

# Nýjasta tækni og málvísindi

## 1. Yfirburðir ensku

Undanfarna tvo áratugi, meðal annars í kjölfar aukins aðgengis að stafrænni tækni, hefur umfjöllun um yfirburði ensku gagnvart öðrum tungumálum fæst verulega í aukana og verið sett í samhengi við þá hættu sem myndast þegar enska er notuð í sífellt fleiri umdæmum annarra mála, þannig að notkunarvið þeirra skerðist og þá ef til vill framtíðarhorfur þeirra einnig (Eiríkur Rögnvaldsson o.fl. 2012, Eiríkur Rögnvaldsson 2016, Iris Edda Nowenstein og Sigríður Sigurjónsdóttir 2021).<sup>1</sup> Segja má að sérstök áhersla hafi verið á stafræn umdæmi og ljóst er að máltækniáætlanir stjórnvalda hafa að hluta til haft það að markmiði að hægt væri að nota nýja tækni á íslensku en ekki einungis ensku (*Samtök atvinnulífsins* 2018). Kapphlaupið við ensku þegar kemur að máltækni framförum er þó strembið, og oft gleymist í umræðu um nýjustu tækni að hún hefur ekki verið útfærð fyrir fjölda tungumála (Bender og Friedman 2018).

Í þessari grein verður lögð áhersla á nýjustu máltæknina sem hefur náð flugi, svokallaða myndandi gervigreind (e. *generative artificial intelligence* (AI)), sem styðst við stór mállíkön eða risamállíkön (e. *large language models*, LLMs). Dæmi um slíka tækni eru til að mynda Bard frá Google og GPT-líkönin sem fyrirtækið OpenAI hefur gert aðgengileg almenningi með spjallmenninu ChatGPT. Í umfjölluninni verður helst miðað við ChatGPT sem

hefur vissa sérstöðu í íslensku samhengi, þar sem ráðist var í sérstakt þjálfunaráttak með mannlegri endurgjöf fyrir íslensku í fjórðu útgáfu þess, GPT-4, en hingað til hafði slíkt aðeins verið gert út frá ensku (*OpenAI* 2023, *Miðeind* 2023). Myndandi gervigreind með stórum mállíkönum hefur orðið sérstaklega fyrirferðarmikil í almennri málvísindaumræðu, en sú umræða verður hér reifuð og skoðuð sérstaklega út frá virkni GPT-4 í íslensku.

## 2. Mikill en umdeildur árangur

Þrátt fyrir skjótan og umfangsmikinn árangur gervigreindarlausna á borð við stór mállíkön er nauðsynlegt að staldra við og vekja athygli á varúðarorðum og gagnrýnisröddum sem hafa sprottið samhliða árangrinum. Segja má að gagnrýni á notkun stórra mállíkana megi flokka í tvennt, annars vegar gagnrýni á framsetningu þeirra (t.d. sem spjallmenna) og möguleikana sem þau bjóða upp á (t.d. í stað leitarvéla) og hins vegar gagnrýni á starfshætti áberandi fyrirtækja í máltæknibransanum. Hér verður farið stuttlega yfir helstu atriði þessarar gagnrýni áður en fjallað er meira um þýðingu stórra mállíkana fyrir málvísindi almennt.

Innan gagnrýni á framsetningu og virkni rúmast fjölbreytt sjónarhorn. Í fréttamiðlum hefur til að mynda talsvert verið fjallað um varúðarorð áhrifamanna sem snúa að því að gervigreindartækni sé að verða of öflug og ógni jafnvel framtíð

1. Við þökkum fyrir gagnlegar umræður um stór mállíkön á málþingi Íslenskrar málnefndar 4. maí 2023; þar áttum við sérlega gagnlegt spjall við Hauk Barra Simonarson. Einnig þökkum við fyrir samræður um efnið við Önnu Sigríði Práinsdóttur í þættinum Orði af orði á Rás 1, Atla Jasonarson, Elmar Unnsteinsson, Steinþór Steingrímsson og Þórhall Eypórsson. Allt sem kann að vera missagt um mállíkön og annað í greininni skrifast þó vitaskuld á höfunda.

mannkynsins (sbr. RÚV.is 2023). Þessari ógn hafa aðrir gagnrýnendur hafnað algjörlega með þeim rökum að framfarir í gervigreind séu alls ekki svo miklar og að um sé að ræða tilefnislausan hræðsluáróður sem ýki virkni gevigreindarlausna og beini athygli almennings frá raunverulegum skaða sem lausnirnar valda nú þegar (Bender og Hanna 2023).

Yfirlit yfir slíkan skaða má að miklu leyti finna í tveimur greinum sem Emily M. Bender, prófessor í máltækni við háskólann í Washington, skrifaði ásamt meðhöfundum sínum 2020 og 2021. Að vísu eru ekki allir höfundar seinni greinarinnar, On the dangers of stochastic parrots: Can language models be too big? (Bender o.fl. 2021), skráðir fyrir henni. Um var að ræða starfsfólk Google sem þurfti að draga til baka þátttöku sína í greininni eftir að fyrirtækið hætti við að samþykkja birtingu hennar. Tveir þessara höfunda, Timnit Gebru og Margaret Mitchell, neituðu að draga nafn sitt til baka; sú síðari lét breyta höfundarnafni sínu við greinina í Shmargaret Shmitchell og dró þannig athyglina að því sem Google fór fram á og vísar þetta þannig til þeirra höfundarnafna sem voru strokuð út (sbr. Weil 2023). Þær misstu báðar starf sitt hjá Google í kjölfarið og vinna nú að hluta til sjálfstætt að svokölluðum siðferðismiðaðum gervigreindarlausnum (e. *ethical AI*). Í greinunum tveimur (Bender og Koller 2020; Bender o.fl. 2021) er sérstaklega varað við bjögum (e. *biases*) sem fram koma í líkönunum og eru mögulega hættulegir jaðarsettum hópum (sjá einnig Atari o.fl. 2023 og Marcus 2022 og 2023) og nauðsyn þess að taka öllu efni sem líkönin framkalla með miklum fyrirvara þrátt fyrir markaðssetningu sem gefur til kynna áreiðanleika sem áþekkur er vandaðri upplýsingaleit með hefðbundnum leitarvélum. Þekkt er að spjallmenni á borð við ChatGPT geti myndað sannfærandi ósannindi (Kidd og Birhane 2023) og þannig magnað upp upplýsingaóreiðu (Bender o.fl. 2021, Marcus 2022 og 2023). Þetta tengist að einhverju leyti gagnrýninni á starfshætti fyrirtækja, en sömu gagnrýnendur hafa ítrekað bent á ókosti þess að upplýsingar um þjálfunargögn stórra mállegra líkana á borð við GPT-4 séu ekki aðgengilegar.

