

PRIJELAZ SA STATISTIČKOG NA NEURONSKI MODEL: USPOREDBA STROJNIH PRIJEVODA SA ŠVEDSKOGA NA HRVATSKI JEZIK

Sandra Ljubas

Sažetak

Cilj je ovog istraživanja provjeriti je li nedavni prijelaz Google prevoditelja (GP) sa statističkog na neuronski model utjecao na poboljšanje kvalitete strojno prevedenih tekstova sa švedskoga na hrvatski jezik. U studenom 2016., kada je GP još rabio statistički model prevođenja, preliminarnom je analizom pogrešaka ustanovljeno da se u prijevodima pojavljuje velik broj morfosintaktičkih i leksičkih pogrešaka koje ih čine gotovo u potpunosti nerazumljivima. Google je, međutim, ubrzo prešao na neuronski model u koji polaže velike nade. Iako neuronski model nije izravno dostupan za ovu jezičnu kombinaciju, istraživanje je pokazalo da novi prijevodi sadržavaju manje pogrešaka, unatoč tome što GP sada izostavlja ili neprevedenima ostavlja veći broj riječi. Ukupno gledano, prijevodi sa švedskoga na hrvatski sada su morfološki točniji i razumljiviji, ali neprecizniji te općenito još nedovoljno kvalitetni. To je u skladu i s rezultatima dosadašnjih istraživanja o neuronskom modelu.

1. Uvod

Godine 2016. analizom je pogrešaka pronađenih u strojno prevedenim tekstovima sa švedskoga na hrvatski jezik ustanovljeno kako je potrebno ulagati značajnije napore u razvoj jezičnih tehnologija za hrvatski jezik kako bi se u budućnosti mogli očekivati kvalitetniji strojni prijevodi na hrvatski jezik, osobito u kombinaciji s drugim manjim jezicima poput švedskog (Ljubas 2017: 60). To je istraživanje provedeno s pomoću Google prevoditelja (GP) koji se u to vrijeme još služio statističkim modelom strojnog prevođenja. Taj model izračunava „vjerojatnost da će se neka riječ prevesti nekom drugom ili da će se prijevodi dviju riječi koje se nalaze jedna pokraj druge također nalaziti jedni pored drugih“ (Ahrenberg i Merkel

2006). GP je svoje rezultate temeljio na sveobuhvatnom korpusu mrežno dostupnih tekstova, a ipak su njegovi odabiri prijevodnih ekvivalenata često bili manjkavi.

U želji da ostane konkurentan na tržištu Google neprestano provodi istraživanja i nastoji unaprijediti algoritme svog Prevoditelja. Krajem 2016. ih je ažurirao pa danas govorimo o Googleovu neuronskom modelu strojnog prevođenja. Riječ je o novome pristupu koji se temelji na načelima dubokog učenja (engl. *deep learning*) i u njega se već polažu velike nade. Rezultati dosadašnjih istraživanja s jedne strane ukazuju na to da bi se neuronskim pristupom mogli riješiti problemi sročnosti te sklanjanja i sprezanja (Castilho i sur. 2017: 111), što je u kontekstu hrvatskog kao morfološki bogatog jezika vrlo važno, ali s druge strane se zamijetilo da neuronski modeli imaju velikih poteškoća s riječima koje se pojavljuju rjeđe, da češće nego prije izbacuju ili neprevedenima ostavljaju dijelove izvornika i da su izlazni podaci općenito neprecizniji (Wu i sur. 2016: 2). Također se očekivalo da će neuronski modeli pružiti lošije rezultate u odnosu na statističke modele kada je riječ o prevođenju dužih segmenata (Popović 2015, cit. pr. Castilho i sur. 2017: 111) odnosno pokazalo se da imaju problema s rečenicama dužima od 35 riječi (Bentivogli et al 2016: 4). Detaljnije su se time pozabavili Toral i Sánchez-Cartagena u istraživanju gdje su proučavali uspješnost neuronskog modela u odnosu na statistički u rečenicama duljine do 5 riječi, do 10 riječi i tako dalje do rečenica duljih od 50 riječi (2017: 7). Rezultati neuronskih modela bili su bolji sve do rečenica duljine 36 do 40 riječi, a nakon toga je kvaliteta značajno opala (ibid.).

U kontekstu istraživanja strojnog prevođenja sa švedsko-hrvatskim jezičnim parom, valja napomenuti kako u trenutku pisanja ovog rada GP-ov neuronski model još ne podržava izravno prevođenje između tih jezika, nego zasad samo podržava kombinacije s engleskim jezikom (SV ↔ EN i HR ↔ EN (Wikipedia 2018)). U prijevodima s manjih jezika GP se tradicionalno služio engleskim jezikom kao jezikom posrednikom, a neuronski modeli danas stvaraju vlastite „kriptične“ međujezike kojima se služe pri prevođenju (Hurskainen 2018). Dodatan je problem što „i neuronski modeli, kao i statistički, uvelike ovise o dostupnosti većih količina podataka i poznato je da oba pristupa daju loše rezultate kada resursa nema dovoljno“ (Ramesh i Sankaranarayanan 2018: 1). I hrvatski se jezik ubraja u one

jezike za koje još nema dovoljno tekstualnih materijala i paralelnih korpusa na kojima bi sustav trenirao, osobito u kombinaciji s drugim manjim jezikom poput švedskog. U nastavku ovog rada provjerit će se je li svejedno došlo do nekih pomaka po pitanju kvalitete prijevoda u ovoj jezičnoj kombinaciji, s obzirom na to da neki istraživači (npr. Östling i Tiedemann 2017: 1) još uvijek smatraju da statistički modeli dalju bolje rezultate od neuronskih kada je riječ o jezicima kod kojih je dostupno manje resursa.

