

## TÍÐNI ORÐA Í TALI BARNNA

Jóhanna Thelma Einarsdóttir, Anna Lís Pétursdóttir og Íris Dögg Rúnarsdóttir. (2019).  
*Tíðni orða í tali barnna*. Háskólaútgáfan, Reykjavík.

Bókin *Tíðni orða í tali barnna* kom út hjá Háskólaútgáfunni 2019. Hún er 534 blaðsíður og er að meginefni orðalistar, auk stutts inngangs. Hún byggir á málsýnabanka Jóhönnu Thelmu Einarsdóttur (2016), dósents í talmeinafræði við Heilbrigðisvísindasvið og Menntavísindasvið Háskóla Íslands, og tveimur meistaraþrófsritgerðum í talmeinafræði þar sem Anna Lís Pétursdóttir (2017) og Íris Dögg Rúnarsdóttir (2018) unnu orðtíðnibækur upp úr gögnum Jóhönnu undir handleiðslu hennar. Í bókinni er hægt að fletta upp tíðni þeirra orða og beygingarmynda sem koma fyrir í tali barnanna í gagnabanka Jóhönnu, en börnin voru á aldrinum 2;5–7;11 ára þegar málsýnum frá þeim var safnað. Bókin var gefin út með styrkjum frá Styrktarsjóði Áslaugar Hafliðadóttur og Rannsóknasjóði Háskóla Íslands. Hún kom út á prentuðu formi en er einnig aðgengileg á rafrænu formi á vef Háskólaútgáfunnar.

Bókin er að stofni til fimm mismunandi orðalistar, auk inngangs. Í inngangi er gerð grein fyrir tilurð verksins og aðferðum við vinnslu þess. Meðal annars er gerð nákvæm grein fyrir afritun, úrvinnslu og greiningu þeirra gagna sem bókin byggir á og ýmis álitamál reifuð sem tengjast skáningu orðanna, flokkun þeirra í orðflokka og framsetningu orðalistanna. Orðtíðnibókin sjálf skiptist í fimm kafla þar sem orðunum og beygingarmyndum þeirra er raðað annars vegar í stafrófsröð og hins vegar í tíðniröð. Í 1. kafla er skrá yfir öll orðin í stafrófsröð, beygingarmyndir hvers orðs, hversu oft orðin og mismunandi beygingarmyndir þeirra koma fyrir og orðflokk þeirra. Í bókinni er reyndar notað hugtakið *flettiorð* í stað orðs en það er uppflertimynd orðs sem leitað er að í orðabók. Í 2. kafla er skrá yfir orðin í stafrófsröð með upplýsingum um tíðni þeirra og orðflokka og í 3. kafla eru sömu upplýsingar gefnar nema orðunum er ekki raðað í stafrófsröð heldur eftir tíðni, þar sem algengasta orðið (sögnin *vera*) kemur fyrst og svo koll af kolli. Í 4. kafla er orðum sem tilheyra hverjum orðflokki fyrir sig raðað í stafrófsröð og loks í 5. kafla eftir tíðni innan hvers orðflokks. Orðflokkunum er raðað þannig að fyrst koma nafnorð, þá lýsingarorð, fornöfn, töluorð, sagnorð, atviksorð og skyldir flokkar, samtengingar og að lokum flokkurinn ógreind orð, en í þann flokk falla þau orð sem ekki var hægt að greina í orðflokka, svokölluð óorð, og einnig erlend orð. Undir flokkinn atviksorð og skyldir flokkar falla ekki aðeins atviksorð heldur einnig forsetningar, upphrópanir og nafnháttarmerkið *að*. Atviksorð eru merkt sem slík en orð sem geta tilheyrt fleiri en einum þessara orðflokka eru merkt með stjörnu og tekið fram um hvaða orðflokka er að ræða. Þannig er t.d. smáorðið *á* stjörnumerkt í flokki atviksorða og skyldra flokka og tekið fram að það geti verið atviksorð, forsetning eða upphrópun.

