Vrednovanje strojnog prevoditelja Google Translate na primjeru jezičnog para turski i hrvatski

Leskovar, Željka

Sveučilište u Zagrebu

Filozofski fakultet

Odsjek za lingvistiku

Vrednovanje strojnog prevoditelja *Google Translate* na primjeru jezičnog para turski i hrvatski

Diplomski rad

Mentorica: dr. sc. Ivana Simeon, viši predavač

Studentica: Željka Leskovar

Zagreb, 2021.

Sadržaj

Sažetak

Ovim radom vrednovan je alat za strojno prevođenje Google Prevoditelj na konkretnom jezičnom paru turski i hrvatski. Analiza polazi od kreiranja korpusa materijala koji se sastoje od originalnih tekstova izvornog jezika, njihova strojnog i konvencionalnog prijevoda kao referentne točke. Odabrani istraživački pristup koristi se kvalitativnom i kvantitativnom analizom pogrešaka koje su prepoznate u izlaznim podacima strojnog prijevoda s ciljem uvida u trenutačno stanje kvalitete prijevoda na manje istaknutim ili korištenim jezicima na globalnoj razini.

Strukturne razlike odabranog jezičnog para dodatno su naglašene razlikom u jezičnim stilovima odabranih tekstova koji su dovedeni u korelaciju s količinom i vrstom pogrešaka izlaznih podataka. Uz statistički prikaz prepoznatih pogrešaka provedena je i lingvistička obrada rezultata s ciljem dublje analize samog jezičnog para i pogrešaka koje su specifične za njega.

Na samom kraju rad donosi nedostatke predmetne analize i prijedlog za promjenu pristupa u budućim istraživanjima.

Ključne riječi: strojno prevođenje, analiza pogrešaka, turski jezik, hrvatski jezik, Google Prevoditelj

Abstract

This thesis evaluates machine translation provided by online tool Google Translate, on a specific language pair Turkish and Croatian. Analysis started off by building a corpus consisted of source language texts as a base and then their machine translation using online tool Google translate and last but not least, conventional translation texts as a reference point. The research part of the thesis is mainly based on qualitative and quantitative analysis of errors identified in machine translation output for gaining insight into current state of translation quality in less prominent languages used on a global scale.

The structural differences between chosen languages are further accentuated by the difference in written style of the selected texts that can be put into correlation with the amount and type of output errors. In addition to the statistical presentation of the identified errors, a linguistic processing of the results was performed with the aim of a deeper analysis of the language pair itself and their characteristical errors.

At the very end, the thesis brings the shortcomings of the analysis and a potential proposal for changing the approach in future research.

Keywords: machine translation, error analysis, Turkish, Croatian, Google Translate

1. Uvod

S obzirom na rastuću potrebu za informacijama na sve većem broju jezika i u što kraćem roku, korištenje jezičnih tehnologija i općenito napredak u polju strojnog prevođenja postali su neizbježni. Samim napretkom strojnog prevođenja uvelike se olakšava posao konvencionalnim prevoditeljima te im se omogućava odabir poslova koji su im potencijalno zanimljiviji u odnosu na zamorne, jednolične tekstove čiju kvalitetu prijevoda već sad mogu osigurati dostupni alati za strojno prevođenje. Dodatni zahtjev na globalnoj razini jest i potreba za komunikacijom na stranim jezicima s ciljem olakšavanja poslovanja, širenja i povezivanja tržišta. U tom kontekstu, o financijskoj uštedi koje razvoj tehnologija s ovog područja donosi bespredmetno je govoriti, no premda njihovo uvođenje zahtijeva možda znatnija ulaganja na početku, zajamčeni su isplativost i relativno brz povrat uloženoga.

Iako je strojno prevođenje preuzelo već dobar dio toga „nezanimljivoga“ posla konvencionalnim prevoditeljima, neprestano ulaganje u ovo područje, provođenje istraživanja i davanje naputaka za poboljšanje rezultata strojnog prijevoda i kvalitete sustava osigurava potencijalno preživljavanje i *malih jezika* na svjetskoj sceni. Kad je riječ o hrvatskom jeziku, ulaganje u jezične tehnologije itekako je prepoznato i neprestano se razvijaju novi alati i resursi kako bi se osigurala vidljivost i konkurentnost na tom znanstvenom području.

Svrha je ovoga rada upravo prikupljanje podataka o hrvatskom jeziku i kvaliteti strojnog prevođenja u kombinaciji s dosad neistraženim jezičnim parom, turskim. Uz neizostavan teorijski dio, definiranje i povijesni pregled strojnog prevođenja u prvom dijelu rada čitatelj se upoznaje s Google Prevoditeljem, alatom koji će biti korišten u istraživačkom dijelu rada, i s konkretnim obilježjima jezičnog para koji će biti analiziran. Iako je analizirani korpus potencijalno premalen i nedovoljno raznolik s obzirom na tip teksta koji je predmet istraživanja, samo prikupljanje informacija potencijalno može poslužiti dobivanju jasnije slike stanja strojnog prevođenja za hrvatski jezik kao ciljni.

2. Strojno prevođenje

Prije definiranja strojnog prevođenja potrebno je odrediti sam pojam prevođenja kao interpretaciju jezične tekstne građe jednog jezika i istovremenu proizvodnju najsličnije jezične tekstne građe u drugom jeziku. Jednostavnije rečeno, prevođenje služi za prijenos tekstova iz izvornog u ciljni jezik uz očuvanje značenja sadržanog u izvornom tekstu.[1] Naizgled jednostavan pojam ima slojevitost koja, prema Lujiću (2007), nikako ne može biti svedena na slijepo prenošenje leksičkih značenja riječi između jezika ili pokušaj prenošenja gramatičkih i sintaktičkih struktura izvornog jezika u ciljni. Kao primarni cilj svakog prevoditelja može se navesti pokušaj da čitatelj prilikom čitanja teksta ne primijeti da se radi o prijevodu, što bi pretpostavljalo dobro razumijevanje teme i konteksta u izvornom tekstu, ali i njihovo još bolje razumijevanje u ciljnom jeziku.

Ukoliko se prevođenje definira kao proces prijenosa značenja s jednog jezika na drugi, utoliko strojno prevođenje (eng. *Machine Translation*), prema Tadiću (2003), najjednostavnije predstavlja prevođenje koje obavlja računalo. Prema Hutchinsu i Haroldu (1992), automatizirati ili mehanizirati proces prevođenja jedan je od najstarijih snova čovječanstva koji se ostvario s pomoću informatizacije i digitalizacije. Dodatno, s obzirom na rastuću potrebu za prijevodom dokumenata iz različitih područja, koji često znaju biti zamorni i ponavljajući, nedostaje prevoditelja, a potreba za asistencijom računala u prijevodu postaje neizbježna. Prema Hutchinsu (1996), ideja o razvoju potpuno automatiziranog sustava opće namjene koji bi bio sposoban producirati prijevod gotovo ljudske kvalitete davno je zaboravljena.

> *Osnovni je cilj istraživanja s područja strojnog prevođenja proizvesti 'pomagala za profesionalne i neprofesionalne prevoditelje koja uporabom računala podupiru ljudske vještine i inteligenciju'.* (Hutchins, prema Tadić: 2003, str. 36)

S time na umu, razvojem i ulaganjem u područje strojnog prevođenja neće nestati potreba za ljudskim prevoditeljima, već će im se omogućiti bolja učinkovitost u radu i smanjiti obujam prijevoda koji su ljudima ionako nezanimljivi. Dovedan i autori (2002) kada govore o strojnom prevođenju podrazumijevaju proces prijenosa informacija s izvornog prirodnog jezika na ciljni prirodni jezik s pomoću računala i drugih elektroničkih pomagala[2] koja skraćuju vremenski tijek samog procesa i olakšavaju ga. Bowker (2002) slaže se s tom definicijom i dodatno ističe

[1] Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2020. Pristupljeno 17. 1. 2021. http://www.enciklopedija.hr/Natuknica.aspx?ID=50270.
[2] Autori navode kao primjer rječnike, tezauruse i baze.

kako se kvaliteta strojnog prijevoda znatno poboljšala tijekom svog postojanja na znanstvenoj sceni te su pogreške i dalje prisutne, ali bitno manje nego ranije. Petrzelka (2011) navodi pak kako je strojnom prevođenju ipak cilj zamijeniti ljudskog prevoditelja računalom i smatra ga specifičnom primjenom znanstvene discipline obrade prirodnog jezika (*NLP, Natural Language Processing*)[3]. Ističe kako kvalitetni rezultati strojnog prijevoda s jednog jezika na drugi nisu česti i da je zapravo u većini slučajeva potrebna revizija prijevoda koju će obaviti ljudski prevoditelj, ali slaže se kako sustavi za strojno prevođenje mogu povećati učinkovitost ljudskih prevoditelja.

> *Kvaliteta prijevoda jedan je od najvažnijih čimbenika. Vrijednost informacija se s vremenom mijenja ili se prevode tekstovi namijenjeni užem krugu ljudi, a kvaliteta prijevoda ne mora biti savršena. Ponekad je dovoljan grubi prijevod teksta kojim se prenose najvažnije informacije.* (Dovedan i autori: 2002, str. 1)

Dovedan i autori (2002) također ukazuju na pripadnost strojnog prevođenja u područje obrade prirodnog jezika, no s obzirom na istaknute karakteristične osobine i dalje se izdvaja kao zasebno područje. Širi pojam koji je usko vezan za to područje neizostavno je i računalna lingvistika, odnosno, kako navodi Tadić (2003), znanstvena disciplina koja se bavi računalnom obradom prirodnog jezika.

Kučiš (2010) ističe pak utjecaj jezičnih tehnologija na prevoditeljsku industriju. Upravo napredak jezičnih tehnologija omogućuje učinkovitije i brže prevođenje, a cilj im je između ostalog uklanjanje jezičnih barijera. Naglašava i bitnu ulogu informacijskih tehnologija u tom području koje služe kao značajna potpora radu suvremenog prevoditelja. Tadić (2003) ide korak dalje i navodi kako jezične tehnologije ovise o informacijskim tehnologijama kao tehnološkoj osnovi i kako bez njih jezične tehnologije ne mogu postojati.

> *Sve ključne znanstvene spoznaje za razvitak jezičnih tehnologija dolaze s područja lingvistike i fonetike, a informacijska tehnologija služi kao sredstvo za njihovu primjenu u daljnjim istraživanjima i konačnim proizvodima.* (Tadić: 2003, str. 13)

Kvaliteti i standardu prevođenja, kako ističe Kučiš (2010), doprinosi timski rad, odnosno razvoj suvremenih informacijsko-komunikacijskih tehnologija s jedne te primjena jezičnih tehnologija s druge strane. Bitno je ipak istaknuti kako je ljudski faktor i dalje neizostavan u izradi profesionalnog i kvalitetnog prijevoda. Međutim, dodatnom uporabom alata i resursa, koji tom

[3] Grana umjetne inteligencije koja spada i u polje računalstva, a bavi se sposobnošću strojeva da razumiju i interpretiraju ljudski jezik kako je pisan ili izgovoren.

istom prevoditelju mogu olakšati posao, zadovoljene su potrebe sve zahtjevnijeg tržišta i održana je kvaliteta samog prijevoda.