2. Cilj i metodologija

Cilj je rada usporediti nalaze dobivene analizom statističkog modela GP-a iz studenoga 2016. godine s rezultatima dobivenima analizom neuronskog modela koji je na snazi u vrijeme pisanja ovog rada, u travnju 2018. godine. Oba je puta analiziran isti korpus koji sadrži 80 kartica tekstova različitih funkcionalnih stilova (tehničkih, političkih, administrativnih i znanstvenih tekstova¹) kako bi se dobio uvid u eventualno poboljšanje u izlaznim podacima. U prvoj fazi istraživanja pogreške su bile klasificirane u deset kategorija: neprevedene, izostavljene i umetnute riječi, leksičke, pravopisne, morfosintaktičke i stilističke pogreške, pogreške u redu riječi i dosljednosti te nejasne rečenice (za detaljniji opis pojedinih vrsta pogrešaka v. Ljubas 2017). Analizom tih podataka uočeno je da se određene vrste pogrešaka često ponavljaju iako bi se mogle eliminirati već i manjim preinakama u algoritmu. U poglavlju 3. bit će prikazano o kojim je pogreškama riječ jer je upravo na njih stavljen naglasak i u drugoj fazi istraživanja. U rezultatima će, međutim, biti prikazani i neki novi primjeri jer je ponovno proučen cijeli korpus, a na njime će se prikazati koliko se neuronski model GP-a razlikuje od statističkog.

¹ Tekstovi su dostupni u odgovarajućim diplomskim radovima studenata Katedre za skandinavistiku pri Sveučilištu u Zagrebu koji se mogu pronaći u repozitoriju Filozofskog fakulteta u Zagrebu (<http://darhiv.ffzg.unizg.hr>)

3. Rezultati

U primjerima koji slijede prvo se u zagradama navodi švedski izvornik, zatim, također u zagradama, prijevod na hrvatski dobiven 2016. godine te naposljetku izlazni podaci dobiveni u travnju 2018. godine. Analizirane pogreške uključuju dva obilježja švedskog jezika: uporabu švedskog formalnog subjekta „det“ kojim se često uvode rečenice u švedskom jeziku (v. potpoglavlje 3.1) te švedske dvočlane glagole, u prijevodu kojih je važno da se glagol i čestica gledaju kao cjelina, a ne da se traže zasebni prijevodni ekvivalenti za njihova dva elementa (potpoglavlje 3.2). Također sam se usredotočila na dva obilježja hrvatskog jezika: povratne zamjenice (potpoglavlje 3.3) i fenomen dvostruke negacije koji švedski jezik ne poznaje (potpoglavlje 3.4). Nadalje je naglasak bio na antonimnim prijevodnim ekvivalentima koje je GP često nudio za svakodnevne, jednostavne riječi koje zacijelo postoje u njegovom korpusu (potpoglavlje 3.5). GP je ranije često griješio i u prevođenju kratica (potpoglavlje 3.6) te složenica (potpoglavlje 3.7). Istaknute su još i pravopisne pogreške, osobito nemotivirano umetanje interpunkcijskih znakova koji mogu promijeniti značenje cijele rečenice (potpoglavlje 3.8) te naposljetku neprevedene riječi, koje je bilo osobito zanimljivo proučiti u svjetlu navoda iz dosadašnjih istraživanja koja su pokazala da se u neuronskom modelu javlja veći broj neprevedenih riječi nego u prijašnjem, statističkom modelu (potpoglavlje 3.9).

3.1 Švedski formalni subjekt „det“

U primjeru 1. formalni je subjekt preveden odgovarajućim pasivnim oblikom. Zanimljivo je da GP nije samo prilagodio oblik glagola, već je i odabrao potpuno drugi glagol („izrađuju se“ umjesto „proizvode se“). U ovom su prijevodu sada svi njegovi dijelovi morfosintaktički ispravni, no smisao se rečenice u cjelini izgubio. Do toga je ponajprije dovela sintagma „prijelazni signal“ koja je odabrana kao ekvivalent švedskom „övergångssignalen“ i nije jasno da je zapravo riječ o *zvučnim signalima na pješačkim prijelazima*.

Primjer 1. (**Det tillverkas** fyra olika modeller av den akustiska Prisma TS övergångssignalen.)
(**To proizvedene** četiri različita modela akustične Prisma TS prijelaznog signala.)
Izrađuju se četiri različita modela akustičnog Prisma TS prijelaznog signala.

I u primjeru 2. i 3. formalni su subjekti dobro prevedeni. U primjeru 2. zabilježene su manje stilističke pogreške, ali nema leksičkih ni morfosintaktičkih pogrešaka.

Primjer 2. (Vid dessa försäljningar **är det du** som säljare som ska redovisa och betala moms.)
(Na toj prodaji, **to je vas** kao prodavatelj prijavljivanja i plaćanja PDV-a.)
U toj prodaji, **vi ste** prodavatelj koji će prijaviti i platiti PDV.

U primjeru 3. zabilježeno je tek jedno morfosintaktičko odstupanje („koje genetske promjene došlo“), ali se pojavila i jedna leksička pogreška koje ranije u prijevodu nije bilo: Trinidad nije prepoznat kao toponim, već je preveden kao da se radi o nečijem osobnom imenu, što dovodi do pogrešne interpretacije teksta.

Primjer 3. (**Det går inte** att säga exakt vilka genetiska förändringar som skedde med guppyfiskarna i samband med försöken på Trinidad, ...)
(**To nije moguće** točno reći što se genetske promjene koje su se dogodile s Guppy riba u vezi s pokušajima na Trinidad, ...)
Nije moguće točno odrediti koje genetske promjene došlo u Trinidadovim pokusima,
...

3.2 Švedski dvočlani glagoli

U švedskom jeziku relativno se malim brojem čestica mogu drastično promijeniti značenja velikog broja glagola. Sustav mora biti istreniran na obuhvatnom korpusu tekstova kako bi prepoznao dvočlane glagole i mora dobro povezivati elemente jer se glagol i čestica ne moraju, naime, uvijek nalaziti u neposrednoj blizini. Upitno je jesu li izlazni podaci u primjeru 4. sada bolji ili lošiji nego ranije. Prijevod na prvi pogled izgleda točno i funkcionira u morfosintaktičkom pogledu, ali leksička se pogreška zadržala, iako glagol i čestica nisu zasebno prevedeni (glagol „komma till“ zapravo znači „postati“, „nastati“). Pogreška „došli do“ ranije je čitatelju mogla signalizirati da se radi o nekom problematičnom elementu, a sada bi čitatelj ili, primjerice, redaktor strojnog prijevoda, pri iščitavanju teksta ovu rečenicu lako mogao previdjeti i percipirati je kao ispravno prevedenu.

Primjer 4. (Vi människor har under alla tider funderat över hur vi har **kommit till**.)
(Mi ljudi imamo u svakom trenutku razmišljao o tome kako smo **došli do**.)
Uvijek smo razmišljali o tome kako smo **došli**.

