

Govoriš nevronske?

Kako ljudje razumemo jezik sodobnih strojnih prevajalnikov

David BORDON

Filozofska fakulteta, Univerza v Ljubljani

Namen prispevka je predstaviti raziskavo preverjanja razumljivosti nerevidiranih strojno prevedenih spletnih besedil. Primarni udeleženci v raziskavi so bili splošni bralci in ne izurjeni prevajalci ali popraviljalci strojnih prevodov. Gre za prvo tovrstno raziskavo, ki je bila izvedena za slovenski jezik. Cilj raziskave je bil preveriti, v kolikšni meri so nerevidirani strojni prevodi razumljivi splošnemu bralstvu, pri čemer sem se posvetil tudi vplivu besedilnega in slikovnega konteksta. Preverjal sem prevode prevajalnikov Google Translate in eTranslation. Raziskava je bila izvedena z anketo, v kateri so udeleženci odgovarjali na vprašanja, ki so preverjala razumevanje spremljajočega besedilnega segmenta, v katerem je bila napaka. Rezultati nudijo vpogled v trenutno stopnjo razvoja strojnih prevajalnikov, ne z vidika storilnosti pri njihovem popravljanju, ampak z vidika, koliko jih razume ciljno bralstvo. Na koncu članka nudim novo evalvacijo izvornih segmentov, ki sem jih v začetku leta 2023 ponovno prevedel, tokrat še s prevajalnikom DeepL.

Ključne besede: nerevidirani strojni prevodi, evalvacija strojnih prevajalnikov, razumljivost pri končnih bralcih, Google Translate, eTranslation, DeepL

Bordon, D.: Govoriš nevronske? Kako ljudje razumemo jezik sodobnih strojnih prevajalnikov. Slovenščina 2.0, 11(1): 138–159.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2023.1.138-159>

<https://creativecommons.org/licenses/by-sa/4.0/>



1 Uvod

Pričujoči razširjeni članek nadgrajuje objavo v zborniku konference JTDH 2022 (Bordon, 2022) in mu dodaja poglavje št 6, v katerem evalviram prevodne rešitve izvirne raziskave z nadgrajenimi prevajalniki v letu 2023. V članku obravnavam raziskavo razumljivosti strojno prevedenih spletnih besedil pri bralcih, ki ne vedo, da prebirajo strojne prevode. Uporabil sem naključno izbrana angleška spletna besedila, slovenske prevode pa sem pridobil z nevronskima strojnima prevajalnikoma Google Translate in eTranslation. Prevodi niso bili revidirani, saj sem želel replicirati okoliščine, v katerih bi jih dejansko lahko našli – na spletu, kjer so zaradi (za nekatere) dovolj visoke kakovosti in cenovne nepremagljivosti (namreč so brezplačni) vedno pogostejša¹, kar velja tudi za prevajalske vtičnike, ki so vgrajeni v sodobne brskalnike in aplikacije. Uporabniki se nasploh vedno več poslužujejo strojnega prevajanja (Vieira et al., 2022).

Vprašanje razumljivosti v taki obliki je postalo aktualno samo v zadnjem času, saj so starejši, statistični modeli prevajalnikov slovnično nekonsistentni in jezikovno okorni, sodobni nevronske prevajalniki pa proizvajajo tekoča besedila, ki so težje ločljiva od človeških, hkrati pa je že profesionalnim pregledovalcem prevodov težje ugotoviti, kje so storili napako (Donaj in Sepesy Maučec, 2018).

Te napake nastanejo predvsem zaradi težav pri razdvoumljanju večpomenskih besed in prevajanju besed, ki jih ni v podatkovni zbirki, s katero smo prevajalnik učili (Thi-Vinh et al., 2019, str. 207; Koehn in Knowles, 2017, str. 28, 31–33; Sennrich et al., 2016, str. 3). Kljub morebitnim posamičnim napačno prevedenim besedam pa lahko ljudje pomen razberemo iz sobesedila. Pri preverjanju razumljivosti sem v vseh primerih vključil še kontekst, saj se v stvarnosti bralci nikoli ne srečujejo z izoliranimi besedami, ampak z zaključenimi besedili, ker pa se osredotočam na spletno okolje, sem besedilnemu kontekstu dodal še slikovnega, ki je pogost element na sodobnih spletnih straneh.

Namen članka je predstaviti grobo oceno razumljivosti prevodov NMT-sistemov (ang. *Neural machine translation*) v času, ko so taka

1 Nekaj primerov za »krompirjeve kline« (potato wedges) na spletu. sl.veg-recipes; sl.hiloved; sl.eathealthyeatgreek; sl.gastromium; sl.atomiyme – strojno prevedene spletne strani za mnogo različnih jezikov.

besedila na spletu vedno pogostejša, pri čemer me zanima predvsem, kako slikovno gradivo v besedilnem kontekstu vpliva na rezultate. Tovrstna raziskava za slovenščino še ni bila izvedena.

2 Sorodne raziskave

Raziskav na področju razumevanja nerevidiranih strojnih prevodov pri končnih bralcih je razmeroma malo, saj je z vidika omejenosti na stroko in gospodarstvu bolj zanimive analize storilnosti pri popravljanju prevodov veliko več raziskav osredotočenih zgolj na prevajalce.

Na Univerzi v Gentu je bila v sklopu projekta AristoCAT izvedena raziskava o razumevanju izmišljenih besed in samostalniških besednih zvez (Macken et al., 2019). Primeri, ki so bili iz angleščine v nizozemščino prevedeni s strojnima prevajalnikoma Google Translate in DeepL, so bili predstavljeni samostojno ali v kontekstu povedi, pri tem pa udeleženci niso imeli dostopa do izvirnega besedila. V povprečju so udeleženci ugotovili pravičen pomen izmišljene besede v 23 % primerov, ko je beseda bila predstavljena brez konteksta. Če ji je bil dodan kontekst, je bilo 41 % odgovorov pravičnih; v scenariju, ko je bila izmišljena beseda predstavljena v povedi in so udeleženci morali izbrati pravičen pomen, je bilo pravičnih odgovorov 56 %.

V sklopu istega projekta je bila izvedena še analiza bralnega razumevanja človeškega prevoda na eni in nepopravljenega strojnega prevoda na drugi strani. Človeški prevodi so bili ocenjeni bolje z vidika jasnosti podajanja informacij, z vidika končnega razumevanja pa je bila razlika manjša (Macken in Ghyselen, 2018).

