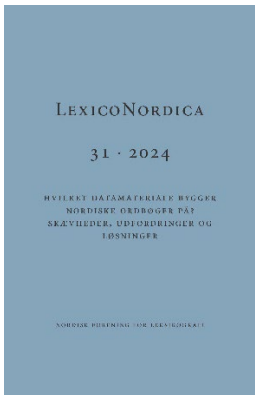


LexicoNordica

Titel:	Talspråkskorpusar som resurs för isländska ordböcker	
Forfatter:	Helga Hilmsdóttir	
Kilde:	LexicoNordica 31, 2024, s. 103-125	
URL:	https://tidsskrift.dk/lexn/issue/archive	

© 2024 LexicoNordica og forfatterne

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Talspråkskorpusar som resurs för isländska ordböcker

Helga Hilmisdóttir

This article focuses on spoken corpora and how they can be used in lexicographic work involving Icelandic. A particular focus is on the *Icelandic youth language* corpus, which gives users access to both texts and recordings. I discuss three domains in which a corpus of spoken language can supplement large text corpora: 1) information on anglicisms, 2) the pragmatic aspects of adverbs, and 3) discourse functions of words or collocations. In addition to this, the paper discusses an alternative method for presenting the peculiarities of spoken language: *Samtalsorðabók*, a dictionary that builds on transcribed excerpts of conversation with sound clips.

1. Inledning

Denna studie fokuserar på talspråkskorpusar och vad de kan bidra med i lexikografiskt arbete på Island. Enligt den europeiska ordbokstraditionen bygger ordboksarbete först och främst på skrivna källor, nuförtiden oftast stora textkorpusar (se diskussion i Eypórs-son 2005; Fjeld 2008; Selback & Svardal 2021). Detta gäller även moderna, isländska ordböcker som *Íslensk nútímamálsorðabók* 'Ordbok över nutida isländska' (ÍNO), som är den enda isländska definitionsordboken som ständigt uppdateras (se Jóhannsson & Úlfarsdóttir, denna volym).

I ÍNO förekommer en del ord som först och främst förknippas med talspråk, t.ex. svordomar, olika dialektala ord, slang och partiklar. Precis som när det gäller andra ordböcker kommer talspråksfenomen dock oftast in i ordboken via stora textkorpusar. Uppslagsorden beskrivs utifrån belägg t.ex. från litterära dialoger eller direkta anföringar i dagstidningar och texter som publicer-

as på webben. Med andra ord beskrivs det som vi uppfattar som talspråk på skriftspråkets villkor (jfr Linell 1982 om ”the written language bias”). På så sätt kan man säga att begreppet *talspråk* tolkas som en stilmarkör (formellt-informellt) snarare än beteckning för språkets medium (talspråk-skriftspråk) (jfr kategorisering i Svensén 2004:379). Samtalsforskningen har dock visat de senaste decennierna att ord och fraser kan ha funktioner i samtal som får en ganska styvmoderlig behandling i ordböcker (Hilmisdóttir 2021). Detta gäller i synnerhet adverb, interjektioner och diskurspartiklar. För att fånga ordens funktioner i det talade språket har Hilmisdóttir (2024) utvecklat en webbordbok som beskriver talspråksfenomen som förekommer i inspelade samtal. *Samtalsorðabók* ’Samtalsordboken’ är en deskriptiv ordbok där användarna har tillgång till såväl ljud som text.

I denna artikel ligger fokus på korpusar som består av talat språk. Syftet med artikeln är att diskutera hur talspråkskorpusar skulle kunna användas i lexikografiskt arbete på Island och eventuellt i övriga Norden och andra ordboksmiljöer. Frågeställningen är som följer: Hur kan talspråkskorpusar användas för att ge en mer nyanserad bild av modern isländska och för att inkludera mer material som är representativt för vardagliga samtal och inte minst ungdomars språkbruk? Vad hör hemma i en traditionell ordbok och hur kan en ordbok som *Samtalsorðabók* komplettera den? Det material som studien baseras på består av tre korpusar som innehåller talat språk: den isländska gigakorpusen *Risamálheildin*, textkorpusen *Íslenskt textasafn* och samtalskorpusen *Íslenskt unglíngamál* ’Isländskt ungdomsspråk’. I diskussionen fokuserar jag på tre olika fenomen i talspråkskorpusarna: 1) anglicismer, 2) pragmatiska funktioner och 3) diskursstrukturerande element.

2. Talspråk i moderna isländska ordböcker

För att ge inblick i hur talspråket har representerats i isländska ordböcker kommer jag att inleda med en jämförelse av tre olika verk: *Islandsk-dansk ordbog* från 1920–1924, *Íslensk orðabók* från 1963 och *Íslensk nútímamálsorðabók* som publicerades på webben 2016.

Islandsk-dansk ordbog av Sigfús Blöndal (1920–1924 och supplement från 1963) var den första ordboken som behandlade modern isländska. Ordboken är deskriptiv till sin karaktär och enligt redaktören var det hans syfte att fånga bl.a. det som han kallar för *almugesprog* 'allmogespråk'. I ordboken som har 154 000 uppslagsord förekommer ett ganska stort antal ord och uttryckssätt som fångats upp i talspråk. Huvudkällorna för talspråksinslagen är först och främst ordinsamlingar på fältet, bl.a. ordlistor som samlades in av redaktörens vänner och kollegor (Blöndal 1920–1924:VIII; Ingólfsson 1997:22; Eyþórsson 2005:16). Talspråksinslagen består av dialektala ord och fraser (Ingólfsson 1997), svordomar, barnspråk, lånord (Óskarsson 1997) och olika diskursmarkörer och interjektioner. En del lånord, dvs. de som ansågs mer problematiska än andra, markerades dock som "Udenlansk Laaneord, alm. i daglig Tale, is. i Byerne; ikke anerkendt i Skriftsprog" (se även Svavarsdóttir & Jónsdóttir 2020:123). Under redigeringsarbetet lades stor vikt vid transkribering av uttal, inklusive lokalt betingade uttalsvarianter (Árnason 1997; Hilmisdóttir & Svavarsdóttir 2023:153–156).