I u 5. je primjeru GP prepoznao dvočlani glagol te se približio njegovu značenju, a i cijela rečenična struktura sada bolje funkcionira: prvi prijevod vrvi od leksičkih, morfosintaktičkih i pravopisnih pogreški, a umetnuta je i riječ „ga“ koje u izvorniku nema te je izostavljena riječ „senast“ (u značenju „najkasnije“). „Senast“ nedostaje i u novom prijevodu, ali ipak se može zamijetiti određeni pomak u prijevodu ove rečenice.

Primjer 5. (Senast år 2005 finns sektorsanpassad information som gör det möjligt **att välja bort** byggvaror/byggkonstruktioner som innehåller...)
 (Do 2005. godine sektor specifične informacije, što ga čini moguće **odabrati** građevinskih materijala / Građevinski programe koji sadrže...)
 Do 2005. godine bit će dostupni sektorski specifični podaci koji će omogućiti **uklanjanje** građevinskih materijala / zgrada koje sadrže...

U primjeru 6. GP je u objema varijantama prijevoda pogriješio pri prijevodu glagola „dra ihop sig“ koji znači „zgusnuti se“. Ipak, u drugoj se verziji zamjećuje veliko poboljšanje u odnosu na prvi GP-ov pokušaj koji je bio mnogo nerazumljiviji, a na mjestu spornog glagola pojavila se imenica „ugovor“.

Primjer 6. (När ett gasmoln börjar **dra ihop sig** genom sin egen tyngd är det viktigt att temperaturen i molnet inte ökar...)
 (Kada je plinski oblak počinje **ugovor** pod vlastitom težinom, važno je da je temperatura oblaka ne raste...)
 Kada se oblak plina počne **udvarati** vlastitom težinom, važno je da se temperatura u oblaku ne povećava...

3.3 Povratne zamjenice

Iako i švedski i hrvatski jezik imaju povratne glagole, oni nisu u istom omjeru zastupljeni među jezicima. GP je u primjeru 7. otpočeka leksički i bio na dobrom tragu u prijevodu glagola „ses“, ali da bi prijevod bio u skladu s hrvatskom gramatikom, za ovakvu je bezličnu uporabu glagola bilo potrebno „dodati“ povratnu zamjenicu, što je učinjeno u novoj inačici prijevoda. Zanimljivo je, međutim, primijetiti kako je u prvoj verziji ispravno bila prevedena riječ „faktiskt“ („zapravo“), koja je u novom prijevodu u potpunosti izostavljena. Semantički gledano tu nije izgubljena neka ključna informacija, a novi je prijevod i u

morfosintaktičkom smislu mnogo bolji. Ipak ovaj primjer govori u prilog navodima o tome da neuronski modeli češće izostavljaju riječi i da su prijevodi neprecizniji.

Primjer 7. (Nu har forskarna skapat en katalog med exempel på hur snabbt effekten av naturlig selektion faktiskt **kan ses**.)

(Znanstvenici su stvorili katalog primjera koliko brzo djelovanje prirodne selekcije zapravo **može vidjeti**.)

Sada su istraživači stvorili katalog primjera kako **se** brzo **može vidjeti** učinak prirodne selekcije.

U sljedećem se primjeru povratna zamjenica javlja i u švedskom izvorniku, a u hrvatskom bi jeziku također odgovarao povratni glagol „(pro)širiti se“. U novoj se verziji ta pogreška više ne javlja, a deklinirani su i toponimi, što 2016. nije bio slučaj. Ovaj primjer pokazuje da je GP napravio velik korak naprijed kada su u pitanju morfološka obilježja hrvatskog jezika i njihovo pravilno usklađivanje.

Primjer 8. (Sedan i början av året **har** ebola **spritt sig** från Guinea till Sierra Leone, Liberia och Nigeria.)

(Od početka godine, ebola **širi** iz Gvineje do Sierra Leone, Liberija i Nigerija.)

Od početka godine, ebola **se proširila** od Gvineje do Sierra Leonea, Liberije i Nigerije.

U primjeru 9. GP je umetnuo povratnu zamjenicu „se“ na mjesto gdje ona nije postojala u švedskom tekstu, a nije bila potrebna ni u hrvatskom tekstu. Ta se pogreška ne ponavlja u novoj verziji koja je ponovno morfološki mnogo pravilnija, ali opet i ponešto nepreciznija u odnosu na prvi prijevod: tako se, primjerice, umjesto „prijelaza na održivije društvo“ sada omogućuje samo „održivo društvo“.

Primjer 9. (Ny teknik och användning av ekologiskt förnyelsebara produkter **kan möjliggöra** en omställning till ett mer hållbart samhälle.)

(Nove tehnologije i korištenje ekološki obnovljivih proizvodi **mogu se omogućiti** prijelaz na održivije društvo.)

Nova tehnologija i uporaba ekološki obnovljivih proizvoda **mogu omogućiti** održivo društvo.

3.4 Dvostruke negacije

Dvostruke negacije često su problematične u strojnim prijevodima sa švedskoga na hrvatski jezik. U tom je pogledu zabilježena pozitivna promjena (primjeri 10. i 11.).

Primjer 10. (**Inget steg** framåt **får** avfärdas som alltför obetydligt.)
 (**Bez korak** naprijed **treba** odbaciti kao previše beznačajan.)
Ni jedan korak naprijed **ne može** biti odbačen kao premalen.

No u primjeru 11., uz leksičku pogrešku „u dugom roku“ koja nije ekvivalent švedskom „i längden“, „dugotrajno“ (što je vjerojatno rezultat interferencije engleskog jezika i doslovan prijevod sintagme „in the long run“), zamijećena je i jedna pravopisna pogreška: „iu“ je napisano zajedno. Ta se pogreška nije pojavila u prvoj inačici prijevoda ove konkretne rečenice, ali jest bila često zabilježena na drugim mjestima, što upućuje na to da se ta pogreška vjerojatno pojavljivala u tekstovima na kojima je sustav treniran.

Primjer 11. (Trots att dessa rättigheter ständigt förvägras människor över hela världen – även i vårt land – **kan ingen** regim i längden ignorera dem.)
 (Iako su ta prava stalno se uskraćuje ljudima diljem svijeta - čak i u našoj zemlji - **može ni** režim u dužini ih ignorirati.)
 Iako su ta prava neprestano uskraćena ljudima diljem svijeta - čak i u našoj zemlji - **nikakav** režim u dugom roku **ne može** ih ignorirati.