Orðin í bókinni eru úr 451 málsýni frá 353 börnum á aldrinum 2;5–7;11 ára, sem safnað var af nemum í talmeinafræði og í kennaranámi við Háskóla Íslands á árunum 2009–2014. Málsýnum er ætlað að endurspeglar raunverulega máltjáningu einstaklinga við eðlilegar aðstæður og þar með gefa þau m.a. upplýsingar um orðaforða. Rannsóknir á máltöku barna sýna að tíðni orða í móðurmáli þeirra hefur mikið að segja um orðaforða þeirra. Þannig virðast börn læra þau orð og þær beygingarmyndir orða sem eru algengar í málumhverfi þeirra fyrr en orð sem eru sjaldgæfari í málinu. Við rannsóknir á tíðni orða í talmáli er því gott að nota málsýni. Orðin í bókinni komu fyrir í tali barnanna í samtölum þeirra við háskólanema sem hittu hvert barn og tóku upp tal þess, oftast á myndband. Höfundar nefna í inngangi að nemarnir hafi leitast við að laða fram sjálfsprottið tal barnanna þannig að málsýnin gæfu dæmigerða mynd af máltjáningu þeirra. Málsýnin úr gagnabankanum eru þrenns konar. Í fyrsta lagi er um að ræða sýni sem nemendur sófnuðu fyrir nemendaverkefni og eru þau flest frá börnum háskólanema. Í öðru lagi voru namar sendir í leikskóla til að safna sýnum frá ákveðnum aldurshópum leikskólabarna og í þriðja lagi var málsýnum grunnskólabarna safnað eftir að tekið hafði verið úrtak úr Þjóðskrá. Málsýnin voru síðan skráð og var afritun þeirra flestra bundin við fyrstu 50 segðir hvers barns. Börnin voru öll eintyngd með íslensku að móðurmáli. Heildarfjöldi orða (svokölluð lesmálsorð) í málsýnunum sem liggja til grundvallar bókinni er 100.107 orð, mismunandi orðmyndir eru 7.883 og fjöldi flettiorða er 3.879.

Markmið bókarinnar er að sýna hvaða orð börnum er tamast að nota þegar þau tjá sig á íslensku við fullorðna. Bókin er sú fyrsta sem gefin er út um tíðni orða, beygingarmynda og orðflokka í máli íslenskra barna og fyrsta íslenska orðtíðniþéttur sem byggir eingöngu á talmáli. Gagnaöflun úr talmáli barna er vandmeðfarnari, tímafrekari og kostnaðarsamari en úr talmáli fullorðinna og úr ritmáli, en gefur mikilvægar upplýsingar um málfærni barna og orðaforða þeirra. Bókin er því mikilvægt framlag til fræðanna enda hafa rannsóknir á orðtíðni barna bæði hagnýtan og fræðilegan tilgang. Hagnýtt gildi bókarinnar er að hún nýtist m.a. við kennslu íslensks orðaforða fyrir börn með málþroskaröskun og börn af erlendum uppruna sem eru að læra íslensku, auk þess sem hún nýtist við gerð íslenskra orðaforðaprófa sem prófa skilning ungra barna á orðum af mismunandi tíðniflokkum. Fræðilegt gildi bókarinnar er einnig ótvírætt þar sem hún nýtist bæði í kennslu um máltöku barna og í máltöku-, málfræði- og sálfræðirannsóknum. Einnig má ætla að algeng orð í bókinni gefi ákveðna mynd af menningarheimi íslenskra barna á árunum 2009–2014, þótt orðin *hrun*, *mótmæli* og *Eyjafjallajökull* komi reyndar ekki fyrir í henni.

Í bókinni eru ekki skrár yfir tíðni orða, beygingarmynda og orðflokka barna á mismunandi aldri í gagnasafni Jóhönnu. Bókin nýtist því ekki við rannsóknir á því hvernig orðaforði barna þróast á aldrinum 2;5–7;11 ára. Gögnin sem bókin byggir á eru þó aðgengileg á rafrænu formi til fræðilegrar notkunar innan Háskóla Íslands (sbr. frétt um bókina á vef Háskóla Íslands 23. mars 2020: [https://www.hi.is/frettir/ny\\_bok\\_um\\_tidni\\_orða\\_i\\_tali\\_barna/](https://www.hi.is/frettir/ny_bok_um_tidni_orda_i_tali_barna/)). Það er mikilvægt fordæmi sem eflaust mun nýtast málfræðingum, talmeinafræðingum og sálfræðingum við rannsóknir sínar. Rétt er þó að benda á að eins og kemur fram í töflu 1 í inngangi er fjöldi barna í einstökum aldurshópum, og sérstaklega fjöldi málsýna sem liggur til grundvallar niðurstöðum í hverjum aldurshópi, mjög mismunandi. Nauðsynlegt væri að nota vigtaða tíðni eftir aldurshópum til þess að leiðrétta þessar skekkjur og er mikilvægt að hafa það í huga þegar gögnin eru nýtt.