Den första enspråkiga isländska ordboken, *Íslensk orðabók*, kom först ut 1963 men kom ut i ny utgåva 1983, 2002 och 2007. Ordboksbasen bygger delvis på Blöndals ordbok och detta gäller inte minst dialektala och arkaiska ord och uttryck som först och främst förknippas med talspråk. Ordboken bygger även på två stora ordboksarkiv som samlades in av det Lexikografiska institutet i Reykjavík (isl. *Orðabók Háskólans*) men förvaras på Árni

Magnússon-institutet för isländska studier: a) skriftspråksarkivet som består av sedlar med belägg som excerperades från manuskript och tryckta böcker från 1540 till 2000, och b) talspråksarkivet som består av sedlar som visar kommentarer om ord och fraser som kom från Lexikografiska institutets informanter. *Íslensk orðabók* har dock färre uppslagsord, ca 88 000. I förordet beskrivs den som en deskriptiv ordbok som omfattar ”det språk som talar vid alla omständigheter” (förf. övers.). Samtidigt understryks ordbokens normativa roll och att den ska ge fingervisningar om erkänd språkanvändning (*Íslensk orðabók* 2002:VI). Den har med andra ord en språkpuristisk utgångspunkt (*Íslensk orðabók* 2007:VI–VII). Språklig variation framhävs t.ex. inte på samma sätt som i Blöndal (t.ex. färre dialektala ord) och den är konservativ angående ord som anses främmande. I den andra och senare upplagor blev inställningen mindre konservativ. In i lemmalistan togs en del anglicismer som tillhör det vardagliga ordförrådet som t.ex. *ælaener* ’eyeliner’ och slangord och uttryck som har sitt ursprung i ungdomsspråk som *bömmur* ’bummer’ och *sjitt* ’shit’. De markerades dock som slang eller tveksamma främmandeord. I den sista reviderade utgåvan av *Íslensk orðabók* från 2007 reviderades även beskrivningar av olika markörer i talspråk som t.ex. interjektionen *jæja* ’jaha, nå’ som används bl.a. för att markera topikövergångar och attityder. I den nya utgåvan framhävdades mer markörernas funktioner i samtal.

Som tidigare nämnts är *Íslensk nútímamálsorðabók* den enda ordboken över modern isländska som uppdateras löpande (se Jóhannsson & Úlfarsdóttir, denna volym). Den har inte samma lemmalista som de två tidigare ordböckerna, utan bygger på en lista som utvecklades för ordböcker från isländska till nordiska språk (Jónsdóttir & Úlfarsdóttir 2011). Målet med ordboken, som har 56 000 uppslagsord, är i första hand att beskriva centralt, nutida språkbruk. Ordboken har t.ex. endast ett fåtal lokalt präglade ord. Vid ordboksarbetet har redaktörerna delvis använt textkorpor

som *Íslenskt textasafn* och *Risamálheildin* (se avsnitt 3). ÍNO är än så länge mindre än de tidigare två ordböckerna vilket kan medföra konsekvenser för beskrivningen av talspråk, i synnerhet dialektala former och nya lånord. Eftersom ordbokens främsta målgrupp är unga språkanvändare och andraspråkstalare är tanken att skriva korta och relativt enkelt formulerade förklaringar, åtminstone i jämförelse med *Íslensk orðabók*. Det betyder att ord som är ganska komplicerade och har många olika funktioner, såsom partiklar, har minimala förklaringar med huvudvikt på semantiskt innehåll. Till exempel beskrivs uppslagsordet *nú* som ett temporalt adverb utan att kommentera ordets funktion som icke-temporal diskurspartikel (jfr Hilmisdóttir 2016). Den icke-temporal funktionen syns dock i artikeln som två fasta eller halvfasta fraser: *er það nú <bíll>* 'ska det här föreställa <en bil>?' och *þó það nú væri* 'naturligvis'. I ÍNO har varje uppslagsord en ljudfil där man kan höra ordets uttal (uppläsning av skådespelare) men inga kommentarer finns om uttalsvarianter. De tre ordböckerna jämförs i Tabell 1.

	Íslandsk-dansk ordbog (1920-1924)	Íslensk orðabók (2002)	Íslensk nútímamáls- orðabók (inget datum)
uppslagsord	154 000	88 000	56 000
källor för talspråk	ordinsamlingar på fältet	Íslandsk-dansk ordbog och ordinsamlingar	lemmalista som gjordes för ISLEX med tillägg
dialektalt ordförråd	stor vikt vid dialektala ord	en del dialektala ord	sporadiskt
lånord	ganska liberal inställning	mest mycket etablerade ord	mest mycket etablerade ord
partiklar	minimal beskrivning, fasta idiomatiska fraser	etablerade partiklar förklaras med synonymer och ibland kommenterar om funktion	enkel beskrivning, semantiskt innehåll och synonymer

uttal	transkription inkl. varianter	inga uttalsanvisningar	uppläsning av isolerade ord
-------	-------------------------------	------------------------	-----------------------------

Tabell 1: Jämförelse av talspråksinslag i *Isländsk-dansk ordbog*, *Íslensk orðabók* och *Íslensk nútímamálsorðabók*.