3.5 Suprotni prijevodi

U ranijim su se prijevodima jednostavne riječi poput „sällan“ („rijetko“) ili „yngre“ („mlađi“) znale pretvoriti u svoje antonime („često“; „kasni“). To je u primjerima 12. do 15. razriješeno, ali zato se zamjećuju druge poteškoće.

Iako je novi prijevod primjera 12. sintaktički ispravniji, nije lako dokučiti značenje sintagme „da je rijetko bilo teško ponuditi imigrantski rad“ i iščitati da se zapravo radi o tome da je rijetko bilo problematično ponuditi useljenicima posao, a i ostatak je rečenice neprecizan (tako primjerice švedski „dessutom“ zapravo znači „štoviše“ i najbolje bi odgovarao na početku rečenice kao tekstni konektor).

Primjer 12. (Tillgången på arbetstillfällen var dessutom så god att det **sällan** förelåg svårigheter att erbjuda invandrare arbete.)
 (Dostupnost poslova je dodatno toliko dobar da je **često** bilo teško ponuditi imigranata posao.)
 Dostupnost radnih mjesta također je bila tako dobra da je **rijetko** bilo teško ponuditi imigrantski rad.

I u primjeru 13. GP je ponudio morfosintaktički točniji, ali neprecizniji prijevod. Ranije je prijevod informirao čitatelja o tome da se potrošnja energije mogla prepoloviti „u posljednjih 50 godina“ (umjesto da se može prepoloviti „u sljedećih 50 godina“), ali novi je završetak rečenice, „može spustiti u 50 godina“, ostao nedorečen. Samim time što se švedski „halveras“ iz pravilnog leksičkog odabira „prepoloviti“ pretvorio u „spustiti“, informacije iz izvornika nisu vjerno prenesene.

Primjer 13. (Sammantaget bedömer man att energianvändningen i bebyggelsen kan halveras **på 50 års sikt.**)

(Sve u svemu, procjenjuje se da je potrošnja energije u zgradama može biti prepolovljen **u posljednjih 50 godina.**)

Sveukupno, procjenjuje se da se potrošnja energije u stambenom prostoru može spustiti **u 50 godina.**

Sličnu situaciju vidimo i u primjeru 14. gdje je suprotan prijevod razriješen, ali ostatak rečenice nikako ne funkcionira. Leksička pogreška koja najviše otežava razumijevanje očituje se u sintagmi „olupina mora“ koja nema nikakvo značenje, a kamoli značenje „morske uvale“. Kratica „e Kr“ („poslije Krista“) nije prevedena, no olakotna je okolnost to što je moguće zamisliti da ni čovjek ne bi nužno preveo tu kraticu u ovom kontekstu jer i samo „god.“ prenosi točno značenje (više o kraticama u potpoglavlju 3.6). Uz navođenje godina često su zabilježene i pravopisne pogreške jer iza njih nedostaju točke, ali se javlja i problem percepcije kraja i početka rečenice kada su u pitanju brojevi i točke. Čini se da je i u ovom primjeru GP točku iza brojke prepoznao kao kraj rečenice pa je sljedeću riječ napisao velikim slovom. Zabilježen je veći broj takvih situacija.

U primjeru 14. zabilježen je i neobičan detalj apozicije „Jezero“ koje u izvorniku nema. GP ju je doduše bezrazložno napisao velikim slovom, ali ju je prvi put uveo kako bi čitatelju objasnio što je to Mälaren, a u novoj je verziji taj pokušaj izostao. Ovdje bi također valjalo usporediti rezultate statističkog i neuronskog modela jer se radi o dužem primjeru, ali nažalost nijedan od ovih prijevoda nije bio dovoljno uspješan.

Primjer 14. (När den havsvik som skulle forma Mälaren under **yngre** järnålder (cirka 550–1050 e Kr) genom landhöjningen skars av från Östersjön och bildade en insjö blev kontrollen av de få in- och utfartslederna av central strategisk betydelse.)

(Kada je ulazni koja će oblikovati Jezero Malaren tijekom **kasnog** željeznog doba (oko 550-1050 AD) je izdizanje bio odsječen od Baltičkog mora i formirao jezero je kontrola od rijetkih ulaznih i izlaznih putova od strateške važnosti.)

Kada je olupina mora, koja je formirala Mälaren tijekom **mlađeg** željeznog doba (otprilike 550-1050 god.) Kroz porast zemljišta, prekinuta s Baltičkog mora i formirala jezero, kontrola nekoliko ulaza i izlaza bilo je od središnje strateške važnosti.

3.6 Kratice

Teškoće koje je GP imao s kraticama očitovale su se ili u neprepoznavanju kratice ili u tome što kratice ne bi deklinirao. Primjer 15. također je povezan i s utjecajem engleskog (među)jezika na prijevode sa švedskog na hrvatski jer je GP kraticu MOSS doživio kao engl. „moss“ („mahovina“), iako se nalazi neposredno iza raspisanog naziva. Kako se ovdje radi o imenu e-usluge švedske Porezne uprave, naziv je moguće samo ostaviti na engleskom s pripadajućom kraticom, što je GP sada i napravio. Drugu kraticu u rečenici („s. k.“) ranije je ispravno preveo kao „takozvani“, a sada ju je zamijenio lošijim odabirom „tj“ i to s pravopisnom pogreškom jer nije dodao točku iza kratice.

Primjer 15. (Det kan antingen göras genom att du momsregistrerar dig i det landet eller att du använder dig av Skatteverkets e-tjänst, den s.k. Mini One Stop Shop (**MOSS**).) (To može ili biti učinjeno po vama registrirati za PDV u toj zemlji ili da koristite e-usluga Porezne Agencije, takozvani Mini Stop Shop Jedan (**mahovina**).)

To se može učiniti registriranjem PDV-a u toj zemlji ili korištenjem e-servisa Porezne uprave, tj Mini One Stop Shop (**MOSS**).

Kada je riječ o kraticama, GP je „popravio“ izlazne podatke samo u primjeru 15. U primjeru 16. kratice „AB“ („d.d.“) i dalje nije prepoznata ni prevedena. Kako se ovdje radi o imenu firme i nema šireg konteksta, čini da je „AB“ samo shvaćeno kao nastavak imena koje je u cijelosti ostavljeno u istom obliku.