Castilho in Guerberof Arenas (2018) sta izvedli primerjalno analizo bralnega razumevanja za statistični in nevronske model strojnega prevajalnika v primerjavi s človeškim izvirnikom. Glede na omejen vzorec (6 udeležencev) in nedoslednost rezultatov je ugotovitev, da sistemi-NMT izkazujejo najboljše rezultate, občasno še boljše kot angleški izvirnik, nedokončna.

Martindale in Carpuat (2018) sta v raziskavi obravnavali odziv bralcev na tekočnost in natančnost nevronske strojne prevode, ob tem pa sta preverjali stopnjo zaupanja informacijam v besedilu. Ugotovili sta, da bralce zelo zmotijo prevodi, ki niso tekoči.

Izsledke potrjuje tudi Popović (2020). V njenem eksperimentu so bralci v 30 % primerov zaradi zavajajoče tekočnosti sprejeli popolnoma napačno informacijo, še 25 % dodatnih primerov pa je bilo skoraj popolnoma (narobe) razumljivih.

Na tem mestu velja omeniti, da so se nedavno začele pojavljati bolj eksperimentalne metode prevajanja, katerih značilnost je upoštevanje multimedijskega konteksta, denimo zvočnega ali slikovnega. Lala in Specia (2018) sta razvila model multimedijskega leksikalnega prevajanja, katerega namen je prevajanje dvoumnih večpomenskih besed s pomočjo slikovnega konteksta. Sulubacak et al. (2020) so predstavili sorodne raziskave, uporabne podatkovne zbirke in metode raziskovanja na področju multimedijskega strojnega prevajanja, ki so vezane na prevajanje z zvokom, sliko in videom. Med novejšimi raziskavami Liu (2021) ponuja nevronske model vizualno-tekstovnega enkodiranja in dekodiranja.

Pričakujemo lahko, da se bo to področje v bodoče še hitreje razvijalo, predvsem zaradi tehnološkega napredka v drugih panogah (prepoznavanje slik, sinteza govora, avtomatsko podnaslavljanje ipd.).

3 Metoda

Raziskava je bila zasnovana okrog vprašalnika, ki je vseboval primere štirih vrst napak v slovenskih strojnih prevodih splošnih angleških spletnih besedil. Preverjal sem prevajalnika Google Translate in eTranslation, pri čemer je bil vsak zastopan z 12 vprašanji. Distribucijo vrst napak opredelim v podpoglavju 3.3. Poseben pomen sem posvetil slikovnemu gradivu v sobesedilu.

3.1 Izbor besedil

Besedila sem zbiral glede na verjetnost, da bi se bralci z njimi lahko dejansko srečali na spletu. Analiza prevajalskega trga je pokazala, da večje prevajalske agencije popolnoma obvladujejo sektorje, ki nudijo največ dobička in hkrati zahtevajo človeško revizijo (tehnika, zdravstvo, pravo, finance ipd.) (Evropska komisija, 2020). V manj dobičkonosnih sektorjih, kjer človeška revizija ni tako pomembna, obstaja večja verjetnost objave nerevidiranih strojnih prevodov.

Pregleda tržnega deleža spletnih iskalnikov, ki jih uporabljamo v Sloveniji, je pokazal, da 96 % vseh uporabnikov spleta uporablja iskalnik Google.² Na osnovi najbolj iskanih pojmov v brskalniku³ sem izločil spletišča, ki nimajo prevodnega potenciala (družbena omrežja, spletni portali v slovenščini, slovenski mediji). S tem sem prišel do končnega izbora besedilnih področij: spletno nakupovanje, turizem, elektronika, multimedija in videoigre, luksuzne storitve, moda, osebno zdravje (telesna vadba in prehrana).

3.2 Prevodi besedil

Pri preizkušanju strojnih prevajalnikov se je izkazalo, da Googlov prevajalnik nudi drugačne prevodne rešitve glede na to, kako besedilo naložimo v obdelavo. Če besedilo prevajamo v pogovornem oknu vmesnika ali v brskalniku prevedemo spletno stran kot celoto, so rezultati boljši kot tisti, ki jih dobimo s funkcijo prevajanja dokumenta. Od štirih različnih specializiranih domen, ki jih nudi eTranslation, je najboljšše rezultate nudil prevajalnik za splošna besedila (General Text). Uporabil sem najboljše možne prevode – omenjeno domeno v eTranslation, v Googlu pa sem prevajal v pogovornem oknu.

Tabela 1: Razlike v prevodih glede na način obdelave; Google Translate

Prevod iz vnosnega polja oz. samodejni prevod strani	Prevod, pridobljen s funkcijo »prevodi dokument«	Izvirnik
Naj bo topla - mikrovalovna pečica ohranja hrano, kot so zelenjava, juhe, jedi, graviža, omake in sladice, topla in okusna v pečici, dokler niso pripravljene za postrežbo.	Naj bo topla funkcijo - Mikrovalovna ohranja živila, kot so zelenjava, juhe, nerazporejenega d'oeuvres, gravies, omake in sladice toplo in okusno v pečice, dokler oni pripravljene, da služijo.	Keep Warm Feature Maintains Food Temperature Keeps foods like vegetables, soups, hors d'oeuvres, gravies, sauces and desserts warm and delicious in the oven until they're ready to serve.

Tabela 2: Prevod enakega segmenta; eTranslation

Prevod modela »General Text« prevajalnika eTranslation
Ohraniti topla funkcijo - Microwave ohranja hrano, kot so zelenjava, juhe, predjed d'oeuvres, omake, omake in sladice tople in okusne v pečici, dokler niso pripravljene za postrežbo

2 <https://gs.statcounter.com/search-engine-market-share/all/slovenia>

3 <https://ahrefs.com/keyword-generator>

3.3 Kategorizacija napak

Prevode sem analiziral in določil štiri kategorije najpogostejših napak, ki niso vezane na jezikovni sistem oz. predpis – raziskava zanemarja slovnico in se osredotoča izključno na leksikalne napake.