Som tabellen visar har utvecklingen gått från att visa språklig variation, inklusive talspråkliga och dialektala drag, till att lägga huvudtyngd på att visa normaliserat språk med huvudtyngd på skriftspråk. Med tanke på den samhällsutveckling som har ägt rum på Island de senaste decennierna där antalet andraspråkstalare har ökat markant finns det dock ett starkt behov för att beskriva talspråksfenomen. Talspråkskorpuser är i det avseendet en viktig resurs.

3. Isländska talspråkskorpuser

I jämförelse med skriftspråksbaserade textkorpuser är talspråkskorpuser begränsade till sitt omfång. På grund av detta spelar det stor roll vilken talare som spelas in, under vilka omständigheter och vilka samtalsämnen det rör sig om. För att ge en bra bild av vardaglig interaktion skulle det vara optimalt att ha en stor korpus med autentiska samtal som spelas in i vardagliga situationer, dvs. samtal som inte är planerade eller styrda. Det kan dock vara svårt att få in tillräckligt mycket material, i synnerhet om korpuser ska vara öppna och finnas tillgängliga på webben.

Än så länge finns det inte allmänt tillgängligt talspråksmaterial för isländska, och det material som finns består mest av transkriberingar av nyhetssändningar och intervjuer som gjorts inom språkforskning. Tabell 2 visar en översikt över tre korpuser som består helt eller delvis av talspråk: talspråksmaterial från den isländska gigakorpuser *Risamálheildin*, talspråksmaterial från textkorpuser *Íslenskt textasafn* och talspråkskorpuser *Íslenskt unglíngamál*.

	Risamál-heildin	Íslenskt textasafn	Íslenskt unglíngamál
ár	2004–2021	2000–2006	2019–2020
genre	nyhetsprogram på radio och tv	vardagliga samtal och gruppintervjuer	gruppintervjuer
storlek	ca 77 miljoner löpord (av 2770 miljoner)	ca 0,5 miljoner löpord (av 65 miljoner)	ca 0,2 miljoner löpord
angiven kontext	en mening, dvs. från stor bokstav till punkt	en rad, 140 tecken	hela inspelningen
ljudfiler	nej	nej	ja

Tabell 2: Översikt över talspråk i isländska korpusar.

Som tabellen visar varierar storleken på de tre korpusarna mycket. Den största korpusen består av en samling nyhetsprogram som finns med i den isländska gigakorpusen. Den består av nästan 77 miljoner löpord och består dels av uppläsning av journalisternas manus, dels av korta ljudklipp från intervjuer. Det finns ingen tillgänglig information om fördelningen mellan planerat och oplanerat tal. Sökorden och meningen som ordet förekommer i visas på skärmen, dvs. från stor bokstav till punkt. Detta betyder att t.ex. interjektioner och partiklar som t.ex. *já* 'ja' eller *nei* 'nej' som ibland förekommer som en egen replik kan förekomma i korpusen utan kontext. Man kan heller inte kontrollera meningarnas yttre kontext, dvs. det som kommer före eller efter repliken i fråga. Användarna har inte tillgång till ljudfilerna för att kontrollera transkriberingen eller för att tolka beläggens prosodiska utformning.

Textkorpusen *Íslenskt textasafn*, som har runt en halv miljon löpord, består av transkriberat talspråk, vardagliga samtal och gruppintervjuer om språkattityder. Sökorden visas i sin kontext, en rad per belägg (140 tecken). När replikerna är korta betyder

det att användaren ibland kan se det som sägs före och efter det aktuella yttrandet, vilket inte är möjligt i gigakorpusen. Ljudinspelningar är däremot inte tillgängliga.

Talspråkskorpusen *Íslenskt unglingamál* samlades in i grund- och gymnasieskolor på Island skolåret 2019–2020. Syftet med insamlingen var att dokumentera unga islänningars samtalsspråk och göra en korpus för framtida forskning. Korpusen gjordes i samarbete med Tekstlaboratoriet i Oslo och finns tillgänglig på webben. Den består av 22 inspelade intervjuer med ungdomar på fyra orter runtom landet. De intervjuade delades in i två åldersgrupper: grundskoleelever i årskurs 9 (14–15 år) och gymnasieelever som gick sitt sista år i skolan (18–19 år). I korpusen deltar 120 olika talare, inklusive intervjuare. I varje intervju deltar fyra ungdomar och en moderator. Under inspelningarna ställer moderatoren några förbestämda frågor till ungdomarna, bl.a. om hur de trivs i skolan, vad de gör på fritiden och hur de ser på framtiden. I slutet spelar moderatoren upp musiksnuttar för att få igång en diskussion om musik, trender och musiksmak.

4. Nedslag i isländska talspråkskorpusar

Men vad kan man använda talspråkskorpusar till i ordboksarbete? Som Selback & Svoldal (2021:104) påpekar beror det på vilka delar av ordförrådet man fokuserar på och vilken typ av ordbok det rör sig om. I traditionella definitionsordböcker är det innehållsorden som dominerar, dvs. främst substantiv och sedan utgör adjektiv och verb något mindre kategorier (Selback & Svoldal 2021:97). För att fånga en stor variation innehållsord behövs dock mycket stora korpusar. Talspråkskorpusar är däremot särskilt användbara när det gäller beskrivningen av funktionsord som t.ex. prepositioner, interjektioner och pragmatiska partiklar (jfr Selback & Svoldal 2021:91).

