjóri Stefán Briem. Hið íslenska bókmenntafélag í samvinnu við Skýrslutæknifélags Íslands, Reykjavík 2005. ISBN 9979-66-164-X. 555 bls.

Í fjórðu útgáfu tölvuorðasafnsins eru um 7700 íslensk heiti og um 8500 ensk heiti á rúmlega 6500 hugtökum sem lúta að upplýsingatækni og tölvunotkun. Orðasafninu er eins og í fyrri útgáfum skipt í tvo hluta. Fyrri hlutinn er íslensk-ensk orðaskrá með skilgreiningum og útskýringum á flestum hugtökunum. Síðari hlutinn er ensk-íslensk orðaskrá. Þar er íslensk þýðing við hvert orð sem um leið er tilvísun til fyrri hluta verksins. Bókin er nú talsvert stærri en þriðja útgáfa og hefur um 1500 hugtökum verið bætt við. Þriðja útgáfa hefur verið aðgengileg í orðabanka Íslenskrar málstöðvar. Ráðgert er að fjórða útgáfa verði sett

í orðabankann þegar þeirri endurskoðun á bankanum lýkur sem nú  
stendur yfir.

*Guðrún Kvaran*