- **Neprevedena beseda;** v prevodu se pojavlja beseda v enaki obliki kot v izvorniku. Dopustil sem možnost spremembe začetnih ali končnih morfemov, če je prevajalnik besedo samo preoblikoval.⁴
 - o Primer 22 – samodejno namakanje – *loosens madeže*.
- **Napaka pri razdvoumljanju večpomenske besede;** denotativni pomen večpomenske besede ali besedne zveze ne ustreza pomenu v izvorniku.
 - o Primer 11 – torba za pedal za bas boben – *primer vključen - »case included«*.
- **Hujša pomenska napaka;** napaka, ki otežuje razumevanje celotnega besedila.
 - o Primer 18 – naprava za hranjenje hišnih ljubljencev. *Baterije vstavimo ali odstranimo*.
- **Izmišljena beseda;** prevajalnik si zaradi kodiranja na enote manjše od besede (subword encoding) pri ponovnem sestavljanju besede v fazi dekodiranja »izmisli« besedo, ki ni v rabi in je denimo ni moč najti v referenčnih korpusih ali v drugih jezikovnih virih – t. i. »nevronščina«⁵.
 - o Primer 15 – dvopojasni Wi-Fi – *dvopasovni Wi-Fi*.

Vsi primeri so vizualno predstavljeni v poglavju 3.6.

Končni nabor je obsegal 12 primerov na vprašalnik, skupno 24. Distribucija števila primerov glede na vrsto napake je bila sledeča:

- neprevedena beseda: 2 primera na prevajalnik;
- napaka pri razdvoumljanju: 4 primeri na prevajalnik;
- hujša pomenska napaka: 3 primeri na prevajalnik;
- izmišljena beseda: 3 primeri na prevajalnik.

4 Denimo, prevod za rob zaslona (ang. *bezel*, je prevajalnik prevedel kot »bezela«).

5 <https://www.alternator.science/sl/daljse/z-nevronsco-v-prihodnost/>

3.4 Kontekst

Izbranim besedilom sem glede na inherentne lastnosti spletne pojavitve dodal kontekst. Kontekst je lahko bil več vrst:

- izključno besedilni,
- besedilni in slikovni; slika ne vpliva na razumevanje,
- besedilni in slikovni; slika vpliva na razumevanje,
- izbor ene izmed več predlaganih slik glede na to, kaj piše v besedilu.

Slikovni kontekst sem vključil pri besedilih, ob katerih so se na spletu pojavljale fotografije, ki so pri nekaterih primerih bile zgolj vizualni dodatek, pri drugih pa je bilo pravilno razumevanje besedila vezano na prepoznavanje pravilnega vizualnega elementa.

V svoji raziskavi besed nisem nikoli predstavil v izolaciji, kot so to denimo storili v raziskavi Macken in drugi (2019), saj to niso realne okoliščine – napake v objavljenih strojnih prevodih bodo vedno del nekoga besedila. Besedil nisem popravljaj, anketirancem sem jih dal v branje vključujoč vse slovnične in pomenske napake, kot bi jih lahko sami prebrali na spletu.

3.5 Oblikovanje vprašalnika, format odgovorov na vprašanja in udeleženci

Anketo sem ustvaril na platformi Google Forms, ki nudi podporo za prikaz slik in dober vmesnik za pregled in izvoz rezultatov. Pomembno je poudariti, da anketirancem nisem razkril, da bodo brali strojno prevedena besedila. Omenil sem, da bodo »prebrali več kratkih besedil, ki so napisana v nekoliko okorni slovenščini«.

Vrste odgovorov so bile omejene s funkcionalnostjo platforme Google Forms in niso sledile nobeni logični metodi; določil sem jih subjektivno glede na vsebino primera in vrsto napake. Gre za najbolj nezanesljivo spremenljivko v metodi, saj bi s formulacijo vprašanja lahko sugeriral pravi odgovor, zanimalo pa me je predvsem to, če prihaja do večjega odstopanja glede na tip odgovora, denimo, če bi bili odgovori odprtega tipa, kjer anketiranci vnesejo svoj odgovor v prazno vnosno polje, bistveno slabši kot tisti, kjer izbirajo med štirimi predlaganimi odgovori. S tem bi lahko preveril konsistenco pravilnosti oz. odstopanja glede na vrsto odgovora.

Vključil sem tri tipe podajanja odgovorov na vprašanja o razumljivosti besedil:

- odgovor odprtega tipa; anketiranci vpišejo odgovor v vnosno polje,
- odgovor zaprtega tipa (A, B, C ali D),
- izbor z razlago (A ali B, zakaj?).

Vprašalnik sem delil na družbenih omrežjih Facebook in Instagram in znance pozval, naj ga posredujejo naprej svojcem in svojim znancem, če je le mogoče starejšim. Demografskih podatkov nisem zbiral, izjema je zgolj podatek, če se oseba, udeležena v anketi, ukvarja s prevajanjem, kar je ena izmed pomanjkljivosti raziskave. Glede na razmeroma majhen vzorec sodelujočih in morebiten efekt odmevne komore bi bilo vsekakor raziskavo potrebno nadgraditi in ponoviti na bolj naključnem in predvsem večjem vzorcu, toda glede na čas zbiranja odzivov, ki je sovpadal s prvo omejitvijo gibanja vezano na epidemijo Covid-19, nisem imel druge izbire.

Na vprašalnik sem prejel 120 odgovorov.

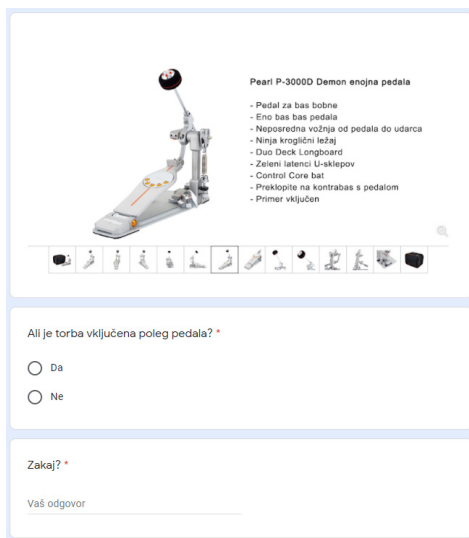
3.6 Primeri vprašanj in odgovorov v vprašalniku

V tem podpoglavju predstavljam par praktičnih primerov vprašanj iz ankete. Izbral sem tri različne primere; vsak ima različno vrsto odgovora, konteksta in napake. Celoten vprašalnik je dostopen na spletu⁶. Metodološki pristop je bil interpretativen – zavedam se mnogih morebitnih pomanjkljivosti, ki so vezane na to, kako so bila vprašanja izbrana, kako so formulirani potencialni odgovori in kako je razumevanje nekega besedila morda pogojeno s poznavanjem tehničnih vidikov področja, kateremu pripada. Bolje bi bilo vprašalnik standardizirati in uporabiti tako zasnovano anketo, ki obenem omogoča tudi boljšo statistično analizo.

