

## „Góð bílstjóri“ og „bestunárangur“: Það sem gervigreindin lærir og lærir ekki í íslensku

Málhæfni stórra mállíkana á borð við GPT, Claude, DeepSeek o.fl. (oft nefnd *gervigreind(in)*) í daglegu tali, sbr. Ágústu Þorbergsdóttur o.fl. 2024) hefur verið mikið til umræðu síðastliðin misseri enda hafa þau náð fordæmalausum árangri í hinum ýmsu skriflegu verkefnum sem fyrir þau hafa verið lögð. Þau virka langbest á ensku en eru brúkleg á fjölmörgum öðrum málum og íslenska er þar ekki undanskilin. Líkönin eru ólík þeim tólum sem máltæknifræðingar notuðu til textavinnslu og spurningasvörunar fyrir ekki svo mörgum árum, að því leytinu til að þau byggja á tauganetum sem læra að greina mynstur í þeim texta sem þau eru þjálfuð á út frá textabrotum sem yfirleitt eru ekki greind, flokkuð eða mörkuð á neinn hátt.

Það er því e.t.v. ekki að furða að ófáir málfræðingar telji þá „málkunnáttu“ sem verður til við útreikninga þessara líkana mikilvægasta rannsóknarefnið innan fræðasviðs gervigreindar um þessar mundir (sjá t.d. yfirlitsgrein Linzen og Baroni 2021). Þannig hefur árangur líkananna á síðustu árum orðið kveikja að fjölda rannsókna bæði málvísindamanna og máltæknifræðinga sem miðar að því að kanna hvaða innsýn líkönin geti veitt okkur í mannlegt mál.

### Tauganet og máltaka

Hér staldrar væntanlega hinn skynsami lesandi við og spyr sig að því hvernig [tölvulíkon sem forrituð eru til þess að spá fyrir um líkindi tóka í textarunu](#) eigi að geta sagt okkur nokkuð um það hvernig við sjálf lærum og notum málið. Er það ekki nokkurn veginn (svo gripið sé til myndmáls sem notað hefur verið í þessari umræðu) eins og að fylgjast með flugvélum til þess að reyna að rannsaka hvernig fuglar fljúga? Mögulega er hugmyndin þó ekki alveg jafnfjarstæðukennd og hún gæti virst í fyrstu. Hún er í það minnsta ekki ný af nálinni.

Máltaka barna hefur verið eitt af helstu viðfangsefnum málvísindamanna frá því um miðja síðustu öld og rannsóknir að miklu leyti beinst að þeirri spurningu hvað sé *mögulegt að læra* á máltökuskeiði og hvað ekki. Svo stiklað sé á gríðarlega stóru, hefur mjög mikið verið ritað og rætt um kenningar þess efnis að það hreinlega geti ekki verið að börn heyri og lesi nægilega mikið af móðurmáli sínu áður en máltökuskeiði lýkur til þess hver málhafi nái jafnsterkum tókum á málinu og raun ber vitni — þar af leiðandi hljóti þekking okkar á tungumálinu að vera að einhverju leyti meðfædd (Chomsky 1965).

Rúm þrjátíu ár eru liðin frá því að tauganet (forverar risamállíkana okkar daga) fór að bera á góma í þeirri umræðu sem andsvar við þessum kenningum um meðfædda málhæfni, á þeim forsendum að tauganet gæti fræðilega séð lært mál einungis út frá sýndum dæmum og án nokkurrar innbyggðrar vitneskju (ein gerð tauganeta var til að mynda upphaflega hönnuð af málfræðingnum Jeffrey Elman (1990) sem tilgáta um það hvernig við lærum málið). Fyrir þrjátíu árum var auðveldara að færa rök fyrir því að tauganet gæfu ekki rétta mynd af máltöku, á þeim forsendum að þau yrðu aldrei fær um að læra ákveðnar setningagerðir. Nú þegar mörg þeirra dæma sem tekin voru virðast leikur einn fyrir líkon á borð við ChatGPT er málið strax farið að flækjast.