I följande tre underkapitel diskuterar jag hur talspråkskorpusar kan användas som resurs vid lexikografiskt arbete och visar samtalsutdrag som stöd: resurs för att identifiera anglicismer som är frekventa i talspråk (4.1.), resurs för att identifiera och definiera uppslagsordens pragmatiska (4.2) och diskursstrukturerande funktioner i samtal (4.3).

4.1. Anglicismer i talspråk

Isländsk språkpolitik har ofta beskrivits som konservativ och puristisk och detta har avspeglats i lexikografiskt arbete (Svavarsdóttir & Jónsdóttir 2020:120; Kristinsson 2021). Lånord och ord som upplevs som ”främmande” är fortfarande kontroversiella i det isländska språksamhället. Detta skapar problem för lexikografer som måste ta ställning till vilka ord som ska inkluderas i en ordbok och hur de bör presenteras (Svavarsdóttir & Jónsdóttir 2020:122).

En del danska och engelska lånord har funnits länge i isländskt talspråk men förekommer sällan i skrift (Kvaran 2007:17; se även diskussion i Hilmisdóttir & Peterson 2024:86–87). Förr betydde detta att det var svårt att uppskatta ordens spridning och frekvens i vardagligt språk. I och med sociala medier har detta dock ändrats och t.ex. vid uppdateringen av ÍNO har redaktörerna kunnat använda den isländska gigakorpusen *Risamálheildin* för att få fram information om förekomster av anglicismer i informellt webbspråk.

Tabell 3 visar en översikt över frekvensen på ett urval anglicismer som används i isländska samtal men finns inte i ÍNO. Materialet som jämförs är som följer: 1) talspråksmaterial från gigakorpusen (*Risamálheildin*, delkorpus som består av nyheter på radio och tv), 2) talspråksmaterial från textkorpusen *Íslenskt textasafn*, och 3) talspråkskorpusen *Íslenskt unglíngamál*.

	Risa- málheildin	Íslenskt textasafn	Íslenskt unglingamál
<i>random</i>	19	0	6
<i>kreisi/crazy</i>	15	0	2
<i>level</i>	36	0	0
<i>sökka</i> 'suck'	9	1	3
<i>beisiklí/basically</i>	29	0	14
<i>jess/yes</i>	4	1	31
<i>what/vott</i>	0	0	23

Tabell 3: Belägg på anglicismer i isländska talspråskorpusar.

Som tabell 3 visar ger talspråskorpusar endast ett fåtal belägg av varje ord. De flesta orden i tabell 3 används frekvent i mycket informella texter som en sökning i gigakorpusens material från sociala medier visar: *random* (N=10 236), *kreisi/crazy* (N=7745), *level* (N=16 610), *sökka* 'suck' (N=13 446), *basically/beisiklí* (N=5449), *jess/yes* (N=1920) och *what* (N=0). Dokumenteringen av frekventa anglicismer är på det sättet inte ett starkt argument för att inkludera talspråskorpusar som resurs för ordböcker som det är nu. Beläggen som där finns ger dock autentiska exempel från talspråk och de visar om och hur orden integreras prosodiskt och morfologiskt. Man kan lyssna på beläggen och höra hur de används i sin kontext. Eventuellt kan talspråskorpusar också ge fingervisningar om vilka anglicismer som är de mest frekventa och hur ordförrådet fördelar sig mellan åldersgrupper och olika minoritetsgrupper. Till exempel verkar substantivet *level* 'nivå' vara ett ord som används mycket av vuxna talare medan *what* förekommer enbart i ungdomsspråskorpusen (som är den nyaste korpusen). Som tabellen visar är det i synnerhet de funktionella orden som är frekventa i talspråskorpusarna, ord som *beisiklí*, *jess* och *what*, dvs. ord som inte är lika kontextbundna som innehållsorden som

stys mer av diskussionsämnet. Detta visar att det finns ett behov för en mycket stor talspråkskorpus som skulle kunna ge en mera nyanserad bild av isländskt talspråk och i synnerhet innehållsord.

4.2. Pragmatiska funktioner

En talspråkskorpus kan ge oss information om ordens pragmatiska funktioner i tal. Ett exempel på det är uppslagsordet *örugglega* som är ett relativt frekvent ord i såväl tal som skrift. Enligt ÍNO har adverbet *örugglega* två betydelser (förf. övers.):

1) *án efa, vafalaust, áreiðanlega* 'utan tvekan, definitivt, som man kan lita på'. Ex. *það verður örugglega skemmtilegt í ferðalaginu* 'det blir **säkert** roligt på resan', *ertu örugglega búinn að læsa útidyrnunum?* 'har du **definitivt** låst ytterdörren?'

2) *á sannfærandi hátt, af öryggi* 'på ett övertygande sätt, med säkerhet'. Ex. *liðið sigraði örugglega á mótinu* 'laget segrade **på ett övertygande sätt** i cupen'.

Söker man på *örugglega* i korpusar ser man dock snabbt att adverbet ofta förekommer i sammanhang där talaren är tveksam eller osäker. I talspråkskorpusen använder ungdomarna adverbet *örugglega* oftast i svar på frågor. En vanlig kontext är t.ex. när de berättar om sina önskemål och planer för framtiden. I följande utdrag frågar moderatoren eleverna om de tänker ansöka om plats vid en gymnasieskola eller något annat motsvarande (rader 1–2).¹

1 I transkriptionerna används följande notationer: (0,3) paus angiven i sekunder; (.) mikropaus; [överlappning inleds;] överlappning slutar; ja- avbr; :: förlängd vokal; hh utandning; .hh inandning; ** med skratt i rösten.