Primer 11: torba za pedal za bas boben – *primer vključen* – »case included«.

- Vrsta napake: napaka pri razdvoumljanju večpomenske besede.
- Vrsta konteksta: besedilni in slikovni, kjer slika vpliva na razumevanje.
- Vrsta odgovora: izbor (A ali B) in razlaga.

⁶ <https://forms.gle/NKnEHrcEgJa7Tydo9>



Pearl P-3000D Demon enojna pedala

- Pedal za bas bobne
- Eno bas bas pedala
- Neposredna vožnja od pedala do udarca
- Ninja kroglični ležaj
- Duo Deck Longboard
- Zelena latenci U-sklepov
- Control Core bas
- Preklopite na kontrabas s pedalom
- Primer vključen

Ali je torba vključena poleg pedala? *

Da

Ne

Zakaj? *

Vaš odgovor _____

Slika 1: Primer 11.

Primer 22: samodejno namakanje »loosens madeže«.

- Vrsta napake: neprevedena beseda.
- Vrsta konteksta: besedilni in slikovni, kjer slika ne vpliva na razumevanje.
- Vrsta odgovora: odprti tip.



Samodejno namakanje - Loosens madeže z namakanjem do 2 uri

Samodejno namakanje

15 Min 30 Min 1 Hr 2 Hr Off

Kaj se zgodi z madeži po dveh urah? *

Vaš odgovor _____

Slika 2: Primer 22.

Primer 18: naprava za hranjenje hišnih ljubljencev. *Baterije vstavimo ali odstranimo.*

- Vrsta napake: hujša pomenska napaka.
- Vrsta konteksta: besedilni in slikovni, kjer slika ne vpliva na razumevanje.
- Vrsta odgovora: izbor (A ali B) in razlaga.

OPOMBA: Predlagamo, da baterije odložite tudi v kolikor ste odsotni, da se elektriki ne bo zgodilo kaj nepričakovanega. - Opomba: Predlagamo, da v svoje baterije vstavite tudi akumulatorje, ko se ne odpravite, da se elektrika ne bo zgodila.


Ali nam zgornje besedilo sporoča, da moramo baterije vstaviti ali odstraniti? *

Vstaviti

Odstraniti

Zakaj? *

Vaš odgovor



Slika 3: Primer 18.

Primer 15: dvopojasni Wi-Fi – *dvopasovni Wi-Fi.*

- Vrsta napake: izmišljena beseda.
- Vrsta konteksta: samo besedilni.
- Vrsta odgovora: zaprti tip.

Hitrejše in bolj priročne povezave - Zelo hiter dvopojasni Wi-Fi 5 (802.11ac) omogoča povezavo Vivo AiO V241 s katerim koli brezžičnim omrežjem z najboljšo možno hitrostjo in z manj motnjami. Z največjo hitrostjo 867Mbps je skoraj tako hitro kot ožičeno omrežje!

Kaj je lastnost tehnologije dvopojasni Wi-Fi? *

Dvakratna hitrost

Deluje na dveh frekvencah

Dvakrat bolj intuitiven

Slika 4: Primer 15.

4 Rezultati

Rezultate predstavljam po naslednjih parametrih:

- splošno razumevanje,
- razumevanje glede na prevajalnik,
- razumevanje glede na tip napake,
- razumevanje glede na tip konteksta,
- razumevanje glede na tip odgovora.

4.1 Splošno razumevanje

Vprašalnik je obsegal 24 vprašanj, s 120 odzivi je bilo vseh možnih odgovorov 2.880. Vseh pravih odgovorov je bilo 1.697 oz. 58,96 %. Daljša razčlemba je na voljo v celotni raziskavi (Bordon, 2021).

4.2 Razumevanje glede na prevajalnik

Odgovori na vprašanja, vezana na prevajalnik Google Translate, so bili pravilni v 51,3 % primerov oz. 739 od 1.440 odgovorov. Prevajalnik eTranslation je pokazal boljše rezultate, delež pravih odgovorov je znašal 66,6 %.

4.3 Razumevanje glede na tip napake

V vprašalniku so bili vključeni štirje tipi različnih napak. V alinejah ni zam tip napake in odstotek pravih odgovorov:

- izmišljena beseda: 48,5 %,
- neprevedena beseda: 64,8 %,
- napačno razdvoumljene večpomenske besede: 65,9 %,
- hujša pomenska napaka: 56,3 %.

4.4 Razumevanje glede na kontekst

V naslednjem segmentu predstavljam delež pravih odgovorov vezanih na kontekst:

- izključno besedilni: 60,4 %,
- besedilni in slikovni; slika ne vpliva na razumevanje: 44 %,
- besedilni in slikovni; slika vpliva na razumevanje: 69,8 %,

- izbor ene izmed več predlaganih slik glede na to, kaj piše v besedilu: 64,2 %.

4.5 Razumevanje glede na tip odgovora

V tem segmentu predstavljam rezultate glede na način izbora odgovora. Primarna funkcija te analize je preveriti konsistenco oz. morebitna odstopanja npr.; če so odgovori odprtega tipa, kjer anketiranci v prazno vnosno polje vnesejo poljuben odgovor, bistveno slabši kot tisti, kjer imajo na voljo denimo štiri predlagane odgovore, izberejo pa enega:

- odgovor odprtega tipa (vnosno polje): 36,3 %,
- odgovor zaprtega tipa (A, B, C ali D): 60,8 %,
- izbor z razlago (A ali B, zakaj?): 68,3 %.

Slabše rezultate pri odgovorih zaprtega tipa v primerjavi z ostalima dvema kategorijama je treba jemati z rezervo, saj so bili primeri s tako vrsto odgovora zgolj štirje. Samo določanje pravilnosti odgovora je pri takih primerih težje, osebno pa sem bil strog ocenjevalec, saj sem vse odgovore, ki niso bili popolnoma pravilni, označil za napačne – poleg tega, da je bil zbran pravi odgovor (A ali B) je razlaga v polju »zakaj?« morala odražati popolno razumevanje, da bi odgovor uvrstil med pravilne.