Slíkar upplýsingar myndu auðvelda rannsakendum að leggja mat á bjaga og aðra skaðlega þætti í virkni líkananna og afla nauðsynlegrar þekkingar á eðli þeirra. Lögð hefur verið áhersla á kosti þess að hafa þjálfunargögn í opnum aðgangi í umræðum um mögulega löggjöf sem snýr að virkni stórra málíkana (Bender og Hanna 2023).

Annað sem gagnrýnt hefur verið varðandi starfshætti stórra fyrirtækja sem framleiða stór myndandi málíkön eru möguleg brot á höfundarréttarlögum (Lucchi 2023), umhverfisáhrif vegna þeirrar miklu orku sem nauðsynleg er til að þjálfna líkönin (Bender o.fl. 2021) og loks notkun ódýrs vinnuafis við óviðunandi starfsaðstæður við þjálfun líkananna (Perrigo 2023). Af þessari umfjöllun er ljóst að stíga þarf varlega til jarðar þegar kemur að notkun og upphafningu stórra myndandi málíkana, og á það einnig við um mögulegt framlag þeirra til fræðilegra kenningasmíða.

### 3. Stór málíkön og málvísindi

Við beinum næst sjónum okkar að málvísindum m.t.t. málíkana og ræðum málgetu þess líkans sem hefur verið mest í umræðunni undanfarið, GPT-4. GPT-líkonin frá OpenAI liggja til grundvallar viðmótinu eða spjallmenninu ChatGPT sem veitir notendum möguleika á að setja fram fyrirspurnir á mjög einfaldan hátt og sjaldnast stendur á svari frá ChatGPT. GPT-4 þykir hafa svo gott vald á ensku (þ.e. mannlegu máli) að innan málvísinda er jafnvel talað í fúlustu alvöru um dauða kenninga Noams Chomskys (sjá einkum Piantadosi 2023) eins og við komum nú að.

#### 3.1 Málgeta málíkana

Chomsky er faðir svokallaðrar málkunnáttufræði (eða generatífrar málfræði, e. *generative grammar*) en þar hefur mikið verið gert úr því að börn ná á skömmum tíma að tileinka sér tungumál þrátt fyrir að ílagið – málið sem þau heyra í máltökunni – sé takmarkað. Á ensku gengur þetta undir heitinu *poverty of the stimulus*; á íslensku hefur verið talað um fábreytileika eða fátækt máláreitisins (Sigriður Sigurjónsdóttir 2013:108, Þórhallur Eyþórsson

2013:152). Chomsky telur hæfni til máls einstaka fyrir manninn; hann hafi sérstakan málhæfileika sem aðrar lífverur hafi ekki. Þetta dregur Sigríður Sigurjónsdóttir (2013:108) saman á eftirfarandi hátt:

Kenning Chomskys um meðfæddan málhæfileika manna gerir [...] ráð fyrir að menn hafi áskapaða málfræðikunnáttu, eða nánar tiltekið að þau málfræðiatríði sem eru algildi í tungumálum séu mönnum meðfædd. Þetta er hin svokallaða algilda málfræði sem málkunnáttufræðingar leitast meðal annars við að varpa ljósi á með rannsóknum sínum.

En hvað þá ef tölvur, eða réttara sagt stór mállíkön, geta náð góðum tókum á tungumálinu, sem á að vera okkur meðfætt?

Innan a.m.k. málkunnáttufræðinnar hafa þróast margvíslegar tilgátur og kenningar um málkerfið og hvernig reglum við beitum þegar við tileinkum okkur málið sem við tölum. Málkunnáttufræðingar hika t.a.m. ekki við að tala um alls konar færslur á setningarhlutum; sem dæmi er gjarnan gert ráð fyrir í *hv*-spurningu eins og *Hvern sá María?* að persónubeygða sögnin færir fram fyrir frumlagið *María* og andlagið *hvern* færir fram fyrir bæði persónubeygðu sögnina og andlagið. Setjum sem svo að mállíkanið GPT-4 eigi ekki í vandræðum með þetta og veltum nú fyrir okkur tveimur sviðsmyndum. Í þeirri fyrri gefum við okkur að í máli manna hafi verið sýnt fram á að alls konar færslur verki, þar á meðal færslurnar sem nefndar voru fyrir spurninguna *Hvern sá María?* Þýðir þetta, úr því að við segjum að mállíkanið eigi ekki í vandræðum með spurnarsetningar á borð við *Hvern sá María?*, að mállíkanið hafi lært færslur og færi persónubeygða sögn og spurnarliðinn *hvern* fram fyrir frumlagið?<sup>2</sup>

Í síðari sviðsmyndinni segjum við að það sé ekkert

meira en tilgáta að færslur verki í mannlegu máli og jafnframt að það séu engar líkur á að mállíkön eins og GPT-4 færi setningarhluta þegar þau mynda spurnarsetningar eins og *Hvern sá María?* Getum við notað þá vitneskju sem við höfum um mállíkön til að hrekja tilgátur um máltökufærli og málkunnáttu mannsins? Svar Stevens Piantadosis virðist vera já en hann skrifaði nýverið grein þar sem hann heldur því fram að árangur stórra mállíkana afsanni kenningar Chomskys.