Það er auðvitað engum blöðum um það að fletta að við mannfólkið lærum tungumál mun hraðar og betur en risamállíkon dagsins í dag, hvort sem um er að ræða ensku, íslensku eða önnur mál, og að máltaka barna og þjálfun mállíkana fer fram með gerólíkum hætti. Eftir situr þó að rannsóknir á því

hvaða málfræðimynstur virðast lærast vel og hvaða mynstur síður vel, með tilliti til ólíkra tungumála, ólíkrar dreifingar í þjálfunargögnum og ólíkrar uppbyggingar líkana, geta bætt enn frekar við þekkingu okkar á því hvernig við virðumst fara að því að læra ákveðna þætti málsins út frá vísbendingum í ílaginu.

## Orðmyndun og samræmi

En hvaða atriði í íslensku getum við kannað með tilliti til þessa? Geta rannsóknir á því hvernig risamállíkon læra íslenska málfræði sagt okkur eitthvað um annaðhvort líkönin eða íslensku, eða þá um mannlegt mál almennt? Rannsakendur hjá Árnastofnun og Háskóla Íslands (undirritaður, ásamt þeim Finni Ágústi Ingimundarsyni, Irisi Eddu Nowenstein og Einari Frey Sigurðssyni) hafa undanfarið unnið að ýmiss konar prófunum á íslenskuhæfni stórra mállíkana og þeim hefur að hluta verið fylgt eftir með samanburðarprófunum á íslenskum málhöfum. Þau próf sneru að þeim atriðum sem hér verða skoðuð nánar: Myndun samsettra orða með ákveðnum viðskeytum og kynjasamræmi innan nafnliðar.

Í orðmyndunarverkefninu skoðuðum við fjögur viðskeyti sem ekki leyfa stofnsamsetningu í íslensku; -*uður*, -*ingur*, -*ing* og -*un* (Guðrún Kvaran 2005:154). Það er, orð sem enda á þessum viðskeytum þurfa á eignarfallsendingu að halda þegar þau mynda fyrri hluta samsetts orðs: Íslenskir málhafar geta t.d. myndað orðin *skoðunarmaður* og *skoðanakönnun* en ekki *\*skoðunmaður* eða *\*skoðunkönnun*. Fyrir hvert viðskeyti, búum við til tíu möguleg orðapör til að þúsla saman, t.d. *bestun* + *árangur*, *elding* + *læti* o.s.frv., og reyndum að hafa þau eins eðlileg og hægt var. Ellefu stór mállíkon voru síðan beðin annars vegar um að mynda nýtt orð úr þessum hlutum og hins vegar að leggja mat á það hvort eignarfallssamsetningar eða stofnsamsetningar væru réttar eða rangar á íslensku (þ.e. fyrir parið *elding* + *læti* voru líkönin látin meta orðin *eldingalæti*, *eldingarlæti* og *eldinglæti*). 108 íslenskir málhafar tóku þetta sama próf að hluta og svör þeirra voru borin saman við svör líkananna.

Í kynjasamræmisverkefninu voru sömu líkön beðin um að leggja mat á 44 stuttar og einfaldar setningar á íslensku þar sem nöfn og nafnorð af ólíkum kynjum voru þöruð saman ásamt lýsingarorði, sem ýmist samræmdu nafnorðinu innan nafnliðarinnar sem það tilheyrði eða nafninu sem var frumlag setningarinnar. Líkönin voru þannig bæði mötuð á eðlilegum setningum á borð við *María er góður bílstjóri*, þar sem lýsingarorðið samræmist karlkynsorðinu *bílstjóri*, og ótækum setningum á borð við *\*María er góð bílstjóri*, þar sem lýsingarorðið samræmist kvenkynsnafninu *María*. Auk þess að leggja mat á setningarnar voru líkönin látin þýða 21 hliðstæða setningu úr ensku (t.d. *María is a good driver*), til þess að athuga hvaða kyn nafnorðsins kæmi út. Líkt og í orðmyndunarverkefninu var sama próf lagt fyrir mennska þátttakendur, alls 188 manns.