4.6 Skupina prevajalcev

Edini demografski podatek, ki sem ga zbiral, je, ali se oseba, ki odgovarja na vprašalnik, ukvarja s prevajanjem, več o tem v naslednjem poglavju. Pritrdilno je odgovorilo 24 udeležencev od 120. Pri teh osebah sem analiziral odgovore glede na vrsto napake in jih primerjal z neprevajalci. Nasploh so bili njihovi rezultati za 6 % boljši (63,7 %), po kategorijah pa:

- izmišljena beseda 53,5 % (+ 6,3 % boljše od neprevajalcev),
- neprevedena beseda 65,6 % (+ 1 %),
- razdvoumljanje večpomenske besede 70,8 % (+ 6,7 %),
- pomenska napaka 63,9 % (+ 9,6 %).

Ostalih demografskih podatkov nisem zbiral, kar je ena od slabosti raziskave. V primeru da bi podatki sovpadali z mojo predpostavko, da

niso relevantni, jih ne bi vključil, sedaj pa preprosto nimam podatkov, na katerih bi lahko utemeljil svojo odločitev.

5 Povzetek rezultatov

Pri pregledu rezultatov sem ugotovil, da je bil odstotek pravih odgovorov pri izboru strojno prevedenih segmentov, ki sem jih ocenjeval, 59 %. Od vseh 2.880 odgovorov je bilo 1.697 pravih.

Na tej točki je potrebno izpostaviti primer št. 6, ki je bil nasploh najslabše razumljen in je znižal povprečje rezultatov v vseh kategorijah, v katerih se je nahajal. Zelo verjetna razlaga zakaj je bil ta primer tako slabši od povprečja je dejstvo, da je bil ta primer glede modaliteta odgovora kategoriziran kot odgovor odprtega tipa (prazno vnosno polje) – če odgovor ni bil povsem točen, sem ga označil za napačnega – pravilna sta bila samo dva. Če bi v tem primeru anketirancem ponudil denimo zaprti tip odgovora, bi bil ta odstotek vsekakor višji, kar potrjuje, da bi iz vidika metodologije lahko bil pristop odlikovanja vprašanj in odgovorov boljši.

Tabela 3: Primer št. 6; »Mednopni vložek«

Slovenski prevod	Angleški izvirnik
En zmagovalec bo prejel grafično kartico GeForce RTX 2080 Ti Cyberpunk 2077 Edition.	One winner will receive the GeForce RTX 2080 Ti Cyberpunk 2077 Edition graphics card.
Vstop v predavanje je enostaven: 1. Prijavite se na forume ali ustvarite forumski račun . 2. Komentirajte to temo (BREZ CITIRANJA TE POSTAJE) in nam povejte, kaj želite narediti najbolj v Cyberpunku 2077. 3. Za potrditev vpisa vpišite svoje uporabniško ime v naš pripomoček za oddajo.	Entering the giveaway is easy: Sign in to the forums or create a forum account. Comment on this thread (WITHOUT QUOTING THIS POST) and tell us what you want to do most in Cyberpunk 2077. Sign your username in our giveaway widget to confirm your entry.
KAKO VSTOPITI: Če želite vstopiti, vnesite mednopni vložek in sledite navodilom za vstop v nagradne igrače.	HOW TO ENTER: To enter, submit your entry during the Sweepstakes Period and follow the directions to enter the Sweepstakes.

eTranslation je bil v povprečju za 15 % boljši od prevajalnika Google Translate, v katerem je bil omenjen primer. Nasploh pa je eTranslation kazal boljše rezultate. Najboljši rezultati glede na tip napake so bili vezani na razdvoumljanje besednega pomena (65,9 %), kar kaže, da

znamo ljudje nasploh dobro razbrati pomen iz sobesedila, na drugem mestu pa so bile neprevedene besede (64,8 %).

Rezultati so bili slabši, ko je prevajalnik napravil hujšo pomensko napako, ki je oteževala razumevanje celotnega segmenta (56,3 %), da-leč najslabše rezultate pa je bilo moč opaziti v kategoriji izmišljena beseda (48,5 %), v kateri je sicer bil prej omenjeni primer št. 6.

Glede na tip konteksta so bili najboljši rezultati pri primerih, kjer je slika vplivala na razumevanje (69,8 %) in kjer so udeleženci morali izbrati sliko, na katero se je nanašalo besedilo (64,2 %). Rezultati so bili nekoliko slabši v izključno tekstovnem kontekstu (60,4 %), najslabši rezultati pa so bili v kategoriji, kjer je bila besedilu priložena slika, ki ne vpliva na razumevanje oz. potencialno zmede udeleženca (44 %) – v tej kategoriji je bil tudi primer št. 6. Izkazalo se je, da slikovni kontekst, ki lahko potencialno vpliva na razumevanje besedilnega segmenta, pri strojnih prevodih v realnih okoliščinah, torej na spletu, z vsem pomožnim gradivom, igra pomembno vlogo.

Udeleženci, ki se sicer ukvarjajo s prevajanjem, so na splošno odgovarjali boljše od povprečja. Njihov delež uspešnosti je bil največji v kategoriji hujša pomenska napaka (+9,6 %), kar bi lahko pojasnili s tem, da zaradi »poklicne deformacije« bolj učinkovito razumejo kontekst.

Pri tem velja omeniti, da je edini demografski podatek, ki sem ga v sklopu raziskave zbral to, če se oseba ukvarja s prevajanjem ali ne. Posledično je problematično sklepati, kakšen je bil denimo nivo znanja angleškega jezika, kakšna je bila starost udeleženih, nivo izobrazbe ipd. V tem primeru gre za veliko pomanjkljivost pri metodologiji, ki grobo omejuje sposobnost poročanja o morebitnih zaključkih. Vse predpostavke bi bile zelo subjektivne in brez empiričnih podatkov je njihovo navajanje brezpredmetno.

6 Stanje leta 2023

V začetku leta 2023 sem besedila po več kot dveh letih in pol (besedila sem namreč zbral in strojno prevedel maja 2020) ponovno strojno prevedel. Zanimalo me je, če so se prevajalniki z nadgradnjami v tem času izboljšali in če so segmenti, kjer so proizvajali pomanjkljive prevode, sedaj bolje prevedeni.