Piantadosi (2023) leggur áherslu á að stór mállíkön feli í sér tilgátur og kenningar um mál; nái tiltekið mállíkan góðum árangri sé það til marks um góða tilgátu. Þegar mállíkan verður til eru svokallaðar færribreytur stilltar og í tilfelli stórra mállíkana er fjöldi færribreyta óheyrilegur. Ef mállíkan stendur sig vel í tilteknu málfræðilegu atriði hefur vel tekist til við stillingu færribreyta og málfræðitilgátan er þá góð, að mati Piantadosis. Stór mállíkön eins og GPT-4 byggja á tölfræði og líkindum; þau eru þjálfuð á því að giska á næsta orð í segð. Þetta er auðvitað mjög ólíkt því sem margar kenningar í málkunnáttufræði ganga út á, þar sem t.d. hafa verið sett fram sérstök bindilögmál (t.d. Chomsky 1981) og eyjahömlur (t.d. Ross 1967) sem eiga jafnvel að vera algildar í málum heims. Það er kannski ólíklegt að stór mállíkön eins og GPT-4 þurfi á slíkum lögmálum og hömlum að halda við að ná tókum á t.d. ensku heldur beita þau öðrum aðferðum með góðum árangri – og þá er spurningin hvort það sýni að við mannfólkið gerum það ekki heldur. Engu að síður virðast dæmi um það að GPT-4 hafi ekki náð fullkomnum tókum á sumum eyjahömlum; Katzir (2023) ræðir eina slíka hömlu sem hann segir aftur á móti að öll enskumælandi börn hafi full tök á.<sup>3</sup> Það vekur sömuleiðis spurningar um hvort það sé e.t.v. talsverður munur á málkunnáttu barna og mállíkana.

En jafnvel þótt Piantadosi (2023) telji góðan

2. Þess má þó geta að GPT-4 virðist ekki hafa full tök á því að mynda *hv*-spurningar. Þegar við báðum um þýðingu úr ensku á íslensku á setningunum *Who did Mary see?*, *Who did Mary meet?*, *Who did Mary fire?*, *Who did Mary hire?* og *Which vase did Mary break?* var fall andlagsins rétt í síðustu þremur spurningunum (*Hvern rak/réð María?* og *Hvaða vasa braut María?*) en í þýðingunni á fyrstu tveimur spurningunum var *hv*-orðið í nefnifalli (*Hver sá/hitti María?*).

3. Katzir (2023) sýnir dæmið *The person that Mary met yesterday and that John will talk to Ed about his new idea tomorrow arrived*. Þessi setning er málfræðilega ótæk. Hér eru tvær samtengdar tilvísunarsetningar. Eyjahamlan sem um ræðir segir að þegar tvær setningar eru samtengdar þar sem önnur innihaldi eyðu verði að vera samsvarandi eyða í hinni setningunni líka. Eyðan í fyrri tilvísunarsetningunni (*that Mary met \_\_ yesterday*) er túlkuð sem *the person* úr móðursetningunni en það er engin samsvarandi eyða í samtengdu tilvísunarsetningunni sem skýrir hvers vegna enskumælandi fólki þykir dæmið ótækt.

árangur stórra mállíkana rothögg fyrir kenningar Chomskys segir hann ýmislegt vera sameiginlegt með þessum mállíkönum og sumum kenningakerfum innan málvísinda, svo sem þeim sem leggja áherslu á mynstur eða konstrúksjónir – mynstursmálfræði (e. *Construction Grammar*) er dæmi um slíkt kenningakerfi (sjá t.d. Goldberg 1995, 2006, 2019; sjá einnig Jóhönnu Barðdal, Þórhall Eypórsson og Dewey 2019 sem dæmi um hvernig mynstursmálfræði hefur verið beitt í rannsóknum á íslensku).

Grein Piantadosis hefur vakið mikla athygli og viðbrögð, bæði jákvæð og neikvæð. Chomsky lætur sér fátt um málgetu ChatGPT finnast og telur ekki að hún segi okkur nokkuð um málhæfileika mannsins (sjá t.d. Chomsky, Roberts og Watumull 2023). Það má e.t.v. setja spurningarmerki, líkt og Katzir (2023) gerir í svari við grein Piantadosis, við það að stór mállíkon séu álitin geta sagt okkur mikið um mannlega hæfni. Viðhorf eins og það sem birtist í grein Piantadosis kallar engu að síður á rannsóknir þar sem borin er saman málgeta mállíkana og mál barna (og raunar einnig fullorðinna).

Úr því að velgengi stórra mállíkana er notuð til að leggja mat á kenningar um hæfni mannsins til að læra og tileinka sér tungumál er mikilvægt að rannsaka fleiri mál en ensku í þessu samhengi. Mikilvægur vinkill í þeirri viðleitni að bera saman mál fólks og tölva í fjölda tungumála er ílagið, eða þær mállegu upplýsingar sem bæði börn og tölvur hafa aðgang að til að læra mannlegt mál. Í samanburðinum viljum við væntanlega að ílag mállíkana og barna sé sem líkast, svo sem að fjöldi orða sé svipaður (sbr. samkeppnina BabyLM sem vikið er að síðar), auk þess sem áhugavert er að bera saman árangur mállíkana eftir því hversu mikið af gögnum liggur að baki þekkingunni á ólíkum tungumálum. Til að mynda er ljóst að orðafjöldinn sem liggur til grundvallar þekkingu GPT-4 á íslensku og ensku er alls ekki sambærilegur, svo sem vegna þess að mun minna er til af íslensku efni, m.a. á netinu. Gerir þetta það að verkum að virkni GPT-4 í íslensku er þeim mun verri en virknin í ensku? Eða nægir magnið af íslensku efni (ílagi) sem er til staðar? Auk þessa má spyrja hvort smærri tungumál eins og íslenska séu ekki

einmitt betur til þess fallin að bera saman færni tölva og barna til þess að læra mál, þar sem ílagsmagnnið sem mállíkanið hefur úr að móða er sambærilegra ílagsmagninu sem nauðsynlegt er börnum á máltökuskeiði. Warstadt og Bowman (2023) skoða fjölda orðtöka í ýmsum líkönum í samanburði við ílag barna. Þar áætla þeir að heildarfjöldinn við 10 ára aldur sé um 100 milljónir en aftur á móti sé fjöldinn 200 milljarðar í GPT-3, fyrirrennara GPT-4. Það er ekki smávægilegur munur.