1. Ansöka till ett gymnasium: IU

(M=moderator, L=Lárus, S=Siggi, T=Tommi)

- 01 M ætliði í einhvern- ætliði í menntaskóla
ska ni gå nán, ska ni gå nán gymnasieskóla
- 02 eða eitthvað svoleiðis
eller något sånt
- 03 L .hnf
- 04 S uh
- 05 T mm já
mm ja
- 06 L **örugglega** bara Emm E: sko
örugglega bara ME
- 07 M mhm
- 08 S hhh já hh
hhh ja hh
- 09 L eða (0,4) Verkmenntaskólann á
eller yrkesskólan i
- 10 Seyðisfirði
Seyðisfjörður
- 11 S ókei
okej

Efter tvekan (rad 4) och ett kort positivt svar (rad 5) utvecklar Lárus svaret och förklarar vilken gymnasieskola han vill söka till. Turen inleds med *örugglega* vilket här signalerar att svaret inte är definitivt. Lárus osäkerhet blir ännu synligare i hans följande tur där han tillägger namnet på en annan skola han kanske kan tänka sig söka till (rad 9–10).

Után *örugglega* i exempel (1) skulle svaret ge ett mera bestämt intryck. Beläggen i korpusen tyder på att i talspråk används adverbet *örugglega* uttryckligen för att indikera att talaren är osäker på sitt svar. Svaren är oftast korta och har formen *örugglega*

+ 'ett möjligt eller sannolikt svar'. För att representera detta i ord-boken borde man tillägga ett betydelsemoment som avspeglar detta.

ég held, ég geri ráð fyrir 'jag tror, jag utgår ifrån' Ex. *-hvað verður í hádegismatinn? -örugglega fiskur* '-vad får vi till lunch? -fisk **tror jag**', *-ertu að vinna á morgun? -örugglega* 'jobbar du imorgon? -jag **utgår ifrån** det'

Osäkerheten som *örugglega* indikerar framgår tydligt när man ser ordets kontext, dvs. man måste ha tillgång till de repliker som kommer före och efter för att kunna tolka funktionen. I en text-korpus som inte visar kontexten kan det vara svårare att sätta fingret på denna användning. Beläggen i gigakorpusen är många, eller drygt 345 000, och man ser inte det som kommer före och efter meningen som innehåller belägget. Genom användningen av en talspråkskorpus får ordboksredaktören tillgång till större kontext och kan lyssna själv på samtalsutdraget. På så sätt är det enklare att tolka ordets egentliga funktion eller innebörd i samtal.

4.3. Diskursstrukturerande funktioner

I talspråkskorpusar kan man hitta funktioner som inte syns i skrivna texter. Detta gäller t.ex. vissa diskursstrukturerande funktioner som bl.a. adverb kan ha. Ordet *allavega* (*alla vega*) och *allavegana* (*alla vegana*) är ett exempel på detta.

Enligt ÍNO har uppslagsordet *allavega* två betydelser (förf. övers.):

1) *af mörgum gerðum, fjölbreytilega* 'av olika slag'. Ex. *hann hlustar á allavega tónlist* 'han lyssnar på all slags musik'.

2) *að minnsta kosti* 'åtminstone'. Ex. *við getum ekki klifrað*

upp í þetta tré, allavega ekki ég 'vi kan inte klättra upp i detta träd, åtminstone inte jag'.

Varianten *allavegana* finns också med i ÍNO men ordboksanvändaren hänvisas direkt till *allavega*. Samma gäller för de särskrivna varianterna *alla vega* och *alla vegana*.

I talspråskorpusen *Íslenskt unglíngamál* förekommer totalt 202 belägg: *allavega* (N=120) och *allavegana* (82). Inget av beläggen används i första betydelsen, dvs. 'av olika slag'. Istället förekommer *allavega* som någon typ av diskursstrukturerande element som används antingen för att precisera eller avgränsa ett tidigare påstående som i exemplet ovan, eller som en återgångsmarkör (jfr Ottesjö 2005 om *iallafall* i svenska).

En återgångsmarkör är en markör som används för att återgå till en berättelse efter ett avbrott, som t.ex. reparationer som har som syfte att utreda oklarheter i det som sagts. Följande utdrag visar ett exempel på detta. I utdraget berättar eleven Embla för de andra hur hennes släkting har byggt robotar som användes i en musikvideo utan att få någon uppmärksamhet för det (rader 1–9). Hon nämner inte namnet på låten men moderatoren kommer med ett förslag (rad 10) som leder till att det blir ett avbrott i Emblas berättelse, dvs. en reparationssekvens som skjuts in. Avbrottet eller reparationen markeras med bokstaven A i marginalen (rader 10–12).

