

Áfram íslenska! – um keppni í vélþýðingum og niðurstöður hennar

Á hverju ári fer fram keppni í vélþýðingum í tengslum við stærstu vélþýðingaráðstefnu í heiminum, sem heitir einfaldlega *Conference on Machine Translation*, en er skammstafað WMT – af sögulegum ástæðum. Í keppninni eru nokkur tungumálapör valin hverju sinni. Í ár voru þau sextán en aðeins önnur þýðingaráttin var skoðuð fyrir hvert par. Eitt tungumálaparanna á þessu ári var enska?íslenska, sem þýðir að mælt var hvernig kerfin stóðu sig í að þýða úr ensku yfir á íslensku en ekki úr íslensku yfir á ensku.

Hvernig er keppt í vélþýðingum?

Keppnin fer þannig fram að rannsóknarhópum, fyrirtækjum eða hverjum sem áhuga hefur er gefinn kostur á að skrá sína þýðingarvél til leiks en að auki eru vinsælustu þýðingarvélarnar og mállíkönin sem eru aðgengileg á netinu mæld af aðstandendum keppinnar. Í aðalkeppninni er leitast við að mæla almenna þýðingargetu kerfanna en til hliðar við keppnina býðst rannsóknarhópum að senda inn sértæk mælipróf, t.d. til að mæla kynjabjaga í þýðingum eða færni í þýðingum á ákveðnum sérsviðum. Í aðalkeppninni, þar sem almenn þýðingargeta er mæld, voru ferns konar textar metnir að þessu sinni: fréttatextar, bókmenntatextar, talað mál sem hefur verið skrifað upp og textar af samfélagsmiðlum. Textarnir eru allir nýir svo að nánast ómögulegt er að þeir hafi verið notaðir við þjálfun þýðingarvéla sem metnar eru. Til að hægt sé að fá samanburð á milli þýðingarvéla og manna sem þýða án hjálpar þýðingarvéla eru allir textarnir þýddir af atvinnuþýðendum. Þýðendurnir fá þau fyrirmæli að nota engar þýðingarvélar sér til aðstoðar en ef þeir gerðu það gæti orðið til skekkja og mælingar bent til þess að þau þýðingarkerfi sem þýðandi notaði sér til aðstoðar væru líkari mannlegum þýðingum en önnur. Allar slíkar skekkjur þarf að forðast.

Hvernig eru þýðingarnar metnar?

Til að fá mælingu sem er líkleg til að vera í samræmi við upplifun notenda mismunandi kerfa dugar ekki að notast aðeins við sjálfvirkar matsaðferðir, en um þær hefur verið fjallað áður hér á Mannamáli, heldur er fólk líka fengið til að fara yfir þýðingarnar. Heildarlengd textanna á ensku er um 37 þúsund orð. Á þessu ári voru 34 þýðingarkerfi borin saman og þannig nánast ómögulegt fyrir hvern og einn að fara yfir allar þýðingarnar. Þeir sem fara yfir þýðingarnar og meta þær skoða þess vegna aðeins lítinn hluta af heildinni en tölfraðileg marktækni er reiknuð út svo að niðurstöðurnar verði skýrar. Þannig fæst röð þar sem sum kerfin geta lent í þyrpingum, t.d. 6.-8. sæti ef ekki er tölfraðilega marktækur munur á þeim. Við mat á þýðingum úr ensku yfir í íslensku voru vanir þýðendur fengnir til að fara yfir en yfirferðin felst í því að finna villur í þýðingunum, merkja við hvort þær séu smávægilegar eða alvarlegar og gefa svo einkunn á bilinu 0-100, þar sem 0 er gefið fyrir gagnslausa þýðingu þar sem merkingin kemst ekki til skila og 100 fyrir setningu sem skilar merkingunni á réttu málsniði og engin ástæða er til að laga neitt.