Við leyfum okkur nú að varpa fram þeirri spurningu hvort málkerfi tölva (eða mállíkana, sem sagt) sem hafa t.d. séð 100 milljónir íslenskra orða sé svipað eða jafnvel sams konar og málkerfi barna sem hafa heyrt sama fjölda orða.

### 3.2 Málkerfi barna og tölva

Þegar því er velt fram, eins og í grein Piantadosis, að stór mállíkon geti verið trúverðugar kenningar um eðli mannlegs máls, liggur beint við að bera saman hvernig mállíkon og börn tileinka sér mannlegt mál. Það er auðvitað mjög vandmeðfarið að ætla sér að bera saman málkerfi manna og tölva. Við verðum að gefa okkur margvíslegar forsendur sem kunna sumar hverjar að reynast hæpnar, svo sem að mállíkon búi yfir málkerfi yfirhöfuð, en of langt mál er hins vegar að útlista þær í svo stuttri grein. Ýmislegt orkar tvímælis, svo sem að bera saman ílag mállíkana, sem samanstandur gjarnan af fullmótuðum textum „á blaði“, og ílagið í máltöku barna sem er oft ófullmótað tal og ófullkomnar segðir (raunveruleg dæmi um það, þar sem faðir talar við börnin sín, eru: *hann var í í í hérna bænum; jæja, núna ætlar, núna skulum við, Sunneva, núna skulum við hlusta á Svövu*). Ekkert barn lærir mál með því að greina hundruð milljóna ritaðra orða líkt og stóru mállíkonin gera.

En getur verið áhugavert fyrir málvísindafólk sem sérhæfir sig í máltöku barna að skoða máltöku tölva? Þessu hefur verið líkt við þá spurningu hvort hægt sé að læra nokkuð um flug fugla út frá því hvernig flugvélar takast á loft, og svarið virðist vera á þá leið að slíkt geti verið athyglisvert (sjá t.d. Linzen 2023, Yang og Piantadosi 2022, Kodner, Caplan og Yang 2022). Þegar þetta er skrifað er beðið niðurstaðna úr



samkeppni þar sem máltækniþólk þjálfar mállíkan þar sem þjálfunargögnin samanstanda af 10–100 milljónum orða á ensku og innihalda mál sem beinist að börnum, texta úr barnabókum o.s.frv. (sjá lýsingu samkeppninnar, sem á ensku er kölluð BabyLM Challenge, hjá Warstadt o.fl. 2023). Mjög áhugavert verður að athuga hvernig til tekst og bera niðurstöðurnar saman við mál og máltöku barna.

Mjög margt er vitað um það ferli þegar börn taka mál enda mikið rannsakað svið innan málvísinda. Jafnvel þótt ílagið, málið sem börn heyra, sé takmarkað gegnir það lykilhlutverki þegar börn læra móðurmál sitt. Magn og gæði ílags skiptir líka höfuðmáli fyrir málgetu mállíkana. Þess vegna er mjög mikilvægt að við höfum a.m.k. upplýsingar um textana sem mállíkönin hafa aðgang að. Upplýsingar um t.d. hvaða íslensku textar eru á bak við GPT-4 og hver orðaföldi íslenskra texta í líkaninu er eru óaðgengilegar og það heftir mjög rannsóknir á málgetu mállíkansins.

Ein leið við rannsóknir á barnamáli er þá að gaumgæfa það sem börn segja og skoða frávík frá máli fullorðinna. Ef barn segir *Ég dettaði* (í stað *Ég datt*) drögum við e.t.v. þá ályktun að það alhæfi í málkerfi sínu veika beygingu sagna, sem er algengari en sterk beyging, en búumst þá jafnframt við að það tileinki sér síðar sterka beygingu þessarar sagnar. Eftir því sem við best vitum notar langflest eða allt fullorðið fólk sem hefur íslensku að móðurmáli *detta* með sterkri beygingu og því eru það skýr frávík frá máli þess þegar barn segir *dettaði*.

Það segir kannski ekkert sérlega mikið um málhæfni mannsins ef mállíkan sem hefur verið matað á fleiri milljörðum orða nær tókum á óreglulegri beygingu sagna og óreglulegri fleirtöluendingu nafnorða í ensku. Aftur á móti er áhugavert að beita prófum sem hafa verið notuð í barnamálsrannsóknum á mállíkön. Weissweiler o.fl. (2023) gerðu það einmitt nýlega en þar athuguðu þau hvernig GPT-3.5 stendur sig á bullorðaprófi eins og því sem Berko (1958) hannaði og er fræg fyrir. Slík próf eru oft kölluð *wug*-próf eftir einum þætti prófsins en þá var sýnd mynd af veru sem börnin höfðu aldrei heyrt um áður en líkist einna helst fugli;

á ensku kallaði Berko hana *wug*. Barninu var því sagt að þetta væri *wug*. Svo var barninu sýnd mynd af einni slíkri veru til viðbótar og það fengið til að mynda fleirtölu orðsins:

1. Now there is another one. There are two of them. There are two \_\_\_\_.

Berko kannaði ekki eingöngu fleirtölu orða heldur einnig t.d. þátíð bullsagna, eins og *bod* en þá sáu börnin mynd af manni sveifla hlut sem var festur við band:

2. This is a man who knows how to bod. He is bodding. He did the same thing yesterday. Yesterday he \_\_\_\_.

Hér er lykilatriði að börnin þekkja ekki orðin sem unnið er með. Þannig er hægt að átta sig á málkerfinu sem börn hafa komið sér upp og þau beita á orð sem þau hafa aldrei heyrt áður. Weissweiler o.fl. könnuðu getu GPT-3.5 í ensku (og einnig tamíl, tyrknesku og þýsku) í bullorðaprófi sem þau byggðu á prófi Berko. Niðurstaða þeirra er að GPT-3.5 standi sig ekkert sérlega vel í slíku *wug*-prófi; þau nefna sem dæmi að líkanið hafi gefið þátíðina *did*, *blushed* og *fried* (sem sagt þátíð annarra þekktra sagna) af bullsögnunum *dedo*, *blus* og *fride*. Þess skal þó getið að í niðurstöðum sínum á samanburði á mállíkaninu og mannlegri hæfni miðuðu Weissweiler o.fl. ekki við getu barna heldur fullorðinna. Þetta þarfnast frekari rannsókna og spennandi væri að skoða þetta m.t.t. íslensku.