Einnig má nefna að flest málsýnin sem orðtíðnibókin byggist á eru samtöl við börn í leik- og grunnskólum sem þekkja ekki háskólanemann sem talar við þau. Auk þess voru yfirleitt nýttar fyrstu 50 segðir hvers barns í hverju málsýni, þegar börnin eru e.t.v. enn feimin við rannsakandann. Þótt engin ástæða sé til að ætla annað en orðtíðnibókin endurspeglar grunnorðaförða barnanna og algengustu orðin sem koma fyrir í máli tveggja til átta ára íslenskra barna á árunum 2009–2014 er hugsanlegt að þessi aðferðafræði gefi eitthvað aðra niðurstöðu en ef foreldri eða einstaklingur nákominn barninu hefði safnað málsýnunum og niðurstöður byggist á heildarafritun langra málsýna sem safnað hefði verið nokkrum sinnum í mánuði í nokkur ár (svokallaðar langsníðsrannsóknir).

Að lokum má nefna að þótt skiljanlegt sé að flokkun orða í orðflokka í orðtíðnibók eins og þessari sé erfiðleikum háð má færa rök fyrir því að ganga mætti á annan hátt frá flokkun sumra kerfisorða í orðflokka í bókinni. Orð tungumála eru gjarnan flokkuð í svokölluð inntaksorð og kerfisorð. Inntaksorð eru orð sem bera uppi merkingu setninga, þ.e. nafnorð, lýsingarorð, sagnorð og ákveðnir flokkar atviksorða, en þessir orðflokkar eru opnir þar sem þeir taka við nýjum orðum. Kerfisorð eru hins vegar orð sem gegna málfræðilegu hlutverki innan setninga, þ.e. fornöfn, töluorð, samtengingar, atviksorð, forsetningar, upphrópanir og nafnháttarmerki, en þessir orðflokkar eru lokaðir þar sem þeir taka yfirleitt ekki við nýjum orðum. Flokkun kerfisorða í bókinni í orðflokka fylgir ekki alltaf þeirri flokkun sem lagt er upp með og gefin er upp í efnisyfirliti, þar sem sum orð sem geta tilheyrt mörgum orðflokum eru aðeins flokkuð undir atviksorð og skylda flokka. Þannig er algengasta smáorðið í flokki atviksorða og skyldra flokka orðið *að* og réttilega er tekið fram að það geti verið nafnháttarmerki, atviksorð, forsetning eða samtenging. Orðið *að* er hins vegar ekki talið upp í flokki samtenginga, þótt aukatengingin *að* sé algeng í íslensku. Samkvæmt höfundum (sbr. bls. 15 í inngangi) stafar þetta af því að þess var gætt í úrvinnslu að tvíflokkar ekki orðin þannig að heildarfjöldi þeirra passaði. Málfræðingar sem hyggjast nýta bókina til málfræðirannsókna þurfa því að hafa í huga að sum margræð kerfisorð sem koma fyrir í tali barnanna eru aðeins talin fram í einum flokki og ekki talin upp í orðflokum sem þau tilheyra einnig.

Þegar á heildina er litið er bókinn *Tíðni orða í tali barna* mikilvægt brautryðjendaverk sem veitir upplýsingar um tíðni þeirra orða og beygingarmynda, ásamt upplýsingum um tíðni orða innan einstakra orðflokka, sem koma fyrir í tali 2;5–7;11 ára barna í samræðum við fullorðna í málsýnabanka Jóhönnu Thelmu Einarsdóttur. Bókin hefur bæði hagnýtt og fræðilegt gildi og er gagnleg öllum þeim sem stunda rannsóknir á orðtíðni og máltöku íslenskra barna.

## UM HÖFUNDINN

Sigríður Sigurjónsdóttir er prófessor í íslenski málfræði við Íslensku- og menningardeild Háskóla Íslands. Hún hefur starfað þar við kennslu og rannsóknir í meira en aldarfjórðung eða frá 1994. Sérvið hennar er máltaka íslenskra barna en hún hefur einnig rannsakað breytingar á íslensku nútímamáli og stafrænt málsambýli íslensku og ensku. Þá safnaði hún óbirtum langsníðsgögnum frá tveimur íslenskum stúlkum á árunum 1998–2007.