Niðurstöður fyrir þýðingar á íslensku

Þrjátíu og fjögur kerfi voru metin eins og fyrr segir. Fyrst fór fram sjálfvirk mat sem notað er til að sigta út allra lökustu kerfin. Ekki þykir ástæða til að fara yfir þau í höndunum. Eftir ítarlega yfirferð sérfræðinganna á úttaki 18 kerfa lá niðurstaðan fyrir. Ekkert vélþýðingakerfi eða mállíkan komst með tærnar þar sem atvinnuþýðandinn var með hælana – og athugið að hann fékk ekki að nota neinar þýðingarvélar sér til aðstoðar. Það segir okkur að þar sem þörf er á vönduðum þýðingum borgar sig að treysta frekar á fólk sem kann til verka en gervigreind.

Enska→Íslenska		
Röð	Kerfi	Einkunn
1-1	Atvinnuþýðandi	87.5
2-2	Gemini-2.5-Pro	77.6
3-4	Erlendur	68.3
3-4	GPT-4.1	68.0
5-5	Shy-hunyuan-MT	63.2
6-6	TowerPlus-9B[M]	57.4
7-7	ONLINE-B	51.8
8-10	Claude-4	47.8
8-10	TowerPlus-72B[M]	46.3
8-10	TranssionTranslate	46.2
11-11	AMI	39.9
12-12	GemTrans	34.8
13-14	SalamandraTA	31.3
13-15	UvA-MT	30.6
14-15	CommandA-WMT	29.0
16-16	NLLB	24.1
17-17	IRB-MT	20.7
18-18	Gemma-3-12B	16.5
19-19	Llama-3.1-8B	10.5
20-35	15 kerfi ekki yfirfarin af fólki	

En hvaða vélþýðingakerfi stóðu sig best? Gemini 2.5 Pro, mállíkanið frá Google, fékk langhæstu einkunnina. Það er þó ekki hefðbundin þýðingarvél heldur mállíkan sem hægt er að biðja um að þýða með skipun á mannamáli, eins og þegar við viljum fá svör frá gervigreindarforritum sem byggja á myndandi gervigreind. Skipunin sem sett er framan við enska frumtextann sem á að þýða var eftirfarandi:

You are a professional English-to-Icelandic translator. Your goal is to accurately convey the meaning and nuances of the original English text while adhering to Icelandic grammar, vocabulary, and cultural sensitivities. Produce only the Icelandic translation, without any additional explanations or commentary. Translate the following text:

Setningunni er ætlað að gera líkaninu það alveg skýrt hvers ætlast er til af því en með þeim hætti að það geti endurspeglað það hvernig notendur þessara verkfæra segja þeim fyrir verkum.

I 2.-3. sæti voru svo GPT-4.1 frá OpenAI og Erlendur frá Miðeind. GPT-4.1 er mállíkan sem var skipað að þýða með sama hætti og Gemini 2.5 Pro en Erlendur er þýðingarvél sem byggist á stóru mállíkani, sem var Claude 3.5 Sonnet þegar keppnin fór fram. Erlendur vinnur þýðinguna í tveimur lotum. Í þeirri fyrri er mállíkaninu send fyrirspurn þar sem það er beðið um að greina frumtextann. Lykilhugtök og nöfn eru dregin út úr textanum, enskum orðum flett upp í ensk-íslenskri orðabók og föst orðasambönd merkt sérstaklega. Í seinni lotunni eru þessar upplýsingar sendar með frumtextanum þegar mállíkanið er beðið um að þýða textann. Að lokum er þýðingin sett í gegnum málfarsleiðréttingakerfi sem ætti að

finna flestar stafsetningar- og málfarsvillur í þýðingunni.

Öll þessi þrjú kerfi hafa breyst síðan mælingarnar fóru fram um mitt þetta ár. Ný útgáfa er komin af Gemini og líklega er Gemini 3 Pro nokkuð betra í að fást við þýðingar en útgáfa 2.5, þó að það hafi ekki enn verið mælt. OpenAI hefur gefið út GPT-5.1 og hugsanlega gildir það sama þar. Erlendur notar ekki lengur Claude 3.5 Sonnet í bakgrunni heldur hefur skipt yfir í Gemini, sem má reikna með að skili einnig betri þýðingum en áður.

Vinsælasta vélþýðingarkerfið á Íslandi, Google Translate, er ekki tilgreint í þessari töflu en nokkur vinsæl vélþýðingarkerfi á netinu eru skráð þar sem ONLINE-kerfi og aðskilin með bókstaf. Einfaldur samanburður bendir sterklega til þess að ONLINE-B sé Google Translate. Það er í sjötta sæti vélþýðingarkerfanna og það efsta sem ekki er víst að byggist á stóru mállíkani. Niðurstöðurnar sýna þannig nokkuð skýrt að á næstu árum munu vélþýðingarkerfi byggjast á þeirri tækni.