2.1. Povijesni pregled

Postoje razilaženja u mišljenju o tome kada je nastala ideja o automatiziranom prijevodu između prirodnih jezika i tko je njezin začetnik, no svi se slažu u tome da je sam razgovor o strojnom prevođenju počeo i prije uporabe prvih računala. Prema Dovedanu i autorima (2002), ozbiljniji pokušaji kreiranja strojeva ili sustava za strojno prevođenje ipak su započeli po pojavi prvog elektroničkog računala ENIAC-a (eng. *Electronic Numerical Integrator and Calculator*) 1946. godine. Razvoj strojnog prevođenja započeo je tako, prema Hutchinsu (2002), 1947. godine razgovorom i korespondencijom između Andrewa D. Bootha i Warrena Weavera, ali ostao je nezamijećen sve do 1949. godine kada je Weaver napisao memorandum „Translation". Spomenuti spis govorio je o mogućnostima korištenja računala u procesu prijevoda dokumenata, ali i ciljevima i metodama njegove automatizacije oslanjajući se na tehnike kriptografije i vojnog šifriranja uz primjenu računala koje su se koristile tijekom II. svjetskog rata. Memorandumom su započeta brojna istraživanja na američkim i europskim sveučilištima s temom strojnog prevođenja, od kojih Dovedan i autori (2002) kao značajnije ističu istraživanje MIT-a (eng. *Massachusetts Institute of Technology*) 1951. godine.

Za taj je period sadržajno bitno to što su od početne ideje stvaranja i integracije cjelokupnih rječnika dvaju jezika između kojih želimo ostvariti prijevod uočene poteškoće:

- *mnoge riječi imaju više prijevodnih ekvivalenata, što ovisi o kontekstu*
- *redoslijed riječi u rečenici u različitim je jezicima različito definiran*
- *lokucije i idiomatski izrazi prevode se kao značenjska cjelina.* (Dovedan i autori: 2002, str. 3)

Već malo konkretnijim pogledom na tu temu došlo je do evidentnog napretka koji, između ostalog, ističe mogućnost da se za proces prevođenja između dvaju prirodnih jezika koristi metajezik ili međukod.

> *Tako su se naizgled jednostavnom rješenju ispriječile brojne teškoće u odabiru odgovarajućeg prijevoda i u poretku riječi u rečenici ciljnog jezika. Naime, uvidjelo se da kvaliteta prijevoda ne ovisi samo o veličini rječnika, te je prevođenje riječ za riječ nestalo, a zamijenila su ga istraživanja s ciljem 'razumijevanja' teksta.* (Dovedan i autori: 2002, str. 3).

Već naredne godine uslijedila je konferencija koju je organizirao Yehoshua Bar Hillel, a okupila je jezične i računalne stručnjake, prema Boothu i Lockeu (1955). Ideja konferencije bila je vođenje neformalnih diskusija, međusobno učenje i upoznavanje vlastitih i tuđih prednosti i mana u pristupu te generalni napredak na tom području. Konferencija nije donijela nikakve bitne zaključke, no sudionici su se ipak složili u tome da su znanja iz područja lingvistike i računalne vještine dovoljno napredovali i naposljetku omogućili strojno prevođenje. U tom periodu dodatno su se iskristalizirali pojmovi pred-urednika i post-urednika, kako navode Dovedan i autori (2002):

> *Pred-urednik priprema tekst za prevođenje kako bi u što većoj mjeri smanjio jezične i strukturne nejasnoće: uklanja višeznačnosti, dugačke rečenice rastavlja na kraće, smanjuje broj zamjenica i jasnije postavlja veze među riječima. Post-urednik nadograđuje prevedeni tekst.* (Dovedan i autori: 2002, str. 3)

Za područje strojnog prevođenja plodna je 1954. godina. Tada je tiskano prvo izdanje časopisa *Mechanical Translation*, koji je zamišljen kao medij za međusobnu komunikaciju zainteresiranih na tom području. Iste godine IBM je demonstrirao računalo opće svrhe koje je bilo programirano za prevođenje određenog uzorka rečenica s ruskog na engleski koristeći vokabular od samo 250 riječi. IBM je na tom projektu sudjelovao sa stručnjacima sa Sveučilišta Georgetown koji su dodatno osmislili pravila sintakse i kodiranje koje je IBM-u omogućilo programiranje. Potaknute tom demonstracijom i lakšim širenjem informacija s tog područja koje je omogućila pojava spomenutog časopisa, različite institucije počele su ulagati u razvoj sustava za strojno prevođenje na području SAD-a.

Sve to završilo je osnutkom ALPAC-a (*Automatic Langauage Processing Advisory Committee*) 1964. i izvješćem objavljenim 1966. koje je označilo kraj novčane podrške vlade SAD-a projektima vezanima za strojno prevođenje, ali i padom morala znanstvenicima s tog područja. Odbor je osnovan kako bi se provjerila kvaliteta rada, troškovi i budući izgledi trenutačnih projekata i istraživanja s područja strojnog prevođenja, s obzirom na postojeće troškove i zahtjeve za prevođenjem na tržištu. Izvješćem je, prema Arnoldu i ostalim autorima (2002), zaključeno kako ne manjka ljudskih prevoditelja, a isto tako nije izgledno da će u neposrednoj budućnosti biti razvijen sustav unutar strojnog prevođenja koji će reproducirati upotrebljive prijevode općih znanstvenih tekstova. ALPAC-ovim izvješćem došlo je do prekida istraživanja strojnog prevođenja na području SAD-a i zatišja te teme na znanstvenoj sceni za čak cijelo jedno desetljeće.

U tom periodu dogodio se premještaj na području strojnog prevođenja iz SAD-a u Kanadu i Europu, koje su ipak imale veće potrebe i koristi od strojnog prevođenja s obzirom na postojanje dvaju jezika i dviju kultura na kanadskom području, a s druge strane višejezičnost unutar tadašnje Europske zajednice i njezinih država-članica. Tako je 1970. u Montrealu započeo projekt TAUM (*Traduction Automatique de l'Université de Montréal*), koji je, kako navodi Hutchins (1995), imao dva velika postignuća: računalni metajezik za baratanje sintaksom kao temelj programskog jezika *Prologa,* koji je vrlo raširen u području obrade prirodnog jezika, i sustav Meteo, osmišljen za ograničeni vokabular i sintaksu meteoroloških izvješća. Svim naporima u tom periodu zajednički je fokus na postojanju međujezika kao metode strojnog prevođenja i dominacija sintaksnog pristupa u odnosu na prethodna istraživanja. Činjenica koja se svakako nameće u vezi s tim razdobljem jest ta da sustavi za strojno prevođenje iz laboratorija ulaze u komercijalne vode i postaju dio druge generacije sustava prevođenja kao što su:

> *ARIANE razvijen na Grenoble Universityju, METAL u Texasu, SUSY u Saarbrückenu, MU na Kyoto Universityju i jedan od najpoznatijih višejezičnih projekata EUROTRA u Evropskoj zajednici.* (Dovedan i autori: 2002, str. 4)

Svojim većim dijelom, prema Arnoldu i ostalim autorima (2002), povijest 80-ih na području strojnog prevođenja zapravo je povijest inicijativa iz prethodnog desetljeća i iskorištavanje poznatih rezultata u susjednim disciplinama. U tržišno orijentiranim sustavima za strojno prevođenje svakako prednjače SYSTRAN kao jedna od vodećih kompanija u svijetu koja se bavi prevođenjem i Logos s partnerskim tvrtkama ALPS i Weidner. Sustavi poput Systrana ili Logosa u principu su zamišljeni za opću primjenu, premda su u praksi, prema Hutchinsu (1995), njihovi rječnici prilagođeni specifičnim jezičnim domenama i okruženjima, što 70-ih i 80-ih godina i nije bila rijetkost. Kvaliteta strojno prevedenih tekstova takvih sustava zasniva se, prema Dovedanu i autorima (2002), na bogatim i dobro razrađenim morfološkim rječnicima, postojanju post-urednika i njegovu posredovanju prije raspodjele dokumenata te naposljetku na kvalitetno osmišljenim programima za obradu riječi i tekstova koji pomažu post-urednicima. Ipak, postaje jasno da je potreban zaista širok spektar znanja za prijevod naizgled jednostavnih rečenica i u toj spoznaji raste ideja da sustavi za prevođenje uz rječnike sadrže i univerzalne enciklopedije.

Sve do 1989. područje rada strojnog prevođenja bilo je u okviru osnovnih lingvističkih pravila vezanih za sintaktičku analizu, leksik, leksički prijenos, morfologiju i dr., odnosno *rule-based.* Osnovna pretpostavka sustava temeljenih na pravilima podrazumijeva analizu i

prikaz značenja teksta izvornog jezika i generiranje jednakog takvog teksta na ciljnom jeziku, a da pritom prikazi budu leksički i strukturno jednoznačni. Pristupi koji su u okviru te metode iz ranije navedenih primjera vidljivi jesu metoda transfera i metoda međujezika. Dominacija tih pristupa te je godine prekinuta značajnijom ulogom korpusa u strojnom prevođenju. Nadolazeće razdoblje od prethodnog je razlikovala primjena statističkih metoda u prevođenju, prevođenja temeljenog na primjerima (eng. *example-based approach*) te primjena korpusa kao baze znanja i izvora lingvističkih podataka. U ovom razdoblju najviše se ističe početak projekta Eurolang, u vlasništvu francuske tvrtke Site, kojemu je cilj bila proizvodnja sustava sa sučeljem koje je s jedne strane lako i intuitivno za korištenje korisnicima, a s druge strane pogodno za prevoditelje uz direktan pristup rječnicima. Većina istraživanja koja su se odvijala u tom razdoblju nije bila u potpunosti fokusirana na samo strojno prevođenje koliko na razvoj sustava opće namjene za obradu prirodnog jezika temeljenih na ujednačavanju i gramatikama koje se baziraju na ograničenjima (eng. *constraint-based grammars*). Sustavi temeljeni na transferu, koji je već ranije spominjan, u prošlosti su bili orijentirani na sintaksu, koja je u ovom razdoblju zamijenjena pristupom koji se više zasnivao na leksiku. Posljedica toga bilo je povećanje informacija koje su se vezale za leksičke jedinice, fokus više nije bio samo na morfološkim i gramatičkim značajkama i prijevodnim ekvivalentima u ciljnom jeziku, već su one sadržavale informacije o sintaktičkim i semantičkim ograničenjima, pa čak i konceptualnim informacijama izvan domene jezikoslovlja općenito. Još jedna od posljedica istraživanja i razvoja u tom razdoblju jest usmjeravanje više pozornosti prema postupcima generiranja dobrih i kvalitetnih tekstova u ciljnom jeziku u odnosu na prethodna razdoblja.