Primjer 16. (© 2004 Prisma Teknik **AB**)
(© 2004 Prisma Teknik **AB**)
© 2004 Prisma Teknik **AB**

U primjeru 17. GP i dalje ne sklanja (a vjerojatno, kao ni u posljednjem primjeru, ne prepoznaje) kraticu „ESA“. Prijevod je i dalje nerazumljiv jer je GP preveo „Sveriges rymdsatsningar“, što se odnosi na „švedska ulaganja u svemir“,



kao da se radi o imenu npr. neke organizacije. Pritom se „Sveriges“ samo odnosi na Švedsku kao državu te je, u skladu sa švedskim pravopisom, ta riječ i napisana velikim slovom. Ako u hrvatskom jeziku govorimo o „švedskim investicijama“, rabimo pridjev koji se po pravilima pravopisa piše malim slovom. Situacija bi, međutim, bila drukčija ako bismo govorili o „investicijama Švedske“ kada ćemo državu ipak pisati velikim slovom.

Primjer 17. (Den allra största delen av Sveriges rymdsatsningar, ca 65 %, går till **ESA**.)
(Velika većina švedskih prostora naporima, oko 65%, ide na **ESA**.)
Velika većina Švedske svemirske investicije, oko 65%, ide u **ESA**.

3.7 Složenice

Kako se složenice mnogo češće javljaju u švedskom negoli u hrvatskom jeziku, njihovo prevođenje često zahtijeva drukčija sintaktička rješenja. U primjerima 18. do 20. vidimo kako je GP griješio u njihovom prevođenju u prijašnjim verzijama, ali i različite pokušaje „popravljanja“ manjkavih prijevoda. U primjeru 18. to je bilo uspješno: „livsbetingelser“ zbilja jesu „životni uvjeti“. U ostatku su se rečenice jedino „ekološki ciklusi“ našli u krivom padežu jer se zapravo radi o „zaštiti prirode i ekoloških ciklusa“.

Primjer 18. (Genom att värna naturens och de ekologiska kretsloppen kan människorna underhålla och utveckla sina egna **livsbetingelser**.)
(Do očuvanje prirode i ekološke ciklusi mogu ljudi održavati i razvijati vlastiti **život**.)
Zaštitom prirode i ekološkim ciklusima ljudi mogu održavati i razvijati vlastite **životne uvjete**.

I u primjeru 19. GP je bio na dobrom tragu što se tiče prevođenja složenice „ljuskaraktär“ sintagmom koja se sastoji od pridjeva i imenice, ali nažalost iz pogrešne perspektive. Ne radi se o „karakterističnoj svjetlosti“, već o „svjetlosnim karakteristikama“ zbog čega u ovom segmentu još uvijek nedostaje informacija koja je ključna za njegovo razumijevanje.

Primjer 19. (**Ljuskaraktär**: Gult sken, ca. 590 nm våglängd)
(**Svjetlo Karakter**: žuto svjetlo, ca. 590 nm valne duljine)
Karakteristična svjetlost: žuta sjaja, cca. 590 nm valne duljine

U primjeru 20. složenica je sustavu za strojno prevođenje još teže prepoznatljiva. Njezini su dijelovi odvojeni spojnicom, što je uobičajeno u švedskom jeziku. Sintagma „klass-, köns- och etniskt förtryck” zapravo se sastoji od dviju složenica: „klassförtryck” („klasno ugnjetavanje”) i „könsförtryck” („spolno ugnjetavanje”), te od pridjevsko-imenične sintagme „etniskt förtryck” („etničko ugnjetavanje”). U hrvatskom bi se to elegantno moglo razriješiti nizom pridjeva, primjerice, da govorimo o „društvu bez klasnog, spolnog i etničkog ugnjetavanja”, ali GP ipak nije uspio u tome da poveže „klasu” i „spol” s riječju „förtryck” (neovisno o tome koji bi točno leksički ekvivalent odabrao za nju), već prijevod glasi „društvu bez klase, spola i etničkog tlačenja”.

Odabir riječi „tlačenje” ulazilo bi u domenu stilističkih pogrešaka, a našlo se i nešto morfosintaktičkih pogrešaka u drugom dijelu rečenice („ostvariti društvu”, „žene i muškarci gradi”, „u slobodi i suradnju”). Iako bi te morfosintaktičke pogreške mogle ići u prilog tezi da neuronski model ima većih poteškoća s dužim rečenicama poput ove, ni rješenja koja je ponudio statistički model nisu bila bolja.

Primjer 20. (Partiets politik och verksamhet bestäms av vårt mål: att förverkliga ett samhälle grundat på demokrati, jämlikhet och solidaritet, ett samhälle befriat från **klass-, köns- och etniskt förtryck**, ett rättvist och ekologiskt hållbart samhälle där kvinnor och män bygger sin egen framtid i frihet och samverkan.)
(politika i aktivnosti stranke određuju se našem cilju ostvarivanja društva temeljenog na demokraciji, jednakosti i solidarnosti, društva slobodnog od **klase, spola i etničke represije**, pravedan i ekološki održivo društvo u kojem žene i muškarci grade vlastitu budućnost u slobodi i suradnja.)
Politike i aktivnosti stranke određene su našim ciljem: ostvariti društvo utemeljeno na demokraciji, jednakosti i solidarnosti, društvu bez **klase, spola i etničkog tlačenja**, pošteno i ekološki održivo društvo u kojem žene i muškarci gradi svoju budućnost u slobodi i suradnju.

3.8 Pravopisne pogreške

Jasno je kako na (ne)razumljivost strojnog prijevoda pravopisne pogreške utječu manje od leksičkih pa čak i morfosintaktičkih pogrešaka. Ipak, neopravdano umetanje ortografskih znakova poput zareza itekako može utjecati na pogrešno razumijevanje teksta. Također se općenito smatra da pravopis ne bi trebao

predstavljati (pre)velik problem za strojnoprevoditeljske sustave jer su trenirani na ljudskim prijevodima koji bi trebali biti očišćeni od takvih pogrešaka, no korpusi često obuhvaćaju i nekvalitetne ljudske prijevode. Pri pripremi korpusa za treniranje sustava također može doći i do pogreški u sravnjivanju fraza, a tako dolazi i do umetanja nepotrebnih interpunkcijskih znakova (kao u primjeru 21.) ili pisanja velikim slovom (v. primjer 24.). Ranije se to učestalo događalo, ali u novim prijevodima nije pronađen nijedan slučaj umetanja nepotrebnih znakova.