2. Robotar för musikvideo: IU

(M=moderator, E=Embla)

- 01 E eitt tónlistarmyndbandið þar er
ett av musikvideorna, där finns
- 02 bjó frændi minn til eitthver
gjorde min släkting några
- 03 ótrúlega flotta hvíta: (.)
otroligt fina vita

- 04 vélmenni fyrir (.) myndavé- (.)
robot för kamer- ...
- 05 myndir og hérna: .hhh hann fékk
bilder och ehm han fick
- 06 ekkert svona kredit fyrir það
inget credit för det
- 07 það er þínu s-[(.)] þínu
det är lite s... lite
- 08 M [já]
ja
- 09 leiðinlegt
tråkigt
- A 10 M love and happiness
- A 11 E já [ábyggilega]
ja säkert
- A 12 E [held ég] að það heiti
tror jag att det heter
- 13 E **allavegana** já hann- hann bjó
allavegana, ja han- han byggde
- 14 E til þann róbót
den roboten
- 15 M [já ókei]
ja okej
- 16 E [og (.)] hann er bara með þá í
och han har dem i
- 17 kjallaranum sínum *heh*
sin källare heh

Efter att reparationen är slutförd återgår Embla till sin berättelse och upprepar delvis det som hon har tidigare sagt (att det var hennes släkting som byggde roboten). Sedan tillägger hon ny information, dvs. att roboten fortfarande finns bevarad i hans källare (rader 16–17). Återgången till den pågående berättelsen markeras med *allavegana*.

Återgångsmarkören *allavega* förekommer som ett initialt annex, dvs. den är inte syntaktiskt integrerad i den återstående delen av yttrandet. Den används som en turinledning som knyter ihop det som talaren kommer att säga med det som hon har sagt tidigare. Med andra ord har *allavegana* en funktion som inte kan hänföras till ÍNO:s beskrivning av uppslagsordet, dvs. det betyder varken 'av olika slag' eller 'åtminstone'. För att inkludera även denna funktion kunde man tillägga ett tredje moment: 'används för att återknyta till huvudtråden efter avbrott eller utvikning'.

Belägg på *allavega* och *allavegana* som visas i exempel (2) är sällsynta i textkorpusar. Dessutom visas de belägg som eventuellt finns i korpusarna inte i tillräckligt stor kontext för att man ska kunna avgöra vad de har för funktion i sammanhanget.

5. Alternativa metoder för att presentera talspråk

En relevant fråga som kommer upp efter denna genomgång är om en definitionsordbok som ÍNO ska inkludera talat språk. Svaret beror först och främst på vad man har för syfte med ordboken och vilken målgrupp den riktar sig till (se diskussion i Selback & Svardal 2021; Eyþórsson 2005). Som tidigare nämnts är ÍNO en medelstor ordbok som fokuserar mest på standardskriftspråk. Den används i grund- och gymnasieskolor och har en normativ roll, även om där också finns uppslagsord som enligt ordboken inte är helt accepterade och markerade som sådana. Eventuellt kunde man fokusera på alternativa lösningar för att beskriva talspråket på ett grundligare sätt.

Den talspråksbaserade ordboken *Samstalsordabók* är ett exempel på en alternativ lösning som kan komplettera ÍNO (jfr *Ordbog over Dansk Talesprog*, se t.ex. Hansen 2015). *Samstalsordabók* fokuserar på interaktiva element som t.ex. pragmatiska markörer, dis-

kurspartiklar, svordomar, hälsnings- och avskedsfraser och tilltal (Hilmisdóttir 2023). Som analysen i avsnitt 4 visar kan markörer i samtal ha en lång räckvidd, dvs. de kan knyta ihop yttranden som inte kommer direkt efter varandra. Enligt traditionen är exempel i vanliga definitionsordböcker korta, oftast en replik eller högst två. I *Samtalsorðabók* används längre utdrag, oftast 6–12 rader. Ordboksanvändaren kan läsa transkriptionen på skärmen, lyssna på hela ljudklippet eller höra själva ordet utan kontext. I ordboken formuleras uppslagsordets funktion och användaren kan slå upp andra ord som har liknande funktioner. Ordboken är deskriptiv och gör ingen skillnad på ord som har funnits länge i språket (t.ex. *ha 'va'*) eller nya lån från engelska (t.ex. *whatever, basically*). Uppslagsorden kan vara enstaka ord (t.ex. *þúst 'du vet'*) eller fraser (t.ex. *allt í lagi*). I ordboken finns även element som inte klassificeras som ord i ordböcker, som t.ex. *ah* och *mhm*.

Figur 1 visar uppslagsordet *ah* som definieras som 'indikerar att talaren har förstått något som han eller hon inte visste tidigare'. Samma exempel används även för att belysa ett av tre betydelsemoment för uppslagsordet *ha 'va'*: 'används som reparationsmarkör som visar att talaren inte har hört det som den andra sa och att han måste upprepa sina ord, dvs. 'skulle du kunna upprepa det du sa just nu?'. Inspelningen är från ett samtal mellan två 15-åriga pojkar.

Uppslagsorden i *Samtalsorðabók* begränsas till ord som har pragmatiska och diskursiva funktioner. Ordboken har använts bl.a. i undervisning för andraspråkstalare vid Islands universitet som är en av ordbokens främsta målgrupper.

Vid arbetet med *Samtalsorðabók* blev det dock tydligt att det finns ett behov av större talspråkskorpusar med varierat material. En korpus som *Íslenskt unglíngamál* är ganska ensidig eftersom alla inspelningarna består av gruppsamtal som styrs av en moderator. När man samlade in och transkriberade ljudklipp för *Samtalsorðabók* stötte man även på ett problem angående personssekretess. Samtal som spelas in som forskningsmaterial kan



Figur 1: Exempel i *Samtalsorðabók*.

användas som korpusmaterial men ljudfilerna kan inte läggas ut öppet på webben. Vid redigeringen av *Samtalsorðabók* samlade man därför in radiosamtal och poddar som redan var tillgängliga på nätet och tog kontakt med moderatorerna för att få tillstånd att använda ljudklipp i en ordbok.