Znova sem uporabil prevajalnika Google in eTranslation, dodal pa sem še prve rezultate iz prevajalnika DeepL, ki se je v zadnjih letih hitro umestil na sam vrh po kakovosti prevodnih rešitev in berljivosti.

Izpostaviti gre, da sem se pri ponovnem pregledu prevodnih rezultatov osredotočal zgolj na izbor primerov, ki sem jih vključil v sklopu raziskave. Rezultati v tem sklopu so zaradi tega morda nekoliko pristranski, saj če bi ocenjevali celotne segmente, bi lahko o razvoju kakovosti prevodov potegnili drugačne (slabše) zaključke, toda sem nad dotičnim izborom primerov, ki je morda nekoliko arbitraren, imel najboljši pregled in je omogočal relativno enostavno primerjalno analizo.

Primeri, ki sem jih uporabil v anketi, so pri obeh prevajalnikih, ki sem ju vključil v prvotno raziskavo, povzročali težave tako enemu kot drugemu sistemu. V anketi sem od 24 primerov vključil 12 primerov na prevajalnik, torej polovico, v tej evalvaciji pa primerjam takratno stanje z aktualnim za vseh 24 primerov. Za vsak izvorni primer sem označil, če je napaka prisotna ali ne – v določenih primerih je denimo napako storil Google in se s tem umestil v anketo, eTranslation pa je dal dobro rešitev. Za to evalvacijo sem označil vse primere za oba prevajalnika in jih po ponovnem prevodu analiziral. Rezultate sem umestil v kategorije:

- izvorna rešitev je dobra, sedanja je enako dobra,
- izvorna rešitev je dobra, sedanja je slabša,
- izvorna rešitev je slaba, sedanja je odlična,
- izvorna rešitev je slaba, sedanja je boljša,
- izvorna rešitev je slaba, sedanja je enako slaba.

6.1 eTranslation 2023

eTranslation je v prvotni raziskavi dal dobro rešitev pri štirih primerih, ostalih 20 je vsebovalo napako. Po ponovnem prevodu so rezultati sledeči:

- 2 izvorni rešitvi sta bili dobri in ostajata enako dobri,
- 2 izvorni sta bili dobri in sta zdaj slabi oz. vsebujeta napako,
- 6 napačnih rešitev je zdaj odličnih,
- 3 napačne rešitve so zdaj nekoliko izboljšane, vseeno pa ne povsem pravilni,
- ostalih 11 primerov ne kaže sprememb in ostaja napačnih.

6.2 Google Translate 2023

Prevajalnik Google je v izvorni raziskavi pravilno prevodno rešitev proizvedel pri dveh od vseh 24 primerov. Dobri dve leti kasneje kaže bistveno boljše rezultate:

- 2 izvorni rešitvi, ki sta bili dobri in ostajata enako dobri,
- 19 napačnih rešitev je zdaj povsem pravih,
- ena rešitev kaže izboljšanje, ni pa povsem pravilna,
- 2 napačni rešitvi ne kažeta sprememb in ostajata napačni.

6.3 DeepL 2023

Rezultate prevajalnika DeepL vključujem prvič, saj ni bil del izvorne raziskave – v tistem času še zdaleč ni užival takega ugleda in prominence kot sedaj, kar nakazuje na to, kako drastično se na tem področju dogajajo spremembe in izboljšave. Prevajalnik sem vključil predvsem zaradi tega, ker so primeri že izbrani in lahko zelo preprosto primerjam sodobne prevodne rešitve z ostalima dvema prevajalnikoma:

- pri 16 primerih je dal odlično rešitev;
- en primer je označen kot mejni, saj je izjemno dvoumen in ga je težko zares jasno umestiti v eno ali drugo kategorijo;
- ostalih 7 primerov je napačnih.

Pri tem gre takoj poudariti, da ima DeepL možnost spreminjanja leksemov v samem uporabniškem vmesniku. S klikom na besedo nam prevajalnik takoj ponudi morebitne alternative, kar je seveda orodje, ki primarno služi prevajalcem. V kontekstu, da bi bila besedila avtomatsko strojno prevedena in na spletu objavljena za končne uporabnike, bi lahko tovrstni tip orodja lahko še dodatno pomagal pri ugotavljanju pravega pomena oz. če bi se to orodje pojavilo v oblaku skupaj s segmentom v izvorniku, bi za osebo, ki ima že povprečno znanje angleškega jezika, verjetno že zadostovalo, da bi v veliki večini primerov prišla do pravega pomena.

Pri napačnih primerih in mejnem primeru sem kliknil na kritično besedo in mi je pri šestih od osmih ponudil pravilno rešitev med alternativami, kar mu daje potencial 22 pravih rešitev od 24.

6.4 Primerjava treh prevajalnikov v 2023

Na podlagi omejenega števila primerov in načina izbora opazovanja je težko z gotovostjo dejati, da se je kakovost strojnih prevodov radikalno izboljšala, je pa kljub vsemu na podlagi omenjenih primerov možno opaziti, da je v nekaj manj kot treh letih vsaj pri določenih primerih moč opaziti izboljšanje.

Če se osredotočimo na primere, ki sem jih vključil v raziskavo, je čas najbolj vplival na prevajalnik Google, ki v praksi kaže 20 pravih rešitev od 24. Prevajalnik DeepL ima trenutno pri danih primerih 8 napačnih rešitev, kot omenjeno pa kaže potencial, da bi lahko z dodatno funkcionalnostjo ponujanja ostalih prevodnih kandidatov dosegal skoraj popolno pravilnost, z izjemo enega mejnega rezultata. Bistveno slabše se je odrezal prevajalnik eTranslation – vsekakor kaže napredek, saj je v prvotni raziskavi dal pravilno rešitev samo v štirih primerih, to število se je januarja 2023 povzpelo na 8, še trije primeri pa kažejo blažjo izboljšavo – najbolj problematično je dejstvo, da je eTranslation pri dveh primerih celo nazadoval, vsi ostali pa so popolnoma enaki.