Staðhæfingar um að málgeta stórra mállíkana geti sagt okkur sitthvað um málhæfni mannsins eru að mestu leyti byggðar á getu þeirra í ensku. Þær vekja strax upp áhugaverðar spurningar um hvernig líkönin standa sig í málum sem þau eru ekki sérþjálfuð á og þar sem færri orð liggja til grundvallar. Það er ekki ofsögum sagt að GPT-4 (og þar áður GPT-3.5 og GPT-3) hafi verið mikið í umræðunni undanfarið. Hérlandis hefur verið rætt mikið um íslenskukunnáttu GPT-4 og þá hefur verið bent á að hún virðist talsvert meiri en hjá fyrirrennaranum GPT-3.5. Þrátt fyrir að gögnin sem liggja að baki, eða upplýsingar um þau, séu ekki aðgengileg getur verið mjög áhugavert að skoða getu GPT-4 í íslensku betur.

Þar getum við skoðað frálagið (sem við getum sagt, samanburðarins vegna, að jafngildi málbeitingu hjá mönnum). Við getum einnig sett fram spurningar í netviðmótinu ChatGPT (þar er hægt að velja um GPT-3.5 og GPT-4) um hvort tiltekna setningar séu eðlilegt íslenskt mál og reynt þannig að fá aðgang að því sem við getum kallað málkunnáttu þess. Þetta má svo setja í samhengi við máltöku barna og athuga t.d. hvort frávik frá máli fullorðinna séu svipuð hjá GPT-4 og börnum.

### 3.3 ChatGPT (GPT-4) og íslenskan

Við höfum á undanförnum mánuðum reynt að átta okkur á íslenskugetu GPT-4 með því að senda margvíslegar fyrirspurnir í viðmótinu ChatGPT. Hér greinum við aðeins frá tveimur atriðum: annars vegar skilningi á spurningum þar sem aukafall var notað og hins vegar málfræðilegu samræmi innan nafnliða. Þetta gerum við til að draga fram mögulegan mun á málkunnáttu barna og málgetu mállíkana en þetta þarfnast mun meiri rannsókna. Við hefðum getað bent á fjölmörg dæmi um það þar sem það kom okkur á óvart hve mikil málgetan virtist vera en einnig ýmis dæmi sem við teljum líklegt að tíu ára börn (hvað þá fullorðnir) myndu ekki gera.

Við spurðum mállíkanið um *Ofsa* og áttum þar við skáldsögu Einars Kárasonar (EFS eru upphafsstafrir annars höfunda þessarar greinar og ChatGPT er spjallmennnið sem er byggt á GPT-4).

#### 3. EFS: Hvað geturðu sagt mér um skáldsöguna Ofsa?

ChatGPT: Ég þekki enga skáldsögu með nafnið Ofsa. Hins vegar er mögulegt að þú sért að leita að skáldsögunni "Ofsi" eftir íslenska rithöfundinn Stein Steinarr. Skáldsagan Ofsi var gefin út árið 1958 og er þekkt fyrir flókinn stíl, táknafræði og tilraunir með málið.

Ofsi fjallar um einstaklinginn Ljótu, sem þjáist af tilveru sinni, og baráttu hans við umheiminn, eigin sál og þau mælskuleg örlög sem hann þarf að takast á við. Sagan fer fram á Íslandi á síðari hluta 20. aldar og endurspeglar innri togstreitu og einhvers konar uppgjöf við samfélagið.

Skáldsagan hefur hlotið margvíslegar túlkunir og er talin eitt helsta verk Steinars. Hún er oft rakin til póstmodernisma og dadaískrar listahreyfingar og er góð dæmi um hvernig íslensk skáldskapur þróast og breytist á miðri 20. öld.

Þetta er athyglisvert svar að mörgu leyti þótt hér verði einblínt á málfræðilega þætti, en benda má að svarið er gott dæmi um þau sannfærandi ósannindi og upplýsingaóreiðu sem spjallmennni geta skapað. Í svari ChatGPT er byrjað á því að taka titilinn eins og hann er skrifaður. Þetta kemur ekki fullkomlega á óvart enda eru titlar oft í nefnifalli, sbr. *Hefurðu lesið (bókina) Stríð og friður?* Hins vegar gerðist þetta líka þegar við spurðum um leikarana Sigga Sigurjóns og Ladda en ekki þegar við spurðum um Eddu Björgvins:

#### 4. EFS: Þekkirðu leikarann Sigga Sigurjóns? Þekkirðu leikarann Ladda? Þekkirðu leikkonuna Eddu Björgvinsdóttur?

ChatGPT: Já, ég þekki þessa þjóðfræga leikara og leikkonu frá Íslandi:

1. Sigga Sigurjóns (Sigurjón Sighvatsson) er íslenskur leikari [...]
2. Ladda er líka leikari, en nafnið er ekki nógu skýrt til að segja meira um hann. [...]
3. Edda Björgvinsdóttir er einnig þekkt íslensk leikkona, þekkt fyrir hluti sinn í sjónvarpsþáttaröðum, kvikmyndum og leikhúsi á Íslandi. [...]

Það að ChatGPT túlki *Sigga Sigurjóns* og *Ladda* sem nefnifall (þegar þolfallið *leikarann* fer á undan) eru skýr frávik frá því sem við byggjumst við að heyra fullorðið fólk segja, jafnvel þótt það hefði ekki heyrt um þessa leikara (að því gefnu að það áttaði sig á því að þeir væru íslenskir eins og ChatGPT virðist gera).

Það væri spennandi að bera þetta saman við svör barna og við höfum raunar gert örlitla tilraun með það í huga. Við spurðum nefnilega einnig hvort ChatGPT hefði lesið *Ofsa*.

## 5. EFS: Hefurðu lesið Ofsa?

ChatGPT: "Ofsa" er skáldsaga eftir íslenska höfundinn Arnaldur Indriðason, sem kom út árið 2004. Bókin er hluti af röð skáldsagna sem fjalla um lögreglumanninn Erlendur Sveinsson. "Ofsa" skartar spennuþræði, þar sem Erlendur, Elínborg og Sigurður Óli reyna að leysa mál um morð sem gerðu ráð fyrir óvenjulegum og flóknum aðferðum.