Til hliðar við aðalkeppnina á WMT eru aðrir þættir í vélþýðingum skoðaðir. Rannsóknarhópum býðst til dæmis að senda inn sín eigin próf og fyrir þýðingar á íslensku voru kerfin t.d. prófuð í þýðingum á Evrópureglugerðum og á ípróttamáli. Þar kom í ljós að engu þýðingarkerfi tókst að ná 80% af hugtökum í Evrópureglugerðum rétt og svipaðar niðurstöður fengust fyrir ípróttapýðingar þar sem bestu kerfin þýddu sértækan orðaforða í ípróttum rétt í einungis 70-80% tilvika. Þær niðurstöður segja okkur að þegar við þurfum að þýða efni með sértækum orðaforða er líklegt að talsvert verði um villur í niðurstöðunum, og það sem mikilvægast er að komist rétt til skila í þýðingunni er líklegra til að vera þýtt vitlaust en annað, ef við ætlum einvörðungu að treysta á þýðingarvél.

Vélþýðingar fyrir íslensku hafa stórbatnað á undanförunum árum og nýjustu mælingar benda til þess að bestu kerfin geti í mörgum tilvikum skilað góðum þýðingum. Hins vegar ber nokkuð á milli þegar við berum vélþýðingar saman við mennskan þýðanda og sá munur eykst talsvert þegar efni með sértækum orðaforða er þýtt. Það er því ljóst að notkun þýðingarvéla á ekki alltaf við – þó að í sumum tilvikum geti þær reynst afar gagnlegar. Í síðasta pistlinum í þessari röð um vélþýðingar, sem birtur verður á Mannamáli 12. desember, verður reynt að svara því hvenær og hvernig við ættum að nota þýðingarvélar og hvenær við ættum að forðast það.

Aðrar greinar í greinaflokknum

Steinþór Steingrímsson. [Hvernig getum við vitað hvaða þýðingarvél er best?](#) Mannamál.is, 10. Nóvember 2025.

Heimildir

Tom Kocmi, Ekaterina Artemova, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Konstantin Dranch, Anton Dvorkovich, Sergey Dukanov, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Howard Lakouga, Jessica Lundin, Christof Monz, Kenton Murray, Masaaki Nagata, Stefano Perrella, Lorenzo Proietti, Martin Popel, Maja Popović, Parker Riley, Mariya Shmatova, Steinþór Steingrímsson, Lisa Yankovskaya and Vilém Zouhar. 2025. [Findings of the WMT25 General Machine Translation Shared Task: Time to Stop Evaluating on Easy Test Sets](#). Í: Proceedings of the Tenth Conference on Machine Translation, pages 355–413, Suzhou, China. Association for Computational Linguistics.

Svanhvít Lilja Ingólfssdóttir, Haukur Jónsson, Kári Steinn Aðalsteinsson, Róbert Fjölfnir Birkisson,

Sveinbjörn Þórðarson og Þorvaldur Páll Helgason. 2025. [Midheind at WMT25 General Machine Translation Task](#). Í Proceedings of the Tenth Conference on Machine Translation, pages 577–582, Suzhou, China. Association for Computational Linguistics.

Selma Dís Hauksdóttir and Steinnþór Steingrímsson. 2025. [Automated Evaluation for Terminology Translation Related to the EEA Agreement](#). Í: Proceedings of the Tenth Conference on Machine Translation, pages 850–855, Suzhou, China. Association for Computational Linguistics.

Einar Freyr Sigurðsson, Magnús Már Magnússon, Atli Jasonarson and Steinnþór Steingrímsson. 2025. [Up to Par? MT Systems Take a Shot at Sports Terminology](#). Í: Proceedings of the Tenth Conference on Machine Translation, pages 856–865, Suzhou, China. Association for Computational Linguistics.

Miðeind. [Frábær árangur Erlends í vélþýðingakeppnum WMT25](#). Miðeind.is, 4. nóvember 2025.