Ipak, najveći pomak u razvoju strojnog prevođenja dogodio se zbog povećanog interesa za prijevod govorenog jezika, koji je pak sa sobom nosio izazove i potrebu kombiniranja prepoznavanja govora (eng. *speech recognition*) s lingvističkom interpretacijom razgovora i dijaloga, prema Hutchinsu (1995). Istovremeno se, prema Dunđeru (2015), povećavao broj digitalnih tekstnih korpusa, što zbog veće proizvodnje digitalnih tekstova na internetu, što zbog samog procesa digitalizacije fizičke dokumentacije. Dodatno, razvoj računala usmjeren je sve više prema široj proizvodnji i primjeni osobnih računala, što potencijalno sa sobom nosi veći broj korisnika strojnog prevođenja i pojavu besplatnih internetskih prevoditeljskih servisa. Svi ti čimbenici pridonijeli su uzletu na području strojnog prevođenja, koje se 2000-ih godina, prema Koehnu (2010), počinje realizirati kroz sustave za statističko strojno prevođenje temeljeno na empirijskim opažanjima.

> *Takvi sustavi se pretežno izgrađuju za određenu domenu, tj. karakteristično područje s ograničenim vokabularom i specifičnim rečenicama, s obzirom da za uže područje namjene generiraju kvalitetnije strojne prijevode (Haddow i Koehn: 2012).* (Dunđer: 2015, str. 14)

Uz statističke prijevode danas se, prema Dunđeru (2015), sve više istražuju hibridni sustavi koji, primjerice, uz sustave za statističko strojno prevođenje upotrebljavaju i neki drugi izvor lingvističkog znanja ili pristup temeljen na primjerima. Također se istražuju i mogućnosti strojnog prevođenja tipa govor u govor (eng. *speech-to-speech translation*) i prijevodne memorije (eng. *translation memory*).

Bez obzira na smjer kojim su istraživanja na području strojnog prevođenja kretala ili će kretati, jedno može biti jasno: strojno je prevođenje, kako ističe Simeon (2008), bez obzira na kvalitetu prijevoda itekako korisno, omogućuje korisniku u prvoj fazi upoznavanja teksta ulaganje minimalnog truda za maksimalnu dobit i olakšava donošenje odluke o korisnosti teksta i eventualnoj potrebi za kvalitetnijim prijevodom.

> *Cilj je većine strojnoprevoditeljskih sustava automatizirati prvu fazu, odnosno uštedjeti prevoditelju zamorno listanje po rječniku i upisivanje teksta, a uz to, ako je moguće i osigurati terminološku dosljednost.* (Simeon: 2008, str. 81)

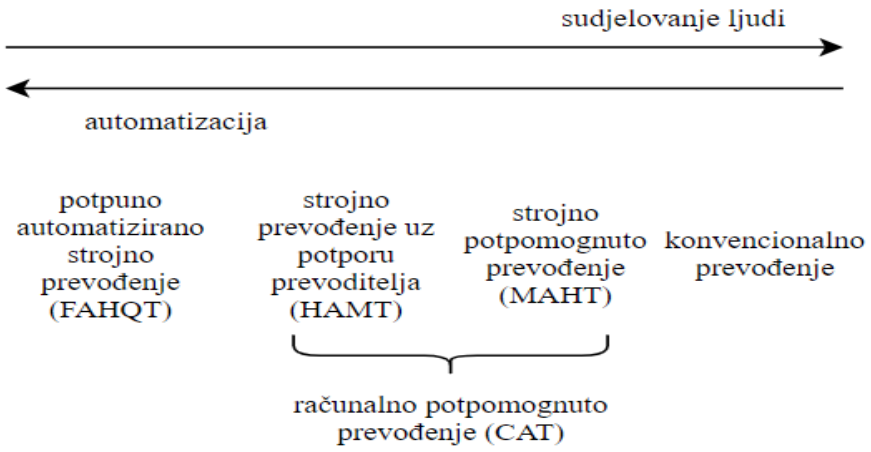
Simeon (2008) smatra da je tržište prepoznalo vrijednost ulaganja u istraživanja s područja strojnog prevođenja i prevoditeljskih sustava koji su njihov plod. Otvaranje novih vidika i razvoja u tom području vidi u napretku ne samo u području jezičnih i računalnih tehnologija već i u području umjetne inteligencije i teorijske lingvistike, koji će naposljetku otvoriti smjer kojim će se kretati u budućnosti.

To se prema Antunović i Pavlović (2019) dogodilo upravo 2016. godine kada je došlo do nove prekretnice na području strojnog prevođenja – uvođenja neuronskog modela sustava za strojno prevođenje u širu primjenu. Taj sustav služi se tzv. dubokim učenjem (eng. *deep learning*) i može se poboljšati dodatnim treniranjem. Prijevodi koji su rezultat takve vrste strojnog prevođenja sadrže manji broj pogrešaka od dosadašnjih, ali ni oni nisu bez nedostataka; pogreške koje ipak generira sustav znatno je teže predvidjeti i uočiti. U praksi strojnog prevođenja i dalje je nezaobilazan, posebice kod tekstova koji se objavljuju u javnosti, proces redakture. Iako postoje napori da se taj proces automatizira, prepravke tekstova kako bi se postigla odgovarajuća kvaliteta i dalje obavljaju ljudi.

Negativni stavovi prema uvođenju sustava strojnog prevođenja u prijevodni proces uz redakturu, prema Antunović i Pavlović (2019), i dalje su prisutni, no autorice smatraju da oni proizlaze iz neznanja i nedovoljno istraživanja teme. Iz sve većeg broja istraživanja teme može se zaključiti da će iz toga zasigurno izniknuti nove informacije koje bi mogle poslužiti potencijalnom daljnjem razvoju na području strojnog prevođenja.

2.2. Vrste sustava za strojno prevođenje

Problemu strojnog prevođenja može se pristupiti iz više gledišta, što nameće i činjenicu postojanja više vrsta strojnog prevođenja, na koje autori opet različito gledaju. Hutchins i Somers (1992) strojno prevođenje vide u spektru prijevodnih metoda na suprotnom kraju od konvencionalnog prevođenja, kao što je vidljivo na Slici 1. S jedne strane imamo potpuno automatizirano visokokvalitetno strojno prevođenje (eng. *fully automated high-quality machine translation, FAHQMT*), odnosno računalni sustav koji producira prijevod visoke kvalitete bez ljudske intervencije, dok s druge strane imamo konvencionalni prijevod koji u procesu prijevoda ne koristi nikakva računalna ili mehanička pomagala. Između se nalaze produkti suradnje ljudi i računala na području prevođenja koja je rezultirala pojmom računalno potpomognutog prevođenja (eng. *computer assisted translation, CAT*). Taj sustav sadrži alate i resurse poput prijevodnih memorija i leksičkih baza podataka s ciljem pomoći prevoditelju prilikom rada, a dijeli se dodatno na strojno prevođenje potpomognuto čovjekom (eng. *human-assisted machine translation, HAMT*) i ljudsko prevođenje potpomognuto strojem (eng. *machine-assisted human translation*).



Slika 1. Konvencionalno i strojno prevođenje (po uzoru na Hutchins i Somers: 1992, str. 148)

Šimić i autori (2010) kategoriziraju pak vrste strojnog prevođenja na sličan način; razlikuju se u činjenici da Hutchins i Somers svoju podjelu temelje na sustavima prevođenja općenito, a ne konkretno na strojnom prevođenju kao u ovom primjeru.

> *Vrste programa za strojno prevođenje najčešće se klasificiraju unutar dvije osnovne kategorije: način uporabe i princip rada.* (Šimić i autori: 2010, str. 82)

Prvu od istaknutih kategorija dodatno dijele na tri grupe koje se pojmovno poklapaju s Hutchinsovom i Somersovom podjelom: strojno prevođenje uz potporu prevoditelja (eng. *human-aided machine translation; HAMT*), strojno potpomognuto prevođenje (eng. *machine-aided human translation, MAHT*) te potpuno automatizirano strojno prevođenje (eng. *fully automatic high-quality machine translation, FAHQMT*). Prve dvije, prema istom izvoru, mogu se svesti na zajednički nazivnik računalno potpomognutog prevođenja (eng. *computer-aided translation*), dok treća živi u teoriji, izuzev izdvojenih slučajeva u praksi ograničenog jezičnog ulaza i rezultata, pa samim time i korisnosti uporabe. Što se pak tiče principa rada i podjele po toj osnovi, autori ističu nekoliko vrsta sustava koji povijesno imaju širu primjenu i veću tržišnu vrijednost: razvijeni na principu rječnika (eng. *dictionary-based*), na principu primjera (eng. *example-based*) i sustavi temeljeni na pravilima (eng. *rule-based*), koje dodatno dijele na sustave s prijenosom (eng. *transfer*) i međujezikom (eng. *interlingua*). Prema istom izvoru, pristup koji spaja dva već navedena pristupa, ona na temelju pravila i na principu primjera, naziva *statističkim strojem* (eng. *statistic machine*):

> *...pri kojem se pri analizi izvornog i sintezi rezultantnog teksta uz gramatička pravila koristi i ogromna baza podataka prevedenih tekstova kojima se metodom statističke usporedbe pronalazi odgovarajuća riječ, fraza ili čak i čitava rečenica s najvećom vjerojatnošću ispravnosti.* (Šimić i autori: 2010, str. 83)

Tadić (2003) navodi dvije osnovne vrste: sustave temeljene na pravilima (eng. *rule based system)* i empirijske sustave ili sustave temeljene na podacima. Prvi istaknuti sustavi služe se skupovima pravila kako bi osigurali odvijanje procesa prevođenja iz izvornog u ciljni jezik, a mogu biti izravni (transformacijski) i neizravni (s jezičnim znanjem). Izravni sustavi povijesno su najstarija vrsta sustava na području strojnog prevođenja te u suštini funkcioniraju po principu zamjene riječi izvornog jezika riječima ciljnog jezika, oslanjajući se pritom na opsežne dvojezične rječnike. Sljedeći je korak u procesu preustroj reda riječi u rečenici prema pravilima ciljnog jezika.

> *Kako ovakvi sustavi zapravo nemaju nikakvih pravila vezanih uz sintaktičku strukturu (osim reda riječi u ciljnom jeziku), oni iznimno ovise o kvalitetnim kontrastivnim gramatikama i detaljnim dvojezičnim rječnicima koji obuhvaćaju prijevodne ekvivalente i iznad razine pojedinačnih riječi.* (Tadić: 2003, str. 38)

Neizravni sustavi, prema istom autoru, najprije prolaze kroz sustav raščlambe rečenica izvornog teksta na sintaktičkoj i semantičkoj razini, nakon toga slijedi apstraktna razina prikaza značenjskih odnosa u izvornom tekstu i pronalazak odgovarajućeg značenjskog ekvivalenta u ciljnom jeziku, u kojem se naposljetku generira rečenica. Sustavi s jezičnim znanjem dalje se dijele na sustave s prijenosom (eng. *transfer*) i sustave s međujezikom (eng. *interlingua*). Prvi od istaknutih tekstove izvornog i ciljnog jezika povezuju pravilima za prijenos, dok drugi tekstove svode na isti zapis jezika posrednika (umjetnog jezika). Empirijski sustavi, prema Tadiću (2003), temelje se na bazama podataka ili usporednim korpusima sastavljenim od izvornih tekstova i njihovih prijevoda.