6. Sammanfattning och avslutande diskussion

I denna artikel har jag diskuterat hur talspråk har inkluderats i moderna isländska ordböcker och hur talspråkskorpusar skulle kunna ge en bättre och mer nyanserad bild av uppslagsordens betydelse eller funktion i talspråk. I artikeln diskuterade jag tre olika fenomen i talspråkskorpusarna: 1) anglicismer, 2) pragmatiska funktioner och 3) diskursstrukturerande element.

Eftersom talspråkskorpusarna tillsvidare är små bygger deras värde först och främst på vad de kan berätta om funktionsord och hur de används i samtal. I ÍNO är funktionsordens semantik ofta underbelyst, eventuellt på grund av att a) den bygger först

och främst på textkorporusar och b) den har som mål att beskriva modern isländska på ett tillgängligt sätt, inte minst med tanke på de yngre generationerna och andraspråkstalare. Ibland fokuseras enbart på uppslagsordens semantiska innebörd (t.ex. *örugglega* 'med säkerhet') även om detta inte nödvändigtvis återspeglar nutida språkbruk. Däremot visar exempelmeningarna ibland belägg som har partikelfunktioner, dvs. de står utanför yttrandets propositionella innehåll (t.ex. *nú* 'nu', *sko* 'ser du'). För att komma åt dessa funktioner måste man ha tillgång till en stor och varierande talspråkskorpus som består av såväl inspelningar som transkriptioner.

För tillfället har ÍNO runt 56 000 uppslagsord vilket betyder att ordboken är betydligt mindre än de äldre verken (156 000 respektive 88 000 uppslagsord). ÍNO är ganska normativ medan *Isländsk-dansk ordbog* från 1920–1924 är utpräglat deskriptiv och *Íslensk orðabók* från 1963 enligt förordet eftersträvar en balansgång mellan att vara normativ och deskriptiv. Syftet med ÍNO är först och främst att fokusera på standardspråk och centralt ordförråd och tanken är att den inte ska innehålla "onödiga" och föråldrade uppslagsord. Eventuellt beror detta på att ÍNO bygger på en lemmalista som har sitt ursprung i ISLEX, som är en isländsk-skan-dinavisk ordbok avsedd för språkinlärare på Island och i Norden (se Jóhannsson & Úlfarsdóttir, denna volym). I lemmalistan finns varken många lokalt präglade ord och betydningar eller nya lånord. En mycket stor och varierande talspråkskorpus skulle kunna användas för att få fram ny och uppdaterad information om centralt ordförråd i modernt isländskt talspråk, men en sådan korpus återstår att samla in.

Avslutningsvis kan man fundera på om allmänna ordböcker kan göra anspråk på att omfatta såväl tal som skrift. Går det att integrera så olika former i ett och samma uppslagsverk? Talat språk kännetecknas först och främst av att det är kontextbundet och dialogiskt (se t.ex. Linell 1982:5–11). De som ingår i ett samtal befin-

ner sig oftast i samma rum och de kan använda rösten, ansiktsuttryck, kroppsspråk och fysiska föremål för att kommunicera med varandra och försäkra att det inte uppstår några missförstånd. I samtal kan vi därför använda mer deiktiska och allmänna ord än vi gör i skrift, och i samtal kan det uppstå ord som är t.ex. bundna till orter (dialektala ord) eller vissa samhällsgrupper (slang). När vi talar måste vi också förhandla med vår samtalspartner om när och hur länge vi får tala och hur det vi säger anknyter till det pågående samtalet, vilket förklarar det frekventa partikelbruket i talspråk. När vi uttrycker oss i skrift används språket däremot utan yttre kontext. Den som skriver tilltalar läsare som oftast befinner sig i en annan tid och annat rum. Skriftspråket är monologiskt till sin natur och skribenten har ett mindre behov av partiklar och andra diskursstrukturerande element än den som talar. Innehållsorden har däremot en tendens att få ökad tyngd i skrift. Ordförrådet blir större och mera beskrivande och precist och vi behöver inte lika stort sammanhang för att tolka meddelandet. Det ser vi t.ex. i traditionella ordböcker där exempelmeningar ofta består av en kort replik. Som min analys har visat behövs ofta en större kontext i talspråk, ibland flera repliker.

Idealiskt ska en allmän ordbok behandla både tal- och skriftspråk så att en vanlig ordboksanvändare kan hitta svar på ett ställe. Men med tanke på att den lexikografiska traditionen först och främst har grundats på skriftspråk finns det också behov för alternativa ordböcker som fokuserar på talspråk och möjliggör en fördjupad analys på talspråkets egna villkor.

Referenser

Ordböcker, korpusar och digitala resurser

Blöndal, Sigfús (1920–1924): *Íslensk-dönsk orðabók*. Reykjavík: Íslensk-dönsk orðabókarsjóður.