Z vidika razvoja najbolj kaže Googlu in prevajalniku DeepL, eTranslation pa je pri tem nekoliko bolj zadržan, se pa vseeno izboljšuje. Če bi želeli ugotoviti, kakšna je resnična stopnja razumevanja pri končnih uporabnikih, bi bilo treba eksperiment ponoviti in razširiti – ponavljam, da so rezultati, ki sem jih tu nanizal, arbitrarno presojeni iz vidika pravilnosti, podobno, kot sem to počel pri izvornih spletnih besedilih, ko sem zbiral »problematični« material za vprašalnik. Z današnjega vidika, vsi primeri, ki sem jih v tem segmentu označil za pravilne, ne bi sodili v anketo, saj se mi zdijo povsem neproblematični.⁷

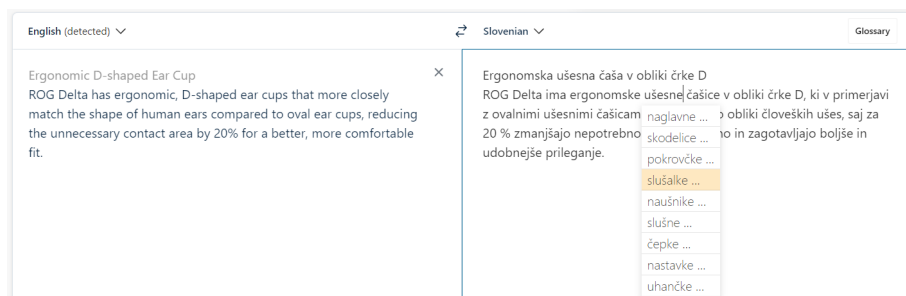
Tabela 4: Primer nazadovanja eTranslation

eTranslation 2020	Londonski stolp in Westminster sosednji lokalni pubi in tržnice ter časovno obdelani rituali, kot je menjava stražarjev, se odvijajo, ko vozači hitijo, da ujamejo Tube .
eTranslation 2023	Stolp v Londonu in Westminster sosednji lokalni pubi in trgi ter časovno obredi, kot je menjava stražarjev, se odvijajo, ko se vozači hitijo, da bi ujeli cev .
Angleški izvirnik	The Tower of London and Westminster neighbor local pubs and markets, and time-worn rituals like the changing of the guards take place as commuters rush to catch the Tube .

⁷ Preglednico z vsemi prevodnimi rezultati sem objavil na povezavi https://bit.ly/Preglednica_Nevronscina_2023.

Tabela 5: Primeri izboljšav prevajalnika Google

Google Translate 2020	<ul style="list-style-type: none"> • Turška rižota; • vnesite mednopni vložek; • krompirjevimi klini; • s kanadskim šminkerjem Simoneom Otisom; • primer vključen; • dvopojasni Wi-Fi 5; • na oprijemni površini je izrezan diamant srednje globine.
Google Translate 2023	<ul style="list-style-type: none"> • Puranja rižota; • oddajte svojo prijavo; • rezinami sladkega krompirja; • s kanadsko vizažistko Simone Otis; • etui vključen; • dvopasovni Wi-Fi 5; • prijemna površina ima srednje globoko diamantno narebrčenje.
Angleški izvirnik	<ul style="list-style-type: none"> • Turkey risotto; • submit your entry; • sweet potato wedges; • Canadian makeup artist Simone Otis; • case included; • dual-band Wi-Fi 5; • the gripping surface features medium-depth diamond knurling.



Slika 5: Primer pravilno ponujene rešitve pri sprva napačnem prevodu – DeepL.

7 Sklep

V članku sem predstavil raziskavo o razumljivost nerevidiranih strojno prevedenih spletnih besedil pri končnih uporabnikih, ki niso bili posebej obveščeni, da prebirajo strojne prevode. Razumevanje besedilnih segmentov, ki so vključevali štiri različne tipe napak, ki nastanejo pri strojnem prevajanju NMT-sistemov, sem preverjal z anketo. Ta je vsebovala strojne prevode splošnih besedil, ki sem jih prevedel s prevajalnikom Google Translate in eTranslation. Besedila so bila nerevidirana,

vsebovala so napake, ki so bile predstavljene v več različnih kontekstih, bodisi s slikovnim gradivom bodisi brez.

Rezultati so pokazali, da je splošna stopnja razumevanja 59 %, pri čemer se je izkazalo, da so prevodi eTranslationa nasploh razumljivejši od prevodov Googlovega prevajalnika. Število pravih odgovorov je bilo najvišje v kategoriji razdvoumljanja večpomenskih besed, kar nakazuje na to, da ljudje lažje razumemo pomen strojnih prevodov, če nam je dan kontekst. Pri tem je bilo najbolj učinkovito slikovno gradivo, s katerim so si lahko udeleženci v raziskavi pomagali razjasniti pomen določenega besedilnega segmenta.

Po analizi se je izkazalo, da je bil nekoliko problematičen način izbire odgovorov, saj sem anketirancem naključno vnaprej določil, na kakšen način bodo odgovarjali. Odgovori odprtega tipa so kazali slabše rezultate kot izbirni odgovori in odgovori zaprtega tipa, toda zaradi majhnega števila vprašanj je težko izpeljati kakšen razumen zaključek. Podobno velja za samo metodo odgovarjanja na anketo, ki je bila pogojena pandemičnemu času. Za bolj relevantne rezultate bi bilo potrebno izvajati test razumljivosti v živo, na razpravljalni način. Enako velja tudi za vzorec sodelujočih – večji in bolj raznolik vzorec bi dal jasnejše rezultate.

V bodoče bi bilo zanimivo raziskati, če se razumevanje nerevidiranih strojno prevedenih besedil izboljšuje skupaj z nadgradnjami strojnih prevajalnikov, hkrati pa bi se lahko osredotočil še na avtomatsko generirana besedila in jezik spletnih robotov.

Menim, da bo v prihodnje nekoliko manj raziskav storilnosti pri popravljanju strojnih prevodov in veliko več raziskav, ki bodo vezane na razumljivost strojno prevedenih ali avtomatsko generiranih besedil v praktičnih situacijah. Končni bralec se vedno bolj pogosto srečuje s takimi besedili, lahko pa pričakujemo, da bo zaradi še dodatnih izboljšav strojnih prevajalnikov, novih metod in razširjenosti prakse tovrstnih potencialnih stikov med stroji in bralci brez vmesnega posega človeškega popravilca vedno več.

Zahvala

Raziskovalni program št. P6-0436 (Digitalna humanistika: viri, orodja in metode) sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije iz državnega proračuna.