> *Rečenično sravnjeni (eng. sentence aligned) usporedni korpusi oni su u kojima se točno zna koja je ciljna rečenica prijevod koje izvorne. Na temelju tako eksplicitno obilježenih prijevodnih ekvivalenata na razini rečenica, pokušava se statističkim metodama doći do prijevodnih ekvivalenata na nižoj razini tj. na razini riječi i sintagmi te tako dobiti korpuse koji su rječno sravnjeni (word aligned).* (Tadić: 2003, str. 39)

Takvi sustavi za strojno prevođenje dakle dodatno koriste razne statističke metode za pronalazak prijevodnih istovrijednica ili pak, ako je riječ o sustavima za strojno potpomognuto prevođenje, traže na strukturnoj i leksičkoj razini rečenice koje su slične izvornima; u tim istim rečenicama na mjestima razlikovanja uklanjaju leksičke jedinice koje im ne odgovaraju kako bi ih naposljetku popunili u ciljnim rečenicama.

Simeon (2008) ističe pak podjelu strojnog prevođenja u užem i širem smislu: u užem smislu ono obuhvaća već istaknuto potpuno automatizirano visokokvalitetno strojno prevođenje i strojno prevođenje uz ljudsku intervenciju (eng. *fully automated high-quality machine translation, FAHQMT*), dok u širem smislu obuhvaća strojno potpomognuto prevođenje (eng. *machine-aided translation, MAT*). Kad je riječ o konkretnim vrstama sustava za strojno prevođenje, dijele se na sustave temeljene na pravilima i statistički utemeljene sustave. Prvom pristupu u centru je jezično znanje koje pokušava formalizirati na način razumljiv računalu i podrazumijeva strojno prevođenje temeljeno na pravilima (eng. *rule-based machine translation*). Proces prevođenja unutar tih sustava odvija se na dvije moguće razine:

međujezičnoj (eng. *interlingua*) razini ili razini prijenosa (eng. *transfer*), koje definira na isti način kao i Tadić (2003). Što se tiče statistički utemeljenih sustava, Simeon (2008) ističe kako se taj pristup strojnom prevođenju oslanja na paralelne korpuse kao izvore podataka, odnosno na dvojezične ili višejezične korpuse koji sadrže niz tekstova na dva ili više jezika. Kao prednost tih sustava Simeon (2008) naglašava njihovu neovisnost o jeziku, a kao nedostatak proces odabira jedinica unutar teksta koje su uključene u postupku traženja koji je već istaknut ranije uz ovaj pristup.

Ljubas (2017) ističe kako su upravo statistički sustavi brojem javno najdostupniji sustavi za strojno prevođenje.

> *Pri statističkom prevođenju sustav pretražuje dostupne arhive prijevoda i izračunava 'vjerojatnost da će se neka riječ prevesti nekom drugom ili da će se prijevodi dviju riječi koje se nalaze jedna pored druge također nalaziti jedni pored drugih'.* (Ahrenberg i Merkel, prema Ljubas: 2017, str 31)

Točnost rezultata i izračuna ovisi o reprezentativnosti materijala koje sustav koristi. Kao prednost tih sustava navodi mogućnost da sam korisnik sustava predloži prijevod koji bolje odgovara kontekstu; na taj način sustav pamti, ažurira i poboljšava rezultate samih prijevoda. Ljubas (2017) unatoč većoj dostupnosti statističkih sustava ističe u svom istraživanju novu paradigmu u svijetu strojnog prevođenja – neuronsko strojno prevođenje (eng. *neural machine translation, NMT*), koje se temelji na neuronskim mrežama i principima dubokog učenja (eng. *deep learning*). Goodfellow i dr. (2016), kako je navedeno u radu Antunović i Pavlović (2019), objašnjavaju taj pojam kao vrstu strojnog učenja koja računalnim sustavima pruža mogućnost napretka temeljenog na velikim količinama relevantnih podataka. Ako se to stavi u kontekst prevođenja, relevantne podatke čine paralelni dvojezični korpusi s izvornim tekstovima i njihovim prijevodima, koji sustavu daju mogućnost učenja procesa prevođenja rečenica ljudskog jezika. Ljubas (2017) također pokušava objasniti taj pristup u svom istraživanju:

> *Sustav s pomoću neuronske mreže pokušava razumjeti kontekst ulazne rečenice X i izračunatu vjerojatnost da prijevod Y odgovara, a da mu pritom nisu potrebne vanjske lingvističke informacije.* (Castilho i sur.: 2017, prema Ljubas: 2017, str. 32)

Neuronsko strojno prevođenje tako se može istaknuti kao novi pristup, ili čak nova vrsta sustava strojnog prevođenja koja dosad još nije bila poznata, i velike se nade polažu u njega. Međutim, već je sada jasno da taj model, jednako kao i statistički, ovisi o pristupačnosti većih količina podataka, tj. usporednih korpusa za željeni jezični par i velikih jednojezičnih modela za ciljni

jezik o kojima izravno ovisi točnost i kvaliteta rezultata. Ipak, glavna je prednost tih sustava mogućnost prepoznavanja tematskog područja, odnosno domene teksta koji prevode.

2.3. Poznati problemi kod strojnog prevođenja

Finka i László (1962) smatraju nezamislivim strojno prevođenje bez potpuna, temeljita i precizna opisa svih jezika koje želimo u tom procesu koristiti.

> *...potrebno je izvršiti goleme predradnje primjenjujući tekovine egzaktnih nauka: statistike, vjerojatnosne i obavijesne teorije, matematičke logike, teorije skupova i grafova i kibernetike.* (Finka i László: 1962, str. 120)

Autori smatraju da obrada jezika ima svrhu i konkretan zadatak – stvaranje ekonomičnog sustava strojnog prevođenja. Ostvarenje tog zadatka imat će društveno, ekonomsko, kulturno i znanstveno značenje, a s obzirom na istaknuta velika očekivanja, istraživanja strojnog prevođenja iziskuju velika ulaganja i napore. Sustavi strojnog prevođenja od svojih prvih inačica pa do danas imaju probleme koji se najčešće svode na kompleksnost samog jezika, pokušaj popisivanja svih jezičnih fenomena i konstrukcija i naposljetku njihov opis koji je prilagođen za čitanje automatskim sustavima, kako ističu Arnold i drugi (2002). Još na samim počecima strojnog prevođenja Bar-Hillel (1953) istaknuo je četiri učestalije kategorije pogrešaka: sintaksu prirodnih jezika, mogućnosti prijevoda prirodnih jezika, idiome te univerzalne sintaktičke kategorije, od kojih su neke prisutne i neriješene još i danas.

Iako se strojno prevođenje sve više istražuje i evidentan je napredak u tom području, pogreške u rezultatima i dalje su prisutne, iako ne u tolikom opsegu i obliku kao prije. Bowker (2002) kao faktor koji je utjecao na taj pozitivni pomak u redukciji pogrešaka nastalih prilikom strojnog prevođenja ističe prije svega napredak tehnologije koja je omogućila postojanje sve većih baza podataka za čuvanje leksičkih jedinica, gramatičkih pravila i u globalu jezičnih i izvanjezičnih znanja. Navodi također i sve bolje razumijevanje teorijske lingvistike u smislu lakšeg kodiranja relevantnih jezičnih pravila i otvaranja mogućnosti stvaranju kontroliranih jezika za smanjenje višeznačnosti (eng. *ambiguity*). Još jedan od faktora koji pridonosi smanjenju pogrešaka u prijevodima koji su rezultat strojnog prevođenja jest sve bolje razumijevanje prednosti strojeva kojima se služe, što dovodi do novih pristupa samom procesu – statističkoga pristupa, pristupa temeljenih na pravilima i sada sve prisutnijeg neuronskog pristupa. Još je bitno istaknuti da su očekivanja od korisnika sustava strojnog prevođenja sve

realnija, odnosno ne smatra se da će procesom strojnog prevođenja biti u mogućnosti, neovisno o domeni, producirati visokokvalitetne prijevode i zamijeniti ljude u potpunosti, već postaje sve izvjesnije kako će upravo računala znatno pomoći ljudima u obavljanju prevoditeljskog posla.

> *Niti jedan strojnoprevoditeljski sustav ne može se smatrati dovršenim, a pogotovo ne savršenim. I tvorci sustava i njegovi potencijalni kupci moraju biti svjesni njegovih ograničenja i mogućnosti za poboljšanje.* (Simeon: 2008, str. 44)

Dovedan i dr. (2002) kao najveći problem u strojnom prevođenju navode već istaknutu višeznačnost riječi ili rečenica te predlažu dva načina za njezino prevladavanje: stvaranje podjezika ograničavanjem gramatike i vokabulara izvornog teksta ili odbacivanje zahtjeva za potpunu automatizaciju posredstvom čovjeka.

Šimić i dr. (2010) priklanjanju se tim navodima naglašavajući ponovo višeznačnost kao najčešći i najteži problem prevođenja općenito uz anaforu, sintaktičku analizu, poredak riječi u rečenici, stilistiku i dr., te izdvajaju dva načina kojima se ti problemi rješavaju.

> *Programi za strojno prevođenje rješavaju gore navedene probleme već spomenutim razvojem tehnologije i teoretskih koncepata. Drugi način rješavanja problema je da ih jednostavno izbjegnu. Time se dakako uvelike smanjuje kvaliteta samog strojnog prevođenja, pa ipak, bolje je imati ikakav rezultat nego nikakav.* (Šimić i dr.: 2010, str. 84)

Hutchins (2003) objašnjava kako su prepreke poboljšanju strojnog prevođenja iste kao na početku te ističe njih nekoliko: problem višeznačnosti, pogrešan odabir riječi u ciljnom jeziku, anafore (zamjenice i članovi[4]), neprimjereno zadržavanje struktura iz izvornog jezika, pogrešnu koordinaciju, morfološke pogreške, pogreške u redu riječi, brojne i raznovrsne poteškoće s prijedlozima, predikatne oblike, umetnute zagrade te općenito probleme sa složenim rečenicama.

Arnold i drugi (2002) kao najistaknutije probleme strojnog prevođenja navode i opisuju sljedeće: dvoznačnost, leksička i strukturna nepodudaranja te višerječne jedinice – idiome i kolokacije. Prvi istaknuti problem dvoznačnosti autori dalje dijele na leksičku, koja podrazumijeva više značenja jedne riječi, i strukturnu dvoznačnost, odnosno postojanje više struktura primjenjivih na određenu rečenicu ili neki njezin dio.