Primjer 21. (Fattiga länders möjlighet **till** utveckling undergrävs genom att de högutbildade dras till de stora företagen i västvärlden.)

(Potencijal siromašne zemlje **'za'** razvoj potkopava visokoobrazovanih privučeni velikim tvrtkama na Zapadu.)

Razvojni potencijal siromašnih zemalja potkopava se privlačenjem visokoobrazovanih na velike tvrtke u zapadnom svijetu.

To ne znači da ne dolazi do pravopisnih pogrešaka druge vrste. Primjer 22. prikazuje slučaj dužeg navoda unutar teksta. Čini se da je GP 'zaboravio' na prve znakove navodnika, pa je druge navodne znakove ponovno protumačio kao otvaranje nekog novog segmenta. Ranije nije napravio istu pogrešku.

I ovaj je segment dugačak, a pogreške su brojne. Osobito se ističe leksička pogreška „*sigurna, sigurna i ekološki prihvatljiva*“. Iako se „trygg“ i „säker“ mogu promatrati kao sinonimi u švedskom jeziku kojima je denotativno značenje „sigurno“, čovjek bi lako razlučio da se ista riječ ne može dvaput uzastopno rabiti u prijevodu. Potrebno bi bilo pronaći, primjerice, neki sinonim, ali GP na tom mjestu još uvijek griješi.

Primjer 22. (Den ovan nämnda visionen och de uppställda effektmålen är baserade på en målbild som säger att: "Det svenska samhället har till stor del utvecklats i enlighet med visionen om ekonomiskt, ekologiskt och socialt hållbar utveckling. Trygg, säker och miljövänlig energi tillhandahålls till konkurrenskraftiga priser. År 2050 klarar Sverige målet att: Utsläppen av växthusgaser i Sverige är lägre än 4,5 ton koldioxidekvivalenter per år och innevånare, och minskande." (Prop. 2004/05:150).) (Vizija i utvrđeni ciljevi učinkovitosti temelji se na viziji koja kaže: "Švedsko društvo je u velikoj mjeri razvijena u skladu s vizijom ekonomski, ekološki i socijalno održivi razvoj. Siguran, siguran i ekološki prihvatljiv energija dostupna po konkurentnim

cijenama. 2050 će Švedska cilj. Emisija stakleničkih plinova u Švedskoj je niži od 4,5 tona ugljičnog dioksida godišnje i stanovnika, te smanjuje "(Prop 2004/05. 150). Gore vizija i utvrđeni ciljevi učinkovitosti temelji se na viziji koja kaže: „Švedska je društvo u velikoj mjeri razvio u skladu s vizijom ekonomski, ekološki i socijalno održivi razvoj. Sigurna, sigurna i ekološki prihvatljiva energija pruža se po konkurentnim cijenama. U 2050. Švedska upravlja cilj: emisije stakleničkih plinova u Švedskoj je manja od 4,5 tona ugljičnog dioksida godišnje, a stanovnici i opada. „(Prop 2004/05. 150).

Kada govorimo o navodnim znakovima, u hrvatskom je jeziku znakove otvaranja navoda ispravno i uobičajeno postaviti „dolje“, ali GP ih često stavlja „gore“, jednako kao i znakove zatvaranja navoda. U primjeru 22. GP je pravilno postavio navodnike, ali u korpusu je zabilježeno i kako GP često griješi u mjestu postavljanja navodnika (v. primjer 23). Ostale se pravopisne pogreške u primjeru 23. očituju i u opetovanoj uporabi spojica koje se nalaze na mjestu gdje se u izvorniku nalaze crtice, a zapravo i dalje nose ulogu crtica s obzirom na to da su od susjednih riječi odvojene bjelinama.

Primjer 23. (Dagens nya fas av aggressiv imperialism – "kriget mot terrorismen" –återskapar på ett handfast sätt ett rasistiskt tänkande.)
(Današnja nova faza agresivnog imperijalizma - "rat protiv terora" - kreira na konkretan način rasist razmišljanja.)
Današnja nova faza agresivnog imperijalizma - "rat protiv terorizma" - jasno rekreira rasističko razmišljanje.

U većini je slučajeva uklonjeno neopravdano umetanje velikih slova ondje gdje ona nisu potrebna, kao što vidimo u primjeru 24.

Primjer 24. (På de nästföljande fyra mötena bedömdes de fem **bedömningsområdena** ett i taget (på första mötet bedömdes två områden) utifrån kraven i programmet och juryns synpunkter sammanställdes.)
(U sljedeća četiri susreta procjenjuje na pet **područja OCJENA** jedan po jedan (na prvi sastanak je ocijenjen dva područja) koji se temelji na zahtjevima programa i komentari žirija bili sastavio.)
Na sljedeća četiri sastanka, pet **područja ocjenjivanja** procjenjivana su jedna po jedna (na prvom sastanku dva su područja ocijenjena) temeljem zahtjeva programa i sastavljenih mišljenja žirija.

Među pravopisnim pogreškama osobito se ističe umetanje zareza. Švedski daleko rjeđe rabi zareze od hrvatskog pa umetanje katkad može biti i poželjno. No u primjeru 25. umetanje je zareza negativno utjecalo na razumijevanje rečenice jer je njime razdvojena cjelina koja funkcionira samo kao cjelovita („vad samhället ska använda tillväxten till“, „za što će društvo koristiti rast“). Zarezi su osjetljiva značajka prijevoda jer ih čitatelj nesvjesno percipira kao logičan dio teksta, dok nemotivirano umetanje npr. navodnih znakova jednostavnije može zanemariti. Iako se novi algoritam i u ovom pitanju bolje iskazao, u istoj je sintagmi došlo do leksičke pogreške koja je ipak presudila o (ne)uspješnosti ovog prijevoda: ne radi se o pitanju „koje društvo će“ već „za što će društvo“ koristiti rast.

Primjer 25. (Frågan om tillväxt måste kombineras med frågan om vad samhället ska **använda tillväxten** till och på vilket sätt tillväxten skapas.)

(Pitanje rast mora biti u kombinaciji s pitanjem o tome što društvo treba **koristiti, rastu** i kako se stvara rast.)

Pitanje rasta mora se kombinirati s pitanjem koje će društvo **koristiti rast** i na koji način se stvara rast.