Ég hef ekki "lesið" bókina á þann hátt sem manneskja myndi lesa, en ég hef aðgang að upplýsingum um bókina og get svarað spurningum um hana. Hafðu það í huga að ég er gervimælgí forrit og skilningur minn á bókum byggir á upplýsingum sem ég hef verið þjálfuð með, en ég skynja ekki skáldskap, menningu eða takt á sama hátt og manneskja myndi gera.

Aftur túlkar ChatGPT titilinn sem *Ofsa* en ekki *Ofsi*. Við spurðum rúmlega sjö ára barn (sem við köllum hér A) þessarar sömu spurningar.

## 6. EFS: A, hefur þú lesið Ofsa?

A: Hvað er það?

EFS: Hvað er hvað?

A: Hvað er þessi Ofsi?

Eins og sést á þessu samtali hefur barnið aldrei heyrt um *Ofsa* (og ekki er ljóst út frá þessu að það geri sér grein fyrir að þetta er titill) en það vefst ekki fyrir því að túlka *Ofsa* í spurningunni sem þolfall, enda segir barnið *Hvað er þessi Ofsi?*

Þegar barnamál er rannsakað er mjög oft lögð áhersla á það sem börn segja en einnig eru börn stundum beðin um að dæma setningar (þó getur þurft að laða slíka dóma fram á annan hátt en í hefðbundnum dómaprófum sem eru lögð fyrir fullorðna). Hvort tveggja getur verið mjög gagnlegt. Við höfum ekki eingöngu velt fyrir okkur málinu í svörum GPT-4 heldur biðjum við einnig mjög oft um dóma á mismunandi setningum. Það sem hér fer á eftir miðast bæði við það sem ChatGPT (GPT-4) hefur látið frá sér og dómapróf þar sem við spyrjum um tiltekna setningar.

Það er auðvitað ekki alltaf ljóst hvort sá skilningur sem við leggjum í að setning sé t.d. málfræðilega eðlileg sé sá sami þegar GPT-4 vinnur úr fyrirspurnum okkar. Við getum þó skoðað eftirfarandi dæmi úr *Setningum* (Höskuldur Þráinsson 2005:4, 10, 90, 552) en segja má að þau séu annaðhvort ótvírætt ótæk eða tæk (stjörnumerktar setningar eru ótækar en tækar setningar hafa enga merkingu).

7. Ég reyni það ekki.

8. \*Ég reyni þess ekki.

9. María barði Harald.

10. \*María Harald barði.

11. Átt þú enga bók eftir Laxness?

12. \*Átt þú neina bók eftir Laxness?

13. Kennarinn hafði borðað kritina.

14. \*Kennarinn hafði borðað krítinni.

Hér eru setningarnar settar upp sem pör þar sem fyrri setningin er tæk en sú síðari ótæk. Við spurðum ChatGPT um þessar setningar en settum þær ekki saman í pörum. Fyrst spurðum við um setningar (7), (10), (12) og (13) (í þessari röð) og síðan (14), (9), (11) og (8).

Við búumst ekki við miklum tilbrigðum í máli fólks varðandi þessar setningar – dómar Höskuldar eru í það minnsta í samræmi við máltilfinningu okkar. Í svörum sínum við spurningum okkar (*Geturðu sagt mér hvort eftirfarandi setningar séu málfræðilega eðlilegar í íslensku?*) talar ChatGPT um að tiltekin setning geti verið „málfræðilega rétt“, „málfræðilega órétt“ eða „ófullkomin“. ChatGPT telur allar tæku setningarnar (7), (9), (11) og (13) vera málfræðilega réttar og telur þrjár ótækar setningar af fjórum ekki málfræðilega réttar, (8), (10) og (14), en segir setninguna *Átt þú neina bók eftir Laxness?* málfræðilega rétta, öfugt við það sem við teljum.

Þetta litla sýnishorn gefur innsýn í það þegar ChatGPT er beðið um að dæma setningar. Þetta er vitaskuld ekki fullkomin leið; við vitum ekki almennilega við hvað ChatGPT miðar í svörum sínum (við vitum t.d. ekki hvort setningar sem spjallmennið telur tækar séu í samræmi við málkerfi

líkansins eða hvort þær séu í samræmi við þær vísbendingar sem felast í ílaginu) en við höfum einnig beðið um dóma á ýmsum setningum með öðru orðalagi og einnig spurt á ensku (t.d. *Are the following sentences grammatical in Icelandic?*).

Við snúum okkur núna að samræmi, sem er mjög ríkulegt í íslensku, alveg sérstaklega innan nafnliða. Þegar karlkynsnafnorð stendur í þolfalli eru ákvæðisorð þess innan nafnliðarins einnig í karlkyni og þolfalli. Þannig segjum við t.d. *Ég las skemmtilegan kafla* (en ekki *\*Ég las skemmtilegur kafla*). Þess vegna kemur það svolítið á óvart að sjá af og til í máli GPT-4 að ekki sé notað samræmi innan eins og sama nafnliðarins. Dæmi um slíkt eru *vegna kjarasamningsbundinna launahækkanir* (nafnorð í nefnifalli eða þolfalli en ákvæðisorð í eignarfalli) og *íslensk skáldskapur* (ákvæðisorð í kvenkyni sem stendur með karlkynsnafnorði). Því datt okkur í hug að kanna þetta aðeins betur:

15. EFS: Geturðu sagt mér hvort eftirfarandi setning sé eðlilegt mál?

María er góð kennari.

ChatGPT: Já, þessi setning er eðlileg og einföld á íslensku.

„María er góð kennari.“

Setningin gefur einfaldlega upplýsinguna að María sé dugleg eða áreiðanleg í kennslustörfum sínum.