[4] Svojstveno za engleski jezik, Hutchins (2003) ističe konkretno određene članove, odnosno *the*.

> *Različiti su uzroci dvoznačnosti u tekstovima: strukturalna dvoznačnost znači da se u tekstu pojavljuju homografi, što stroju otežava raščlambu rečenice, zatim postoje dvoznačnosti na razini leksema (što se odnosi na polisemiju i metafore), dvoznačnosti na razini anafore (kada se teže razumije na što se neka zamjenica odnosi), a katkad nastaju problemi i zbog toga što u jezicima postoji više načina da se izrazi ista stvar.* (Bennett i sur.: 1986, prema Ljubas: 2017, str. 34)

Problemi vezani za leksička nepodudaranja među jezicima, kako objašnjavaju Arnold i dr. (2002), odnose se na razlike u načinima na koje jezici klasificiraju svijet, odnosno koje pojmove podupiru izrazima, a koje ipak ne odluče leksikalizirati. Oni koji su pak vezani za strukturna nepodudaranja očituju se u različitim jezicima koji koriste različite strukture za istu svrhu ili iste strukture za različite svrhe, što neizbježno stvara probleme u procesu prevođenja. Prevoditelji kod konvencionalnih prijevoda u tim slučajevima mogu biti i donekle kreativni pa odabrati između izravnog posuđivanja, korištenja neologizma ili pružanja opisa ili objašnjenja specifičnih leksičkih jedinica. Kad je riječ o idiomima, prema istom izvoru, oni se najjednostavnije mogu definirati kao izrazi čije se značenje ne može u potpunosti razumjeti iz značenja sastavnih dijelova. Prema Hrvatskom jezičnom portalu i mrežnom izdanju Hrvatske enciklopedije, problematičnost koju ti izrazi predstavljaju za strojno prevođenje istaknuta je u opisu samog pojma u oba spomenuta izvora:

> *Idiom – osebujna riječ ili izraz, ob. teško prevodiv ili neprevodiv na drugi jezik [šilo za ognjilo]* (http://hjp.znanje.hr/index.php?show=search, 26. 1. 2021.)

> *Idiom se katkada (prema engl. idiom) pogrješno upotrebljava u značenju 'teško prevodiv izraz specifičan za pojedini jezik'.* (idiom. Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2020. Pristupljeno 27. 1. 2021. http://www.enciklopedija.hr/Natuknica.aspx?ID=26925)

U kontekstu strojnog prevođenja problematično je to što oni ne mogu biti podložni istim pravilima u prijevodnom procesu; iako primjena postojećih pravila ponekad može rezultirati jednakim značenjima, u većini slučajeva tome nije tako. Stoga Arnold i dr. (2002) predlažu tretiranje idioma kao zasebnih jezičnih jedinica ili s posebnim pravilima koje će prepoznati idiomatsku strukturu izvornog jezika i pronaći odgovarajući par u ciljnom. Druga od istaknutih višerječnih jedinica koju autori spominju jest kolokacija, kod koje se značenje može shvatiti iz značenja pojedinih riječi koje ju sačinjavaju, ali nepredvidivi su odabir i primjena određenih riječi.

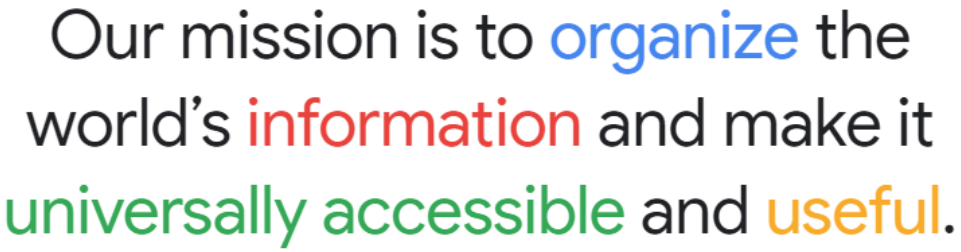
Kolokacija – lingv. obvezatna ili uobičajena veza riječi koja nije određena gramatikom [podnijeti molbu ali uložiti žalbu] (http://hjp.znanje.hr/index.php?show=search, 26. 1. 2021.)

S obzirom na istraživanja i radove vezane za tematiku vrednovanja strojnog prevođenja, često se klasifikacija poznatih ili anticipiranih pogrešaka, od kojih je većina navedena, odabire s obzirom na jezični par koji je predmet istraživanja ili pak na subjektivno znanje istraživača. Takav pristup, bio naposljetku u rezultatima prikladan kategorizacijom ili ne, svakako daje dodatni uvid u jezični par i pogreške koje su manje ili više karakteristične za njega, te postaje dio rješenja.

Daljnji je razvoj strojnog prevođenja nezaustavljiv, postoji sve veća potreba za njime, a kvaliteta sustava i njihovih rezultata raste zahvaljujući teorijskom lingvističkom znanju i interdisciplinarnosti pristupa. Poznavanje problema svakako je dio tog rješenja, stoga je potrebno percipirati važnost ulaganja u područje jezičnih tehnologija i sustavno raditi na njihovu razvoju.

2.4. Google Prevoditelj

Googleova je misija organizirati sve informacije svijeta i učiniti ih opće dostupnima te dakako korisnima.



Slika 2. Misija Googlea (https://about.google/intl/en/ [pristupljeno 26. 1. 2021.])

S tim na umu, javno dostupna i besplatna usluga Google Prevoditelja, u svijetu rastuće potrebe za informacijama na različitim jezicima bez previše napora, čini se kao logičan izbor proizvoda. Google je 2006. godine objavio puštanje u javnost servisa Google Prevoditelja, koji se tada oslanjao na statističko strojno prevođenje bazirano na frazama kao ključnom algoritmu za rad.[5] S obzirom na broj korisnika koji se njime od tada neprestano služi i globalno jačanje potrebe za višejezičnom komunikacijom, ulaganje u taj proizvod bilo je neizbježno, a važno je istaknuti kako rezultati također ne izostaju. Deset godina od lansiranja tog servisa za automatsko prevođenje u javnost, Google je najavio korištenje najsuvremenije tehnike učenja, dubokog učenja (eng. *deep learning*) za postizanje dotad najvećeg poboljšanja kvalitete strojnog prevođenja vezanog za alat Google Prevoditelj. Napredak u strojnom učenju (eng. *machine learning, ML*) potaknuo je napredak u sektoru strojnog prevođenja pojavom novog modela neuronskog strojnog prevođenja (eng. *neural machine translation, NMT*), koji je Google prilagodio za razvoj svog sustava *Google Neural Machine Translation (GNMT)*, tada primjenjivog za više od 100 jezika.[6]

Broj podržanih jezika Googleova servisa sve je veći i danas broji 108 jezika. Prednost tog *online* alata za strojno prevođenje zacijelo je oslanjanje na najveći i sveobuhvatan korpus na svijetu – internet, odnosno mrežno dostupne tekstove kojih je sve više. Njegova je najveća prednost, prema Ljubas (2017), ipak u činjenici da nije usmjeren na neko specifično područje,

[5] https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html [pristupljeno 28. 1. 2021.]
[6] Ibid.

već je sveobuhvatan, čak i nauštrb preciznosti i točnosti. Dio onoga što danas karakterizira sustav Google Prevoditelja u pozadini jesu:

- hibridni model arhitekture sustava koji se pokazao bolji u kvaliteti, stabilniji u učenju s manje istaknutom latencijom
- mrežno indeksiranje (eng. *Web Crawl*) unaprijeđeno novim sustavom za prikupljanje podataka koji je osjetljiviji na samu kvalitetu podataka i fokusira se više na preciznost nego na odziv (eng. *recall*)
- modeliranje iskrivljenih podataka koje se rješava tako da sustav najprije trenira/„uči“ na svim podacima, bez obzira na njihovu čistoću, a zatim se postupno primiče manjim i čišćim setovima podataka
- povratni prijevod koji se pokazao posebno korisnim za povećanje tečnosti u prijevodima jezika s manjim bazama podataka na kojima sustav može učiti.[7]

Iako je napredak Googleova servisa za strojno prevođenje vidljiv, kvaliteta prijevoda daleko je od savršene. Model je i dalje žrtva tipičnih pogrešaka strojnog prijevoda, uključujući loše rezultate u pojedinim tematskim domenama, preplitanje različitih jezičnih dijalekata, stvaranje pretjerano doslovnih prijevoda te lošije performanse sustava u području neformalnog i govorenog jezika. Bez obzira na istaknute pogreške, taj servis za automatsko prevođenje i dalje sadržava relativnu koherentnost u prijevodima koje producira.[8]

[7] https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html [pristupljeno 28. 1. 2021.]
[8] Ibid.

3. Jezični par i njegove karakteristike

Kako objašnjava Jakobson (1959), svaka usporedba dvaju jezika podrazumijeva i ispitivanje njihove međusobne prevodivosti.

> *Prevoditelj bi u najširem smislu riječi trebao dobro poznavati gramatička pravila i sintaktičke sklopove vlastitoga jezika, ali ništa manje i onoga jezika s kojega prevodi. Osnovna je postavka da jezik pretpostavlja jezik, i da prevođenje s jednoga pretpostavlja dobro poznavanje onoga drugoga. Osim toga, neophodno je solidno poznavanje kulturno-duhovnoga i povijesno-društvenoga konteksta u kojem je pojedini tekst nastao.* (Lujić: 2007, str. 60)

Između jezičnog para koji je predmet ovog rada postoji određeni jaz u vidu genetske nesrodnosti i tipološke različitosti, što prema Čauševiću (2018) stvara poteškoće izvornim govornicima turskog jezika prilikom učenja hrvatskog jezika. Dok s jedne strane izvorni govornici turskog jezika hrvatski percipiraju jezikom *gramatičkih iznimaka,* različitosti u gramatičkim strukturama jezičnog para također zahtijevaju ulaganje dodatnog napora za svladavanje određenih kategorija turskog jezika izvornim govornicima hrvatskoga. Halliday (1970), kako je navedeno u radu Baker (1992), svjestan je navedenog te ističe kako jezik daje strukturu iskustvu i pomaže u određivanju načina na koji gledamo na svijet, stoga je potreban i opravdan određeni intelektualni napor za sve što se razlikuje od onoga što nam naš materinski jezik sugerira.