- Blöndal, Sigfús (1963): *Íslensk-dansk ordbog. Supplement*. Reykjavík: Íslensk-danskur orðabókarsjóður.
- Íslensk nútímamálsorðabók*. Þórdís Úlfarsdóttir & Halldóra Jónsdóttir (red.). Árni Magnússon-institutet för isländska studier. <islenskordabok.arnastofnun.is> (februari 2024).
- Íslensk orðabók handa skólum og almenningi* (1963). Árni Böðvarsson (red.). Reykjavík: Bókaútgáfa Menningarsjóðs.
- Íslensk orðabók handa skólum og almenningi*. Önnur útgáfa, aukin og bætt (1983). Árni Böðvarsson (red.). Reykjavík: Bókaútgáfa Menningarsjóðs.
- Íslensk orðabók*. Þriðja útgáfa, aukin og bætt (2002). Mörður Árnason (red.). Reykjavík: Edda.
- Íslensk orðabók*. Fjórdða útgáfa byggð á 3. prentun frá 2005 með allnokkrum beytingum (2007). Mörður Árnason (red.). Reykjavík: Edda.
- Íslenskt textasafn*. Árni Magnússon-institutet för isländska studier. <corpus.arnastofnun.is> (apríl 2024).
- Íslenskt unglíngamál*. Helga Hilmisdóttir (red.). Tekstlaboratoriet vid Universitetet i Oslo. <tekstlab.uio.no/glossaz/iyl> (februari 2024).
- Ordbog over Dansk Talesprog*. Carsten Hansen (red.). <odt.hum.ku.dk/om/talesprogtilordbog> (februari 2024).
- Risamálheildin*. Árni Magnússon-institutet för isländska studier. <malheildir.arnastofnun.is> (februari 2024).
- Samtalsorðabók*. Helga Hilmisdóttir (red.). Árni Magnússon-institutet för isländska studier. <samtalsordabok.arnastofnun.is> (apríl 2024).

Annar litteratur

- Árnason, Kristján (1997): Hljóðfræðiathuganir Jóns Ófeigssonar. I: *Orð og tunga* 3, 71–78.
- Eyþórsson, Aðalsteinn (2005): Hver er kjarni orðaforðans? I: *Orð og tunga* 7, 9–20.

- Fjeld, Ruth Vatvedt (2008): Talespråksforskningens betydning for leksikografien. I: Janne Bondi Johannessen & Kristin Hagen (red.): *Språk i Oslo: ny forskning omkring talespråk*. Oslo: Novus. 15–28.
- Hansen, Carsten. (2015): Beskrivelsesproget i *Ordbog over Dansk Talesprog*. I: *LexicoNordica* 22, 57–75.
- Hilmisdóttir, Helga (2016): *Nú* in Icelandic conversation. I: Peter Auer & Yael Maschler (red.): *NU/NÁ. A family of discourse markers across the languages of Europe and beyond*. Berlin/Boston: de Gruyter. 409–441.
- Hilmisdóttir, Helga (2021): Talspråkskorpuser, diskurspartiklar och lexicografi. I: *LexicoNordica* 28, 79–100.
- Hilmisdóttir, Helga (2023): Pragmatiska markörer i samtal – en webbaserad ordbok för isländskt talspråk. I: Louise Holmer et al. (red.): *Nordiska studier i lexicografi* 16. Rapport från 16:e konferensen om lexicografi i Norden, Lund 27–29 april 2022. 127–139.
- Hilmisdóttir, Helga & Ásta Svavarsdóttir (2023): Icelandic dialect classification. I: *Dialectologia Special issue* 11, 147–175.
- Hilmisdóttir, Helga & Elizabeth Peterson (2024): Language contact and language change: Impact on the languages of the Nordic countries. I: Elizabeth Peterson & Kristy Beers Fägersten (red.): *English in the Nordic Countries. Connection, Tensions, and Everyday Realities*. London: Routledge. 84–103.
- Ingólfsson, Gunnlaugur (1997): Mállýskuorð. I: *Orð og tunga* 3, 21–24.
- Jóhannsson, Ellert & Þórdís Úlfarsdóttir. *Ordbog over moderne islandsk – udvikling og tilføjelser*. I: *LexicoNordica* 31 (denne volum).
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2011): ISLEX – en flersproget nordisk ordbog: I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexicografi* 11. Rapport från Konferensen om lexicografi i Norden, Lund 24–27 maj 2011. Lund: Skrifter utgivna av Nordiska föreningen för lexicografi 12. 353–366.

- Kristinsson, Ari Páll (2021): Et indblik i islandsk deskriptiv og præskriptiv leksikografi: Importord i to nye digitale islandske ordbøger. I: *LexicoNordica* 28, 121–138.
- Kvaran, Guðrún (2007): Undersøgelse af afløsningsord i de nordiske sprog. Indledning. I: Guðrún Kvaran (red.): *Udenlandske eller hjemlige ord? En undersøgelse af sprogene i Norden*. Oslo: Novus. 9–18.
- Linell, Per (1982): *The written language bias in linguistics*. Linköping: University of Linköping.
- Ottesjö, Cajsa (2005): Iallafall som diskursmarkör. I: Jan Anward & Bengt Nordberg (red.): *Samtal och grammatik – Studier i svenskt samtalspråk*. Stockholm: Studentlitteratur. 201–229.
- Óskarsson, Veturliði (1997): Tæk og miður tæk í Blónðsal-sorðabók. I: *Orð og tunga* 3, 25–34.
- Selback, Bente & Terje Svardal (2021): LIA-korpuset som ressurs i revisjonen av tre ordbøker. I: Kristin Hagen, Gjert Kristoffersen, Øysten A. Vangsnes, Tor A. Åfarli (red.). *Språk i arkiva. Ny forskning om eldre talemål frå LIA-projektet*. Oslo: Novus forlag. 87–107.
- Svavarsdóttir, Ásta & Halldóra Jónsdóttir (2020): Kontroversielle ord i purismens land. I: *LexicoNordica* 27, 127–136.
- Svensén, Bo (2004): *Handbok i lexicografi: Ordböcker i teori och praktik*. Stockholm: Norstedts akademiska förlag.

Helga Hilmisdóttir
Forskningsdocent, fil.dr.
Árni Magnússon-institutet för isländska studier
Edda, Arggrímsgata 5
IS-107 Reykjavík
helga.hilmisdottir@arnastofnun.is