Í sem stystu máli — en nánar verður rýnt í niðurstöðurnar á öðrum vettvangi — voru svörin frá mennskum málhöfum afdráttarlaus. Í 99 prósentum tilvika notaði fólk kynjasamræmi innan nafnliðar (*góður bílstjóri*), bæði í dómum sínum og í þýðingum úr ensku, á meðan þau líkön sem komust næst mennskri frammistöðu voru með rétt rúmlega helming dóma réttan og ekkert líkan þýddi meira en 15 af 21 setningu rétt með tilliti til þessa. Í orðmyndunarhlutanum fannst ekki eitt einasta dæmi um stofnsamsetningu í mennsku svörunum og yfirleitt voru 80-90 prósent svarenda sammála um eina eða tvær orðmyndir, sem alltaf fólu í sér eignarfallssamsetningu. Í úttaki líkananna var þessu öfugt farið, aðeins fjögur líkön notuðu eignarfallssamsetningar í meira en helmingi tilfella (og þá bara rétt meira en helmingi) og jafnvel bestu líkönin buðu upp á ótækar orðmyndir á borð við *\*bestunárangur*, *\*eldinglæti*, *\*snyrtingstóll*, o.s.frv.

## Hvað segir þetta okkur?

Nú kann að vera að einhverjum finnist við vera að einblína á fullmikil smáatriði í íslenskri málfræðikunnáttu í annars oft þokkalega sannfærandi orðaflaumi þeirra mállíkana sem best standa sig í íslensku. Skiptir máli þó að líkan myndi ekki alltaf orð nákvæmlega eins og við myndum búast við að íslenskur málhafi gerði það? Já og nei: Eftir situr að bestu líkönin sem standa til boða eru til margs nýt en það sem við höfum áhuga á að skoða er hæfni þeirra til að koma sér upp mynstrum og reglum út frá þjálfunargögnum — og það samanborið við mannfólk. Það að átta sig á takmörkum stórra mállíkana (sem kunna að vera mismunandi eftir tungumálum) er enda eitt af því sem mestu máli skiptir upp á ábyrga notkun þeirra að gera.

Erfitt er að áætla hversu mörg orð barn heyrir á máltökuskeiði til þess að koma sér upp fullmótuðu málkerfi en oft er miðað við að það sé í kringum hundrað milljónir orða (Gilkerson o.fl. 2013). Til samanburðar var síðasta opna mállíkanið frá OpenAI, GPT-3 (forveri ChatGPT-forritsins) þjálfað á fimm hundruð milljörðum orða, þ.e. um fimm þúsund sinnum meiri gögnum en mannsbarn er „þjálfað“ á, og þótt hernaðarleynd ríki yfir gögnum stærstu líkananna í dag, má gera því skóna að gagnamagnið sé margfalt meira en svo. Hér skiptir máli að próf okkar leggja upp með að tryggja að þær íslensku setningar sem líkönin eru prófuð á ættu hvergi að koma fyrir í þjálfunargögnum líkananna (þ.e. að líkönin ættu ekki að hafa „séð þær áður“). Til dæmis var þess gætt að engin þeirra samsetninga sem við notuðum í orðmyndunarverkefninu fyndist við leit á Google, Tímarit.is eða í Risamálheildinni — í nokkurri mynd. Prófin snúa því að getu líkananna til þess að alhæfa málfræðireglur, sem virðist hvergi nærri því nógu öflug til þess að hægt sé að líta á þau sem kenningar um það hvernig við lærum mannlegt mál — líkt og tilhneiging er fulloft til.

## Heimildir

Ágústa Þorbergsdóttir, Starkaður Barkarson og Steinþór Steingrímsson. 2024. Orð ársins 2023: Gervigreind(in). *Hugrás*.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. M.I.T. Press, Cambridge.

Gilkerson, Jill, Jeffrey A. Richards, Steven F. Warren, Judith K. Montgomery, Charles R. Greenwood, D. Kimbrough Oller, John H.L. Hansen og Terrance D. Paul. 2017. Mapping the Early Language Environment Using All-Day Recordings and Automated Analysis. *American Journal of Speech-Language Pathology* 26,2:248-265.

Guðrún Kvaran. 2005. *Orð. Handbók um beygingar- og orðmyndunarfræði*. Íslensk tunga II. Almenna bókafélagið, Reykjavík.

Jeffrey L. Elman. 1990. Finding Structure in Time. *Cognitive Science* 14,2:179-211.

Linzen, Tal, og Marco Baroni. 2021. Syntactic Structure from Deep Learning. *Annual Review of Linguistics* 7:195-212.

## Mynd

Minnie Zhou. Unsplash.