3.9 *Neprevedene riječi*

Kao što je to spomenuto u uvodu (poglavlje 1.), smatra se da neuronski modeli češće izbacuju ili neprevedenima ostavljaju dijelove teksta (Wu i sur. 2016: 2). Izostavljene su riječi i dosad predstavljale veliki problem u prijevodima GP-a, a neprevedene tek djelomično. Često je GP, naime, zastao na međukoraku i riječi ostavio prevedenima na engleski jezik. Ako govorimo o prosječnom govorniku hrvatskog jezika koji se poslužio GP-om kako bi razumio određeni tekst, mnogo je vjerojatnije da će razumjeti englesku riječ, nego da je ona ostala u potpunosti neprevedena, odnosno da se u tekstu pojavila na švedskom jeziku. Naravno, u redakтури prijevoda takve se pogreške svakako moraju eliminirati.

U korpusu su uočeni najrazličitiji oblici pojavljivanja neprevedenih riječi zbog kojih se nekada čini da GP nazaduje u koracima, a nekada da ipak napreduje. U 26. primjeru GP je sada cijelu riječ preveo na hrvatski, ali u primjeru 27. s engleskog se prijevoda ponovno vratio na švedski, a u primjeru 28. je čak riječ koju je ranije dobro preveo na hrvatski ponovno ostavio neprevedenu na švedskom jeziku.

- Primjer 26. (Klassernas sammansättning och **självuppfattning** ändras med den ekonomiska utvecklingen.)
(Sastav klase i **self-percepcija** mijenja s gospodarskog razvoja.)
Sastav nastave i **samopoštovanje** se mijenjaju s gospodarskim razvojem.
- Primjer 27. (**Inom** angöring/parkering finns risk för problem när många föräldrar lämnar samtidigt som det kommer leveranser.)
(**In** slijetanje / parking je rizik od problema kada mnogi roditelji napuste dok će dostaviti.)
Inom ribolov / parkiranje postoji rizik od problema kada mnogi roditelji napuste dok se isporuke dogode.
- Primjer 28. (**Rasismen** är ensam vare sig orsak till eller verkan av nykolonisering och imperialistiska krig.)
(**Rasizam** je sama ili uzrok ili posljedica kolonizacije i imperijalističkog rata.)
Samo **Rasismen** nije ni uzrok niti posljedica nove kolonizacije i imperijalističkog rata.

U primjeru 28. zanimljivo je i da je GP riječ „Rasismen“ koju je ipak ostavio na švedskom, također napisao velikim slovom kao da se radi o početku rečenice. To zaista jest početak švedske rečenice, ali u prijevodu je ispred „Rasismen“ legitimno umetnuta riječ „Samo“ koja je također napisana velikim slovom.

Mnogo je primjera neprevedenih riječi u korpusu, a i dalje se mnoge javljaju na engleskom. U primjeru 29., koji je i inače u potpunosti nejasno preveden, javljaju se dvije riječi na engleskom za koje je teško zaključiti zašto ih je GP odabrao i što bi one u ovom kontekstu uopće mogle predstavljati.

- Primjer 29. (SEPARAT PROTOKOLL: BRYTNING AV NAMNSEDLAR)
(Poseban protokol refrakcije NAZIV novčanica)
ODVOJENI PROTOKOL: BRIDGING NAMES

Za kraj je ostavljen primjer u kojemu je GP stavljen pred izazov prevođenja kazala. U njemu se četiri puta ponavlja fraza „sid 2/ sid 3“ („stranica 2/ stranica 3“), a GP je od toga dvaput „stranicu“ preveo odgovarajuće, jednom je ostavio neprevedenu na švedskom, a jednom umetnuo na drugo mjesto. U novom je prijevodu ipak uspješno preveo „stranicu“ na svim mjestima.

Primjer 30.	(Generell beskrivningsid 2
	Sprängskiss sid 2
	Installation.....sid 3
	Kabelanslutning sid 3)
	(Opći opis-..... stranica 2
	Eksplodirala sid 2
	Instalacija stranica 3
	Kabelski priključak stranica 3)
	Opći opis stranica 2
	Eksplodirani prikaz stranica 2
	Instalacija stranica 3
	Kabelska veza stranica 3

4. Zaključak

Cilj je ovog rada bio provjeriti koliko su promjene koje su se dogodile u algoritmima Google prevoditelja u razdoblju od studenog 2016. do travnja 2018. godine utjecale na kvalitetu prijevoda sa švedskoga na hrvatski jezik. Google je u tom periodu uveo veliku promjenu prelaska sa statističkog modela prevođenja na neuronski, ali taj novi pristup još ne podržava sve jezične kombinacije, već je samo dostupan za prevođenje u kombinacijama s engleskim jezikom. Nametnulo se pitanje hoće li promjene u Googleovu algoritmu svejedno utjecati na prijevode i u drugim jezičnim kombinacijama. Prethodno istraživanje prijevoda sa švedskoga na hrvatski (v. Ljubas 2017) pokazalo je poražavajuće rezultate s izlaznim podacima koji su vrvjeli od najrazličitijih pogrešaka, dok su prijevodi s engleskoga na hrvatski jezik bili nešto bolji (v. Pavlović 2017).

Rezultati ovog istraživanja u načelu su u skladu s nadama u to da bi neuronski pristup mogao riješiti mnoge poteškoće s kojima se statistički model teško nosio ako se usredotočimo na konkretne poteškoće koje su se dosada javljale u statističkom modelu. Bolji rezultati zabilježeni su u prijevodima formalnog subjekta „det“, povratnih zamjenica, dvostrukih negacija, složenica i suprotnih prijevoda. Više se ne pojavljuju ortografski znakovi koji su ranije bili umetnuti bez razloga, ali se i dalje pojavljuju raznovrsne pravopisne pogreške. Prijevodni ekvivalenti za dvočlane glagole i složenice djelomično su bolji. Kratice još uvijek nisu

zadovoljavajuće prevedene, a izlazni su podaci otkrili mnoge neprevedene riječi te veći broj izostavljenih riječi u odnosu na ranije prijevode. Upravo to često dovodi do nepreciznosti prijevoda koji na prvu i mogu djelovati ispravno, ali u njima ipak nedostaju određene informacije.