3.1. Turski jezik i gramatika

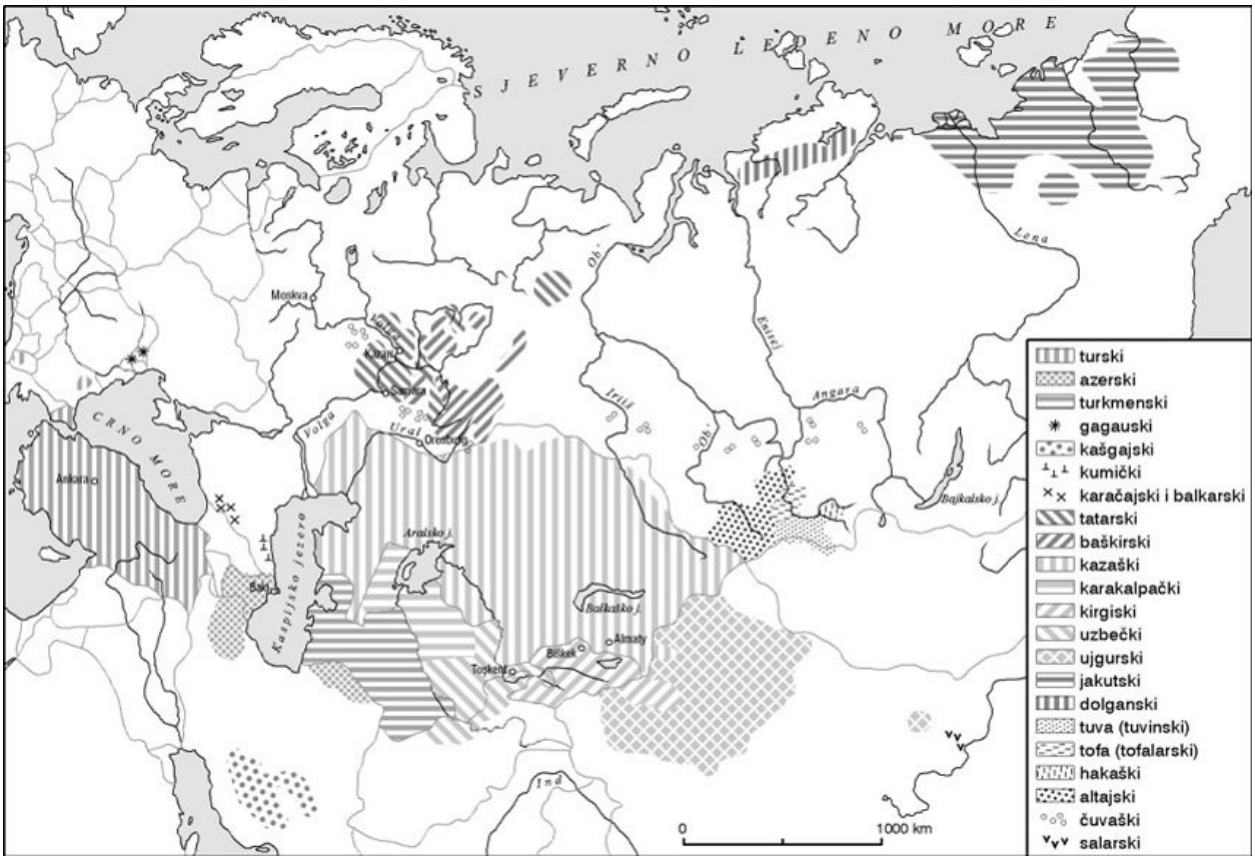
Suvremeni turkijski jezici, kao zasebna jezična porodica genetski srodnih i tipoloških bliskih jezika, prema Čauševiću (1996) imaju svoje osobitosti, a to su:

- vokalna harmonija[9] i eufonična struktura[10] riječi na fonološkoj razini
- aglutinativni tip morfologije, sustav deklinacije i konjugacije koji ne poznaje gramatičke izuzetke, odsustvo gramatičkog roda, korištenje poslijeloga (postpozicija) umjesto prijedloga (prepozicija), mnogo infinitivnih formi, malo veznika i često nejasne granice među vrstama riječi na morfološkoj razini

[9] Pojam podrazumijeva usklađenost vokala i u korijenu/osnovi riječi i u sufiksima za tvorbu i fleksiju.
[10] Odnosi se na skladan odnos glasova unutar riječi koji u pravilu ne tolerira konsonantske skupine.

- osobit poredak riječi unutar sintagmi i rečenica, gramatička nepodudarnost određenih članova, *ljevostrano* nizanje zavisnih elemenata u odnosu na upravni član sintagme, posebni sklopovi infinitnih formi i dr. na sintaktičkoj razini.

S obzirom na istaknute tipološke osobitosti, turkijske jezike često pribrajaju altajskoj jezičnoj porodici, no njihova genetska srodnost još uvijek nije posve uvjerljivo dokazana. Povijesno gledajući, turski jezik formirao se na temelju jezika oguskih plemena[11] koja su naselila Malu Aziju te govorila i pisala starim anadolijskim turskim jezikom, temeljem suvremenog turskog.



Slika 3. Turkijski jezici (https://enciklopedija.hr/Natuknica.aspx?ID=62774 , [pristupljeno 28. 1. 2021.])

Prema Čauševiću (1996), spomenuti je jezik tijekom stoljeća evoluirao u sljedećim razdobljima:

- *stari anadolijski turski jezik (od 13. do 15. st)*
- *ranoturski (u glavnim se crtama formirao na prijelazu iz 15. u 16. st.)*
- *srednjoturski (od 17. do sredine 19. st)*
- *novoturski (od sredine 19. st. do prelaska na latinično pismo 1928. god.)*

[11] Plemena koja su u razdoblju od 12. do kraja 13. st. imala državu pod vodstvom seldžučke dinastije čiji su potomci Turci, Azerbajdžanci, Turkmeni, djelomice i Gagauzi, prema Čauševiću (1996).

- *suvremeni turski* (Čaušević: 1996, str. XII)

Evoluciju jezika u istaknutim fazama važno je naglasiti zbog promjena koje je jezik doživljavao, a koje su zaslužne za današnje stanje turskog jezika. Prema navedenoj periodizaciji jezika, ranotursko i srednjotursko razdoblje najviše karakterizira udaljavanje književnog od govornog jezika. Navedena diferencijacija, kako objašnjava Čaušević (1996), s jedne strane ima elitni književni jezik (*Fasih Türkçe*), koji je pod snažnim utjecajem arapskog i perzijskog jezika sadržavao čak 90% posuđenica i tuđica iz tih jezika, a utjecaj se izvan leksika širio i na gramatičko ustrojstvo. S druge strane postoji druga varijanta književnog jezika, koja se smatrala jezikom srednje obrazovanog sloja (*Orta Türkçe*), bliža govornom jeziku. Oprečno tim varijantama stoji govorni jezik puka (*Kaba Türkçe*), koji je omogućio koegzistenciju više turskih dijalekata i nije poznavao ni razumio prethodne dvije varijante jezika. U narednom razdoblju dolazi do pokušaja jezičnih reformi kojima se nastojalo književni jezik približiti govornom, a osmansko pismo latiničnom, no bez mnogo pomaka. Situacija se mijenja padom Osmanskog Carstva i proglašenjem republike, kada se nastavlja s dotadašnjim pokušajima, međutim oni tada daju i konkretne rezultate[12]. Godine 1928. prelazi se na latinično pismo, 1932. osniva se Tursko lingvističko društvo (*Türk Dil Kurmu*), kojemu je primarna zadaća provođenje jezičnog purizma odnosno stvaranje *čistog turskog jezika (*tzv. *Öztürkçe)* procesom eliminacije, prvenstveno arapskih i perzijskih riječi i frazeologizama. Kapović (2010) ističe upravo turski jezik kao primjer vrlo silovitog jezičnog purizma koji još uvijek živi:

> *Ondje su na glavnom udaru u prošlosti bili arabizmi i perzijanizmi, a u novije vrijeme i druge posuđenice, ponajviše anglizmi.* (Kapović: 2010, str.82)

Jedan od načina provođenja procesa čišćenja jezika, svojstven tom razdoblju, svakako je uvođenje novotvorenica, odnosno značenjskih ekvivalenata arapskim i perzijskim tuđicama. Unatoč nastojanjima suvremeni turski jezik i danas sadrži mnogobrojne riječi iz arapskog i perzijskog jezika te održava dvojni i paralelni leksik turskog i arapsko-perzijskog jezika.

Turski jezik službeni je jezik Republike Turske, a pripada jugozapadnom ogranku turkijskih jezika[13] te grupi aglutinativnih jezika. Prema mrežnom izdanju Hrvatske enciklopedije,[14]

[12] Sve istaknuto dalo je poticaj značajnijoj jezičnoj reformi koju Turci zovu tzv. jezičnom revolucijom, *Dil devrimi.*
[13] Prema klasifikaciji koju navodi Čaušević (1996), suvremeni turkijski jezici dijele se na četiri ogranka; *jugoistočni (uzbečki i novoujgurski), jugozapadni (turski ili osmanskoturski, azerbajdžanski, turkmenski, gagauski), sjeverozapadni (karaimski, kumičkim karačajsko-balkarski, tatarski, baškirski, kazaški, karakalpački, kirgiski i nogajski) i sjeveroistočni (hakaski, altajski i tuvinski). Osim spomenutih u ovu porodicu spadaju i tri izolirana jezika (jakutski, čuvaški i halački), koje nije moguće svrstati ni u jedan od četiri osnovna ogranka*. (Čaušević, 1996, str. IX)
[14] https://www.enciklopedija.hr/natuknica.aspx?id=795 [pristupljeno 28. 1. 2021.]

aglutinativni jezici definiraju se kao tipovi jezika u kojima se sve gramatičke kategorije i gramatički odnosi izražavaju dodavanjem jednoznačnih i standardnih gramatičkih (tvorbenih ili flektivnih) morfema na korijen riječi. Pri procesu aglutinacije korijen i osnova riječi u turskom jeziku nisu generalno podložni promjeni, izuzev vokalne harmonije, a sufiksi se dodaju prema morfotaktičkim pravilima, pri čemu neki mogu biti izostavljeni uz zadržavanje utvrđenog linearnog slijeda. Poredak leksičkih i gramatičkih morfema ostvaruje se prema principu *konkretno pa opće*, a isti se princip aglutinacije prenosi i na sintaktičku razinu. Kad je riječ o klasifikaciji riječi po vrstama, Čaušević (1996) objašnjava kako turski jezik nema posebnih granica među vrstama riječi, ali navodi sljedeće: imenice, pridjeve, zamjenice, brojeve, glagole i priloge kao punoznačne vrste, postpozicije i kvazipostpozicije, veznike, čestice i uzvike kao pomoćne (gramatičke) riječi te naposljetku izdvojene onomatopeje i onomatopejske izraze. Turskom jeziku također je svojstveno nepoznavanje kategorije gramatičkog roda. U slučajevima kada je ipak potrebno u jeziku izraziti prirodni spol ili eksplicirati rod ljudskog bića ili životinje, to se izražava konkretnim setom leksema, najčešće imenica, koji označavaju prirodni spol i koji u jezičnoj uporabi dolaze ispred ili iza imenice na koju se odnose. Kategorija i izražavanje broja u turskom jeziku ovise o situaciji i kontekstu, odnosno namjeri da se izrazi mnoštvo različitih nerazdvojnih jedinki ili mnoštvo koje se da prebrojiti. Turski jezik, prema Čauševićevoj kategorizaciji (1996), sadrži osam padeža: apsolutni padež, genitiv, dativ, akuzativ, lokativ, ablativ, instrumental i relativ-ekvativ.