Við spurðum jafnframt um setninguna *María er góður kennari* sem ChatGPT sagði ekki vera eðlilega. Við báðum um fallbeygingu orðsins *kennari* og þótt hún væri ekki alveg fullkomin var ljóst að þetta var karlkynsorð í meðförum ChatGPT (og svo sagði spjallmennið einnig að þetta væri karlkynsorð þegar spurt var sérstaklega út í kynið á *kennari*). Við vildum ekki eingöngu skoða málfræðilega dóma mállíkansins svo að við báðum það einnig – á ensku – að skrifa sögu um Mary sem væri talin „good teacher“ til þess að laða fram *góð/góður kennari* í málnotkun.

16. EFS: Could you write a story in Icelandic about Mary who is considered to be a good teacher? Later she becomes a doctor and is also considered to be a good doctor.

ChatGPT: Certainly! Here's a short story in Icelandic:

#### Mary: Frá Kennara til Læknis

Einu sinni var kona sem hét Mary. Hún var kennari í grunnskóla og ávallt mjög vinsæl meðal nemendanna og foreldra þeirra. Mary var þekkt fyrir að vera góð kennari. [...]

17. EFS: Could you write a story in Icelandic about Mary who is considered to be a good teacher?

ChatGPT: María var talin vera góður kennari. [...]

Eins og sést í (16) og (17) höfum við bæði fengið sögur frá ChatGPT þar sem María (eða Mary) er sögð vera „góð kennari“ og þar sem hún er „góður kennari“.

Í nafnliðnum *góð kennari* felast skýr frávik frá máli fullorðinna jafnvel þótt við segjum *María er góð* og jafnvel þótt við gætum sagt um kvenkyns kennara: *Nýi kennarinn er góð* (sbr. umræðu hjá t.d. Önnu Helgadóttur 2011, Guðrúnu Þórhallsdóttur 2015, Einari Frey Sigurðssyni 2017 og Einari Frey Sigurðssyni o.fl. 2022). Dæmið er sérstaklega áhugavert vegna þess að það dregur mögulega fram grundvallarmun á málkerfi tölva og manna þar sem tölvur miða meira við línulega yfirborðsröð orða til þess að ákverða tengsl þeirra á milli á meðan þekkt er að mannlegt mál einkennist fyrst og fremst af stigveldisskipulagi (sjá einnig t.d. Þórhall Eyþórsson 2023).

Þessi sýnishorn um íslenskuvirkni ChatGPT eru áhugaverð en erfitt er draga ályktanir um mun á máli og máltöku barna og tölva út frá þessu þar sem við vitum ekki hvaða íslenska leynist í þjálfunargögnunum. Hugsanlegt er að þar sé frekar lítil íslenska og þar af jafnvel slatti af mjög illa vélþýddum textum.



#### 4. Lokaorð

Út frá ofangreindri umfjöllun er ljóst að fyrst og fremst er aragrúa spurninga enn ósvarað um annars vegar getu myndandi gervigreindar og stórra mállíkana og hins vegar gildi þeirrar getu, bæði fyrir samfélagið sem heild og fyrir kenningar um eðli mannlegs máls. Alþjóðleg umræða um stór mállíkön er lituð af enskuskekkju og í því samhengi er mikilvægt að afla upplýsinga um virkni stórra mállíkana í smærri tungumálum. Með slíkum rannsóknum skýrist hvers eðlis kapphlaupið við ensku er, svo sem hvort stór mállíkön verði nothæf í íslensku málsamfélagi, en einnig skapast mikilvæg þekking á virkni og eðli slíkra mállíkana almennt. Sú þekking skiptir sköpum

í umræðunni um mögulega nýtingu á gervigreind í samfélaginu öllu auk þess sem hún er orðin grundvallaratriði í málvísindakenningum. Áhugavert verður að fylgjast með þróun og framgangi mállíkana sem eru sérþjálfuð á íslensku og þar sem upplýsingar um þjálfunargögnin liggja fyrir.

Umfjöllun okkar gefur aðeins takmarkaða innsýn í það hvers eðlis málvísindalegar rannsóknir á virkni stórra mállíkana í íslensku gætu verið auk þess sem komið var inn á mögulegt mikilvægi þeirra. Okkur þykir þó ljóst, áður en lengra er haldið í slíkum rannsóknum, að samfélagsleg umræða þurfi að eiga sér stað um ýmis siðferðisleg álitamál í notkun stórra mállíkana og myndandi gervigreindar.

#### Heimildaskrá

- Anna Helgadóttir. 2011. *Notkun málfræðilegra kynja í máli ungs fólks. Rannsókn á kynjanotkun í íslensku máli*. MA-ritgerð, Háskóla Íslands, Reykjavík.
- Atari, Mohammad, Mona J. Xue, Peter S. Park, Damián E. Blasi og Joseph Henrich. 2023. Which humans? Handrit, <https://doi.org/10.31234/osf.io/5b26t>.
- Bender, Emily M., og Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics* 6:587–604.
- Bender, Emily M., og Alex Hanna. 2023. AI causes real harm. Let's focus on that over the end-of-humanity hype. *Scientific American*, 12. ágúst. <https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/>.
- Bender, Emily M., og Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, bls. 5185–5198. Association for Computational Linguistics, Online. <https://aclanthology.org/2020.acl-main.463>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major og Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? FAccT'21: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, bls. 610–623. Association for Computing Machinery, New York, NY.
- Berko, Jean. 1958. The child's learning of English morphology. *WORD* 14:150–177.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris, Dordrecht.
- Chomsky, Noam, Ian Roberts og Jeffrey Watumull. 2023. The false promise of ChatGPT. *New York Times*, 8. mars. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.
- Einar Freyr Sigurðsson. 2017. *Deriving Case, Agreement and Voice Phenomena in Syntax*. Doktorsritgerð, University of Pennsylvania.
- Einar Freyr Sigurðsson, Finnur Ágúst Ingimundarson og Matthew Whelpton. 2022. Samræmi við hulin nafnorð. *Orð og tunga* 24:21–56.
- Eiríkur Rögnvaldsson. 2016. Um utanaðkomandi aðstæður íslenskrar málþróunar. *Skírnir* 190:17–31.
- Eiríkur Rögnvaldsson, Kristín M. Jóhannsdóttir, Sigrún Helgadóttir og Steinþór Steingrímsson. 2012. *Íslensk tunga á stafrænni öld / The Icelandic Language in the Digital Age*. META-NET White Paper Series, Springer, Berlín.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. The University of Chicago Press, Chicago, IL.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, Oxford.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton University Press, Princeton, NJ.
- Guðrún Þórhallsdóttir. 2015. Gender agreement in 19th- and 20th-century Icelandic. Jürg Fleischer, Elisabeth Rieken og Paul Widmar (ritstj.): *Agreement from a Diachronic Perspective*, bls. 267–286. Mouton de Gruyter, Berlín.
- Höskuldur Þráinsson. 2005. *Setningar. Handbók um setningafræði*. Íslensk tunga III. Meðhöfundar: Eiríkur Rögnvaldsson, Jóhannes Gísli Jónsson, Sigríður Magnúsdóttir, Sigríður Sigurjónsdóttir og Þórunn Blöndal. Almenna bókafélagið, Reykjavík.