Iako se u izlaznim podacima više ne javljaju određene vrste pogrešaka, ne možemo jednostavno zaključiti da je neuronski model (koji trenutačno još nije ni dostupan za izravno prevođenje sa švedskog na hrvatski jezik) riješio sve probleme i da ulazimo u doba ispravnih, razumljivih i za objavljivanje spremnih strojnih prijevoda. Ako se odmaknemo od zacrtanog cilja i proučimo nove izlazne podatke iz šire perspektive, još uvijek će nas dočekati velik broj pogrešaka, nerijetko i potpuno nejasne rečenice. I dalje je velik broj pogrešnih leksičkih odabira, osobito kada se radi o stručnim terminima, a ni izostavljene riječi ne pridonose boljem razumijevanju. Možemo reći da je prijelaz na neuronski model korak u pravom smjeru: što se morfologije i sintakse tiče, vidimo da se GP u brojnim situacijama iskazao čak i u prijevodima na hrvatski jezik koji je morfološki veoma složen, no također ima i brojnih primjera gdje se zbog grubih morfosintaktičkih pogrešaka ne mogu dokučiti sintaktički odnosi u rečenici. Rezultati ukazuju na važnost pripreme obuhvatnijih korpusa na kojima bi se sustav trenirao. U kombinaciji sa švedskim i hrvatskim jezikom paralelni je broj korpusa još premalen i prije svega je potrebno raditi na povećanju broja dostupnih materijala kako bi sustav bio uspješniji pri odabiru ispravnih leksičkih ekvivalenata.

Zanimljivo je pitanje kako će Google nastaviti razvijati svoj novi model. Danas gotovo na dnevnoj bazi možemo proučavati pomake i promjene u izlaznim podacima i može se očekivati da će dosadašnja pozitivna iskustva s neuronskim modelom potaknuti proširenje na veći broj jezičnih kombinacija koje bi on podržavao. Također, kako su istraživanja pokazala da ključ dobrih izlaznih podataka leži u dobro istreniranim sustavima, daljnji bi se napori prvenstveno trebali usmjeriti na prikupljanje što većeg broja kvalitetnih paralelnih korpusa.

Literatura

Primarni izvori

Repozitorij Filozofskog fakulteta u Zagrebu. URL: <http://darhiv.ffzg.unizg.hr/>. Pristup: 18. 4. 2017.

Grubješić, Iva. 2015. Översättning av passiv från svenska till kroatiska: magisterarbete. Diplomski rad. Zagreb: Filozofski fakultet.

Hršak, Mirna. 2015. Översättning av kulturspecifika element i svenska tidningsartiklar: magisterarbete. Diplomski rad. Zagreb: Filozofski fakultet.

Matijanić, Anja. 2013. Översättning av socio-politiska kulturspecifika begrepp från svenska till kroatiska. Diplomski rad. Zagreb: Filozofski fakultet.

Mileta, Iliana. 2016. Några särdrag hos filmtextning från svenska till kroatiska. Diplomski rad. Zagreb: Filozofski fakultet.

Momirski, Lana. 2016. Översättning av arkitektonisk terminologi från svenska till kroatiska. Diplomski rad. Zagreb: Filozofski fakultet.

Nestić, Elizabeta. 2015. Översättning av institutionsnamn från svenska till kroatiska: magisterarbete. Diplomski rad. Zagreb: Filozofski fakultet.

Veršić, Vanja. 2014. Ortnamn i svensk-kroatisk översättning: magisterarbete. Diplomski rad. Zagreb: Filozofski fakultet.

Žic, Luka. 2014. Översättning av neologismer och verkspecifika termer. Diplomski rad. Zagreb: Filozofski fakultet.

Sekundarni izvori

Ahrenberg, Lars i Merkel, Magnus. 2006. „Maskinöversättning – Automatisk översättning”. URL: <http://sprakteknologi.se/vad-aer-sprakteknologi/maskinoeversaettning/automatisk>. Pristup: 16. 4. 2018.

Bentivogli, Luisa, Bisazza, Arianna, Cettolo, Mauro i Federico, Marcello. 2016. Neural versus Phrase-Based Machine Translation Quality: A Case Study. URL: <https://arxiv.org/abs/1608.04631>. Pristup: 22. 8. 2018.

- Castilho, Sheila, Moorkens, Joss, Gaspari, Federico et al. 2017. „Is Neural Machine Translation the New State of the Art?”. *The Prague Bulletin of Mathematical Linguistics* 108(1): 109-120. URL: <https://www.degruyter.com/downloadpdf/j/pralin.2017.108.issue-1/pralin-2017-0013/pralin-2017-0013.pdf>. Pristup: 8. 4. 2018.
- Hurskainen, Arvi. 2018. „Sustainable Language Technology for African Languages”. U: Agwuele, Augustine i Bodomu, Adams (ur.), *The Routledge Handbook of African Linguistics*. New York: Routledge Handbooks Online. URL: <https://www.routledgehandbooks.com/doi/10.4324/9781315392981-19>. Pristup: 16. 4. 2018.
- Ljubas, Sandra. 2017. „Analiza pogrešaka u strojnim prijevodima sa švedskog na hrvatski”. *Hieronymus - časopis za istraživanja prevođenja i terminologije* 4: 28-64.
- Östling, Robert i Tiedemann, Jörg. Neural machine translation for low-resource languages. URL: <https://arxiv.org/pdf/1708.05729.pdf>. Pristup: 22. 8. 2018.
- Pavlović, Nataša. 2017. „Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka”. U: Stolac, Diana i Vlastelić, Anastazija (ur.), *Jezik kao predmet proučavanja i jezik kao predmet poučavanja*. Zagreb: Srednja Europa. 279-295.
- Popović, Maja. 2015. „chrF: character n-gram F-score for automatic MT evaluation”. U: *Proceedings of the Tenth Workshop on Statistical Machine Translation*. 392-395.
- Ramesh, Sree Harsha i Sankaranarayanan, Krishna Prasad. 2018. Neural Machine Translation for Low Resource Languages using Bilingual Lexicon Induced from Comparable Corpora. URL: <http://aclweb.org/anthology/N18-4016>. Pristup: 22. 8. 2018.
- Toral, Antonio i Sánchez-Cartagena, Victor M. 2017. A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions. URL: <https://arxiv.org/pdf/1701.02901.pdf>. Pristup: 22. 8. 2018.
- Wikipedia. 2018. „Google Neural Machine Translation”. URL: https://en.wikipedia.org/wiki/Google_Neural_Machine_Translation#Languages_Supported_by_GNMT. Pristup: 16. 4. 2018.
- Wu, Yonghui i sur. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. URL: <https://arxiv.org/pdf/1609.08144v2.pdf>. Pristup: 16. 4. 2018.