> *Bitna karakteristika padeža nije njegov vanjski oblik izražen sufiksom ili odsustvom sufiksa, nego njegov smisaoni sadržaj i funkcije koje on može imati u strukturiranju rečenice.* (Čaušević: 1996, str. 75)

U turskom jeziku ne postoje određeni i neodređeni članovi, već se ta kategorija izražava brojem jedan (*bir*) i raznim gramatičkim sredstvima. Još jedno od obilježja turskog jezika i kategorizacije koja mu je svojstvena odnosi se na morfološke kategorije glagola: lik, lice, broj i način. Potencijalno svim jezicima svojstveno je postojanje istog ili sličnog gramatičkog ustrojstva bez obzira na tipološke razlike, a razlike se u tom slučaju očituju samo u pravilima određivanja mjesta u redoslijedu pojavljivanja člana u rečenici.

> *Osnovno gramatičko ustrojstvo rečenice čine njezini sljedeći članovi: predikat, subjekt, objekt, adverbijalna oznaka.* (Čaušević: 1996, str. 452)

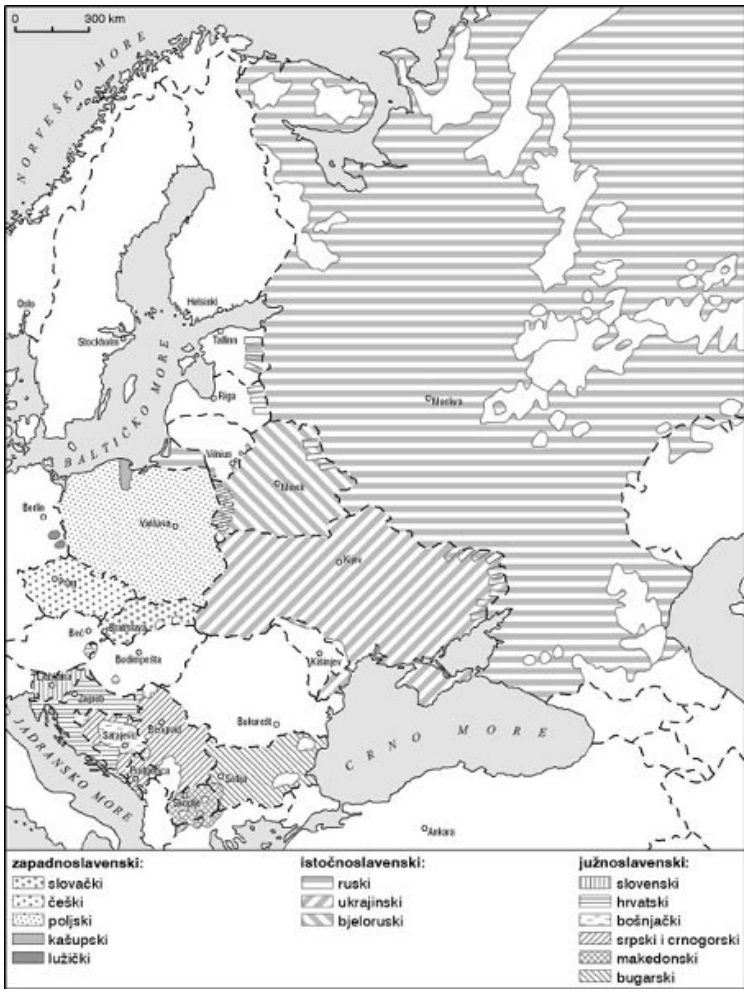
Pravila nizanja istaknutih rečeničnih elemenata u turskom jeziku za jednostavnu tursku rečenicu prema Čauševiću (1996) podliježu istaknutom modelu: subjekt (S) – adverbijalna oznaka

vremena i mjesta (AO) – objekt (O) – predikat (P). S obzirom na istaknuti model turski jezik prema sintaktičkom ustroju kategorizira se kao SOV jezik.

3.2. Hrvatski jezik i gramatika

> *Kao što je ljudsko tijelo u starosti ono isto s kojim se rađamo kao novorođenčad, premda izgrađen većinom od posve različitih stanica, tako i jezik nosilac jedne pisane kulture može kroz povijest ostati isti, čak i ako mu se postupno mijenjaju bitni sastavni elementi. Kontinuitet u razvitku ono je što oboma daje identitet.* (Matasović: 2008, str. 34)

Hrvatski jezik pripada indoeuropskoj jezičnoj porodici i razvio se iz praslavenskog jezika koji se u doba seobe Slavena na područje današnje Hrvatske i Bosne (6. i 7. st.), prema Matasoviću (2008), još uvijek koristio. Hrvatski jezik pripada stoga slavenskoj jezičnoj skupini jezika i tijekom povijesti pokazuje kontinuitet kulture pisanog jezika kojim su se služili Hrvati, iako su se konkretni glasovi i oblici samog jezika tijekom povijesti mijenjali.



Slika 4. Slavenski jezici (https://www.enciklopedija.hr/natuknica.aspx?ID=56592 [pristupljeno 30. 1. 2021.])

Nekoliko stoljeća od dolaska Hrvata na današnja područja hrvatski jezik razmjerno se brzo mijenjao, dok su se na području srednjovjekovne hrvatske države istaknula tri narječja: čakavsko, štokavsko i kajkavsko, koja su tada postojala kao zasebni književni jezici, kako objašnjava Raguž (1997). Hrvatski kao suvremeni standardni jezik i službeni jezik Republike Hrvatske nastao je na jednom dijalektu – štokavskom, i to reformama koje su proveli ilirci u prvoj polovici 19. st., iako su težnje prema općem jeziku svih Hrvata, čak i bez samostalne države, jačale još od početka 17. st. Hrvatski je jezik za svoga postojanja bilježen trima pismima – glagoljicom, ćirilicom i latinicom, koja se kasnije zadržala kao službeno hrvatsko pismo.

Hrvatski jezik prema morfološkoj klasifikaciji pripada skupini fuzijskih jezika, koje karakterizira promjena oblika riječi u morfološkim paradigmama, odnosno primjena fuzioniranih flektivnih morfema u izražavanju gramatičkih kategorija.

> *Za hrvatski i druge jezike, fleksija (promjena, mijenjanje) imenskih riječi (imenica, pridjeva, zamjenica, brojeva) po padežima zove se sklonidba (sklanjanje, deklinacija), fleksija glagolskih riječi po licima (osobama) zove se sprezanje (konjugacija), i za pridjeve i priloge imamo još stupnjevanje (komparaciju).* (https://proleksis.lzmk.hr/1571/ [pristupljeno 31. 1. 2021.])

Hrvatski jezik karakteriziraju i glasovne promjene koje mogu biti uvjetovane fonološki, s ciljem olakšavanja izgovora, ili morfološki, prilikom već istaknutih promjena riječi i tvorbi novih riječi. Riječi hrvatskoga jezika po svom se osnovnom značenju, prema Težaku i Babiću (1994), mogu svrstati u deset vrsta: imenice, zamjenice, pridjevi, brojevi, glagoli, prilozi, prijedlozi, veznici, čestice i usklici. Dodatna podjela odnosi se na promjenjivost koja može biti izazvana potrebom da se označe neka od sljedećih obilježja:

- *rod, tj. muška, ženska ili neutralna određenost*
- *broj, tj. količina*
- *padež, tj. odnos prema drugim riječima obično povezan sa službom riječi u rečenici*
- *lice, tj. odnos prema onome tko govori*
- *vrijeme i način.* (Težak, Babić: 1994, str. 77)

Obilježje hrvatskog jezika koje je potrebno istaknuti jest kategorija roda, koja u jeziku odražava prirodne razlike u spolu živih bića – ljudi i životinja – ili nepostojanje spola kod stvari i pojava. Riječi muškog roda u pravilu označavaju ili se odnose na muško biće, jednako kao što to čine riječi ženskog roda za žensko biće, dok riječi srednjeg roda označavaju mlado biće kojemu ne treba isticati spol ili stvari i pojave koje u prirodi nemaju spola, kako navode Težak i Babić

(1994). Isti izvor jasno ističe da se gramatički rod ipak ne podudara nužno sa spolom; neke riječi koje označuju muško gramatički mogu biti ženskog ili srednjeg roda ili čak mogu imati oba, isto vrijedi i za potvrđene riječi koje označuju žensko ili stvari i pojave bez spola; rod se u takvim slučajevima može istaknuti upotrebom pokaznih zamjenica. Kod riječi hrvatskog jezika razlikuju se dva broja – jednina i množina, a očituju jedinke ili zbir u jednini i mnoštvo. Hrvatski jezik, prema Težakovoj i Babićevoj kategorizaciji (1994), ima sedam padeža: nominativ, genitiv, dativ, akuzativ, vokativ, lokativ i instrumental, a prema njima se mijenjaju imenice, zamjenice, pridjevi i brojevi. Red rečeničnih dijelova u hrvatskom jeziku koji je neobilježen jest subjekt – predikat – objekt, atribut ispred imenice, posvojni genitiv iza nje, što hrvatski jezik prema sintaktičkom ustroju kategorizira kao SVO jezik. Težak i Babić (1994) objašnjavaju kako se riječi u rečenici mogu slagati po obliku, odnosno gramatički, ili po smislu, odnosno logički.

> *Slaganje riječi u rečenici po rodu, broju, padežu i licu zove se sročnost ili kongruencija.* (Težak, Babić: 1994, str. 247)

3.3. Usporedba jezičnog para

Prema istaknutim obilježjima turskog i hrvatskog jezika nameće se činjenica da je najočitija razlika istaknutog jezičnog para u morfološkoj klasifikaciji. Hrvatski naime pripada skupini fuzijskih jezika koji se za izražavanje gramatičkih kategorija služe procesom stapanja nekoliko gramatičkih kategorija u jednom morfemu, dok turski spada u skupinu aglutinativnih jezika u kojima morfemi generalno označavaju samo jednu gramatičku kategoriju. Za svaku gramatičku kategoriju prilikom procesa aglutinacije upotrebljavaju se isti sufiksi u više alomorfnih oblika,[15] što turski, za razliku od hrvatskog, čini jezikom koji, prema Čauševiću (1996), ne poznaje izuzetke u deklinaciji i konjugaciji. Razlike u gramatičkom i obavijesnom ustrojstvu rečenice te redoslijedu rečeničnih dijelova također nisu zanemarive. Turski je SOV jezik, dok je hrvatski SVO jezik. U turskom jeziku nova obavijest uvijek dolazi ispred predikata, a u hrvatskom uvijek iza njega. Dodatna je razlika u jezicima višeznačnost gramatičkih morfema u hrvatskom jeziku, što u turskom nije slučaj jer gramatički morfemi nose jednu funkciju i jedno značenje. Čaušević (2019) ističe kako postoji čak 17 strukturno-tipoloških obilježja različitosti u ustroju turskog jezika naspram hrvatskog, od kojih su neke istaknute u Tablici 1. Autor ih navodi kao

[15] Takvi oblici odnose se na ozvučenje finalnog fonema ili transformaciju finalnog /k/ u / ğ/.

relevantne i otežavajuće u slučajevima učenja i usvajanja hrvatskog jezika kod izvornih govornika turskog jezika, međutim mogu poslužiti i za usporedbu istaknutog jezičnog para.