- Iris Edda Nowenstein og Sigríður Sigurjónsdóttir. 2021. Stafrænt málsambýli íslensku og ensku. Áhrif ensks ílags og málnotkunar á málfærni íslenskra barna. *Ritið* 21(3):11–56.
- Jóhanna Barðdal, Þórhallur Eyþórsson og Tonya Kim Dewey. 2019. The alternating predicate puzzle: DAT-NOM vs. NOM-DAT in Icelandic and German. *Constructions and Frames* 11:107–170.
- Katzir, Roni. 2023. Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi (2023). Handrit, <https://lingbuzz.net/lingbuzz/007190>.
- Kidd, Celeste, og Abeba Birhane. 2023. How AI can distort human beliefs: Models can convey biases and false information to users. *Science* 380(6651):1222–1223. <https://doi.org/10.1126/science.adi0248>.
- Kodner, Jordan, Spencer Caplan og Charles Yang. 2022. Another model not for the learning of language. *PNAS* 119(29):e2204664119. <https://doi.org/10.1073/pnas.2204664119>.
- Linzen, Tal. 2023. What, if anything, can large language models teach us about human language acquisition? Handrit.
- Lucchi, Nicola. 2023. ChatGPT: A case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*. <https://doi.org/10.1017/err.2023.59>.
- Marcus, Gary. 2022. A few words about bullshit. How MetaAI's Galactica just jumped the shark. *Marcus on AI*, 16. nóvember. <https://garymarcus.substack.com/p/a-few-words-about-bullshit>.
- Marcus, Gary. 2023. Inside the heart of ChatGPT's darkness. Nightmare on LLM Street. *Marcus on AI*, 11. febrúar. <https://garymarcus.substack.com/p/inside-the-heart-of-chatgpts-darkness>.
- Miðeind. 2023. Miðeind í samstarfi við OpenAI um GPT-4. 14. mars. <https://mideind.is/frettir.html>.
- OpenAI. 2023. How Iceland is using GPT-4 to preserve its language. 14. mars. <https://openai.com/customer-stories/government-of-iceland>.
- Perrigo, Billy. 2023. Exclusive: OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic. *TIME*, 18. janúar. <https://time.com/6247678/openai-chatgpt-kenya-workers/>.
- Piantadosi, Steven T. 2023. Modern language models refute Chomsky's approach to language. Handrit, <https://lingbuzz.net/lingbuzz/007180>.
- Ross, John Robert. 1967. *Constraints on Variables in Syntax*. Doktorsritgerð, Massachusetts Institute of Technology.
- RÚV.is = Ísak Gabriel Regal. 2023. Guðfaðir gervigreindar hættir hjá Google og varar við framþróun tækninnar. *RÚV.is*, 2. maí. <https://www.ruv.is/frettir/erlent/2023-05-02-gudfadir-gervigreindar-haettir-hja-google-og-varar-vid-framthroun-taekninnar>.
- Samtök atvinnulífsins. 2018. Almennarómur bjargar íslenskunni frá stafrænum dauða. 27. ágúst. <https://www.sa.is/frettatengt/frettir/almannaromur-bjargar-islenskunni-fra-stafrænum-dauða>.
- Sigríður Sigurjónsdóttir. 2013. Máltaka barna og meðfæddur málhæfileiki. Höskuldur Þráinsson og Matthew Whelpton (ritstj.): *Chomsky: Mál, sál og samfélag*, bls. 107–127. Hugvísindastofnun og Háskólaútgáfan, Reykjavík.
- Warstadt, Alex, og Samuel R. Bowman. 2023. What artificial neural networks can tell us about human language acquisition. Shalom Lappin og Jean-Philippe Bernardy (ritstj.): *Algebraic Structures in Natural Language*, bls. 17–59. CRC Press, Boca Raton, FL.
- Warstadt, Alex, Leshem Choshen, Aaron Mueller, Ethan Wilcox, Adina Williams og Chengxu Zhuang. 2023. Call for papers – The BabyLM Challenge: Sample-efficient pretraining on a developmentally plausible corpus. <https://doi.org/10.48550/arXiv.2301.11796>.
- Weil, Elizabeth. 2023. You are not a parrot. *New York*, 1. mars. <https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html>.
- Weissweiler, Leonie, Valentin Hofmann, Anjali Kantharuban, Anna Cai, Ritam Dutt, Amey Hengle, Anubha Kabra, Atharva Kulkarni, Abhishek Vijayakumar, Haofoei Yu, Hinrich Schütze, Kemal Oflazer og David R. Mortensen. 2023. Counting the bugs in ChatGPT's wugs: A multilingual investigation into the morphological capabilities of a large language model. Handrit, <https://doi.org/10.48550/arXiv.2310.15113>.
- Yang, Yuan, og Steven T. Piantadosi. 2022. One model for the learning of language. *PNAS* 119(5):e2021865119. <https://doi.org/10.1073/pnas.2021865119>.
- Þórhallur Eyþórsson. 2013. Málmyndun, málkunnátta og algildismálfræði. Höskuldur Þráinsson og Matthew Whelpton (ritstj.): *Chomsky: Mál, sál og samfélag*, bls. 149–171. Hugvísindastofnun og Háskólaútgáfan, Reykjavík.
- Þórhallur Eyþórsson. 2023. Chomsky og spjallmennnið. *Morgunblaðið*, 25. mars, bls. 28.