Literatura

- Bordon, D. (2022). Govoriš nevronska? Kako ljudje razumemo jezik sodobnih strojnih prevajalnikov. V D. Fišer & T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 286–291). Ljubljana: Inštitut za novejšo zgodovino. Pridobljeno s https://nl.ijs.si/jtdh22/pdf/JTDH2022_Proceedings.pdf
- Bordon, D. (2021). »Razumevati nevronska: Kako si ljudje razlagamo jezik strojnih prevajalnikov«. Magistrsko delo. Ljubljana: Univerza v Ljubljani. Pridobljeno s <https://repozitorij.uni-lj.si/IzpisGradiva.php?id=125328>
- Castilho, S., & Guerberof Arenas, A. (2018). Reading Comprehension of Machine Translation Output: What Makes for a Better Read?. V J. A. Perez-Ortiz, F. Sanchez-Martinez, M. Espla-Gomis, M. Popovič, C. Rico, A. Martins, J. Van den Bogaert, M. L. Forcada (ur.), *Proceedings of the 21st Annual Conference of the European Association for Machine Translation* (str. 79–88). Alacant, Španija. Pridobljeno s <http://doras.dcu.ie/23071/>
- Donaj, G., & Sepesy Maučec, M. (2018). Prehod iz statističnega strojnega prevajanja na prevajanje z nevronskimi omrežji za jezikovni par slovenščina-angleščina. V D. Fišer & A. Pančur (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika 2018* (str. 62–68). Ljubljana: Filozofska fakulteta, Inštitut za novejšo zgodovino. Pridobljeno s http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018_Donaj-et-al_Prehod-iz-statisticnega-strojnega-prevajanja-na-prevajanje-z-nevronskimi-omrezji-za-jezikovni-par-slovenscina-anglescina.pdf
- Evropska komisija (2020). European Language Industry Survey 2020 Before & After Covid-19. Pridobljeno s https://ec.europa.eu/info/sites/default/files/2019_language_industry_survey_report.pdf
- Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. V *Proceedings of the First Workshop on Neural Machine Translation* (str. 28–39). Vancouver, Canada: Association for Computational Linguistics. Pridobljeno s <https://arxiv.org/pdf/1706.03872.pdf>
- Lala, C., & Specia, L. (2018). Multimodal Lexical Translation. V *Proceedings of the 11th international conference on language resources and evaluation (LREC)* (str. 3810–3817). Miyazaki, Japonska: European Language Resources Association (ELRA). Pridobljeno s <https://www.aclweb.org/anthology/L18-1602/>
- Lelner, Z. (2022). Machine Translation vs. Machine Translation Post-editing: Which One to Use and When?. Pridobljeno s <https://blog.memoq.com/machine-translation-vs.-machine-translation-post-editing-which-one-to-use-and-when>

- Liu, J. (XX) Multimodal Machine Translation. Pridobljeno s <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9547270>
- Macken, L., & Ghyssele, I. (2018). Measuring Comprehension and User Perception of Neural Machine Translated Texts: A Pilot Study. V *Translating and the Computer 40 (TC40), Proceedings* (str. 120–126). Geneva: Editions Tradulex. Pridobljeno s <https://biblio.ugent.be/publication/8580951>
- Macken, L., Van Brussel, L., & Daems, J. (2019). NMT's wonderland where people turn into rabbits. A study on the comprehensibility of newly invented words in NMT output. V *Computational Linguistics in the Netherlands Journal*, 9, 67–80. Pridobljeno s <https://www.clinjournal.org/clinjournal/article/view/93>
- Martindale, M. J., & Carpuat, M. (2018). Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT. Pridobljeno s <https://arxiv.org/abs/1802.06041>
- Nunes Vieira, L., O'Sullivan, C., Zhang, X., & O'Hagan, M. (2022). Machine translation in society: insights from UK users. *Language Resources & Evaluation*. Pridobljeno s <https://doi.org/10.1007/s10579-022-09589-1>
- Popović, M. (2020). Relations between comprehensibility and adequacy errors in machine translation output. V R. Fernández & T. Linzen, *Proceedings of the 24th Conference on Computational Natural Language Learning (CoNLL 2020)* (str. 256–264). Pridobljeno s <https://aclanthology.org/2020.conll-1.19.pdf>
- Sennrich, R., Haddow, B., & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. Pridobljeno s <https://arxiv.org/abs/1508.07909>
- Sulubacak, U., Caglayan, O., Grönroos, S.-A., Rouhe, A., Elliott, D., Specia, L., Tiedemann, J. (2020). Multimodal machine translation through visuals and speech. Pridobljeno s <https://arxiv.org/abs/1911.12798>
- Thi-Vinh, N., Ha, T.-L., Nguyen, P.-T., & Nguyen, L.-M. (2019). Overcoming the Rare Word Problem for Low-Resource Language Pairs in Neural Machine Translation. V *Proceedings of the 6th Workshop on Asian Translation* (str. 207–214). Hong Kong, Kitajska: Association for Computational Linguistics. Pridobljeno s <https://arxiv.org/abs/1910.03467>
- Voroniak, D. (2022). Post-Editing of Machine Translation: Best Practices. Pridobljeno s <https://blog.crowdin.com/2022/03/30/mt-post-editing/>
- Zdarek, D. Machine Translation Post-editing Best Practices. Pridobljeno s <https://www.memsource.com/blog/post-editing-machine-translation-best-practices/>

Do you Speak Neuralese? How People Comprehend the Language of Modern MT Systems

The aim of this paper is to present a study on the comprehensibility of unedited machine-translated web texts. The primary participants in the study were general readers, not trained translators or post-editors, and it is the first study of its kind to be conducted for the Slovene language. The aim of the study was to examine the extent to which unedited machine translations are comprehensible to general readers, while giving focus to the influence of textual and pictorial context. The translations were obtained from Google Translate and eTranslation. The survey was conducted by means of a questionnaire, in which participants answered questions that tested their understanding of a text segment that included an error. The results provide an insight into the current state of development of machine translation engines, not from the point of view of PEMT, but from the point of view of how well machine translations are understood by the target readership. At the end of the article, I provide a new evaluation of MT output in the year 2023, including results for the DeepL MT engine.

Keywords: unedited machine translation, MT evaluation, Understandability by end readers, Google Translate, eTranslation, DeepL