HRVATSKI	TURSKI
1. Morfonološka razina	
Glasovne promjene na granici osnove i sufiksalnog morfema: nom. jd. *oko* / nom. mn. *oč-i*	Nepromjenjivost korijena/osnove, odnosno tvorbenih i gramatičkih morfema: *göz / göz-ler* (oko – mn.)
2. Morfološka razina	
Različiti tipovi sklonidbe imenica, zamjenica i pridjeva muškoga, ženskoga i srednjega roda; šest glagolskih vrsta s ukupno 29 razreda	Jedna deklinacija i jedna konjugacija, nema gramatičkih iznimaka.
3. Gramatička kategorija roda	
+	-
4. Sintaktička razina	
Zavisni tagmem u spojevima riječi dolazi ili ispred ili iza glave sintagme: *[velik(i)] prozor* *prozor [sob-e]*	Ljevostrano nizanje svih odredbenica u odnosu na određenicu: *[büyük] pencere* *[oda-m] pencere-si* Soba – gen. prozor-njezin – posvojni sufiks, 3. lice jd.
SVO jezik: *Ivan voli Petru.*	SOV jezik: *Ivan Petra´yi seviyor*. Ivan Petra – ak. voli – prezent, 3. lice jd.
Zavisne surečenice uvrštavaju se u temeljnu s pomoću subjunktora, strukturni model [SVO + [subjunktor + (S)VO]]: *[Mirna zna [da sam otišao na more]]*	Imenske, participske i konverbne skupine riječi (tzv. infinitni ili nelični glagolski oblici) uvrštavaju se u ustroj temeljne rečenice po modelu [S + [infinitna skupina] + V] *[Mirna, [denize gittiğim]-i biliyor]*. [Mirna – gen. [more – dat. odlazak-moj – particip perfekta, posvojni sufiks, 1. lice jd.] – ak. zna]

Tablica 1. Neke tipološke razlike između hrvatskog i turskog jezika (po uzoru na Čaušević: 2019, str. 74)

Kako je vidljivo u Tablici 1., turski jezik, za razliku od hrvatskog, ne poznaje kategoriju gramatičkog roda, a i poimanje kategorije broja, prema ranije istaknutim značajkama jezika

pojedinačno, bitno se razlikuje u odabranom jezičnom paru. Iz tablice je dodatno vidljiv i različit sintaktički ustroj rečenica:

> *U hrvatskim spojevima riječi sročne odredbenice[16] prethode određenici[17], a nesročne dolaze iza nje. Tipično strukturno obilježje turske sintagme jest da sve odredbenice stoje ispred određenice.* (Čaušević: 2019, str. 76)

Čaušević (2019) objašnjava kako se istaknuta struktura u turskom i nizanje određenica s lijeve strane odnosi na sve vrste riječi ili sintagme koje se mogu naći u odnosu s odrednicama, što je obrnut redoslijed od hrvatskog jezika. Dodatno, sintaktička veza među njima u hrvatskom ostvaruje se sročnošću, odnosno slaganjem prema gramatičkim kategorijama (rod, broj, padež i sl.), što u turskom nije slučaj; u njemu se svi gramatički morfemi pridružuju određenici. Iako se hrvatski jezik, kao i turski, u sintaktičkim vezama također koristi pridruživanjem i dodatnim upravljanjem, razlika ostaje u već istaknutom pozicioniranju riječi, što je za turski ponovo karakteristično ljevostrano. Što se pak kategorizacije glagola na morfološkoj razini tiče, hrvatski jezik ima kategorije vida (svršeni i nesvršeni), stanja (aktiv, pasiv), načina (imperativ, kondicional), roda (muški, ženski, srednji), lica i broja, dok turski sadrži sve istaknute kategorije izuzev vida i stanja. Osobitost hrvatskog jezika, prema Težaku i Babiću (1994), uz mogućnost izricanja vremena, lica i osoba koje vrše s pomoću glagola jest dodatno izricanje podatka o svršenosti i nesvršenosti radnje. Ekvivalenta za vid u turskom jeziku nema, ali se ta kategorija, prema Čauševiću (1996), ipak izražava na sljedeće načine: glagolskim vremenima, perifrastičnom[18] konjugacijom i analitičkim glagolima[19] za izražavanje modalno-vidskih karakteristika radnje. Čaušević (2018), koji se iscrpno bavio usporedbom sintakse tog jezičnog para, dodatno objašnjava njegovu najveću istaknutu razliku – sredstva i načine uvrštavanja zavisnih surečenica u glavnu rečenicu prikazane u Tablici 2..

[16] Odnosi se na glavnu sastavnicu sintagme, dalje u tekstu određenica.
[17] Odnosi se na zavisnu sastavnicu sintagme, dalje u tekstu odrednica.
[18] Perifrastični oblici neizravno ili opisno izriču ono što ostali oblici mogu izravnije ili jednom riječju.
[19] Najčešće se izražava formom gerund + pomoćni glagol.

GRAMATIČKA SREDSTVA		VRSTE REČENICA
I. ASINDETSKO sklapanje nizanjem rečenica bez uporabe veznika	jukstapozicija, intonacija, strukturni paralelizam, leksički korelati i dr.	A) asindetske rečenice s obilježjima implicitne koordinacije B) asindetske rečenice s obilježjima implicitne subordinacije
II. SINDETSKO	veznici	
1. sklapanje dviju rečenica u nezavisnosloženu rečenicu s pomoću veznika	konjunktori	nezavisnosložene rečenice
2. sklapanje uvrštavanjem ishodišne rečenice u temeljnu (glavnu) rečenicu	subjunktori	zavisnosložene rečenice

Tablica 2. Načini sklapanja nezavisnosloženih i zavisnosloženih rečenica u hrvatskom i turskom (po uzoru na Čaušević: 2019, str. 1)

Jasno je da jezični par turski i hrvatski ima različitosti koje među istaknutim gramatičkim i lingvističkim strukturama zacijelo uključuju i drugačiju percepciju izvanjezične stvarnosti sagledane kroz jezik.

4. Jezični stilovi i njihove karakteristike

Iako svaki standardni jezik ima stilski neutralan dio, svakako ima i obilježje višefunkcionalnosti, svoje zakonitosti i ono što ga čini posebnim.

> *To znači da je njegova funkcija višestruka, da je on jezikom književnosti, ureda, medija i znanosti te da se u skladu s tim funkcijama raslojava na funkcionalne stilove.* (Mihaljević: 2003, str. 39)

Prema Siliću (2006), hrvatski standardni jezik definira se kao *jezik hrvatske polifunkcionalne javne komunikacije* koji se različito koristi u različitim okruženjima; od znanosti, ureda, novina, radija i televizije, književnosti do svakodnevnog govora jezik preuzima različitu funkciju. Kao što je već istaknuto, svaki od stilova ima svoje zakonitosti, a razlikuju se međusobno u svom odnosu prema normi odnosno, kako Mihaljević (2003) objašnjava, prema stupnju dopuštene individualne slobode. Odgovor na pitanje zašto je standardni jezik polifunkcionalan jasno daje Silić (2003):

> *Zato što je život koji prati polifunkcionalan. U skladu s time (pogotovu u ovome vremenu) moramo reći nešto što je vrlo važno za (su)odnos jezik – život: nije život onaj koji prati jezik, nego je jezik onaj koji prati život. Nikako ne život poslije jezika, nego jezik poslije života.* (Silić: 2003, str. 38)

Funkcionalne stilove hrvatskog jezika, prema Siliću i njegovoj terminologiji (2006), možemo podijeliti na znanstveni, administrativno-poslovni, novinarsko-publicistički, književnoumjetnički (beletristički) i razgovorni, a dalje u tekstu bit će pobliže istaknuta i opisana samo dva koja su predmet ovog rada i dio istraživanja.

4.1. Novinarsko-publicistički stil

Silić (2006) ovaj stil ističe kao najsloženiji funkcionalni stil hrvatskog standardnog jezika, koji obuhvaća područje novinarstva i publicistike. Autor ističe razliku između tih dvaju pojmova kako bi opravdao široku primjenu predmetnog stila. Neke od zadaća koje ovaj stil treba ispuniti svakako su obavještavanje čitatelja, slušatelja ili gledatelja o suvremenim zbivanjima, širenje znanja o društvu, kulturi, politici, vjeri i dr., usmjereni rad na pridobivanju ljudi za kakvu aktivnost, poučavanje, odgajanje, pa i zabava. U skladu s time Silić (2006)

kategorizira funkcije ovog stila u okviru općih funkcija novinarskih medija, pa tako razlikujemo informativnu, propagandnu, popularizatorsku, prosvjetiteljsku, agitativnu, pedagošku i zabavnu. S obzirom na djelatnost koju obuhvaća, jezična sredstva kojima se pritom služi svakako su prikladna, a uključuju uporabu neutralnih (stilski neobilježenih) i ekspresivnih (stilski obilježenih) jezičnih sredstava. Primjena neutralnih sredstava načelno je zastupljenija u informativnim, popularizatorskim, prosvjetiteljskim i pedagoškim žanrovima ili, konkretno, u vijestima, komentarima, recenzijama, intervjuima, anketama i reportažama. Što se pak ekspresivnih sredstava tiče, ona su primjerenija za propagandne, agitativne i zabavne žanrove: kratke priče, eseje, feljtone, pamflete, parodije, groteske i sl. Istaknuti žanrovi razlikuju se dodatno prema strukturi, ako govorimo o vrsti medija kojim se prenose poruke javnosti, pa se dodatno specificira ju kao radijski ili televizijski za razliku od novinskog, a zajednički im ostaje opći sadržaj. Neke od stilskih figura kojima se služi ovaj stil, ističe Silić (2006), jesu: poredba, metonimija, alegorija, simbol, aluzija, antifraza, antiteza, kontrast, paradoks, emfaza, antonomazija, eufemizam, ironija, litota, perifraza i igra riječima. Istaknute figure značenje dobivaju tek u odnosu s ostalim članovima rečenice, no i sama struktura rečenice može služiti kao „stilska figura“: ponavljanje određenih dijelova rečenice ili cijele rečenice, korištenje retoričkih pitanja, grafičko obilježavanje teksta i sl. U novinarsko-publicističkom stilu svakako je bitan i najprije privlači pozornost čitatelja ili gledatelja upravo naslov, stoga on mora biti snažan i izazovan te pobuditi zanimanje. Silić (2006) navodi kako po načinu i sadržaju naslove možemo podijeliti na nominativne, informativne i reklamne. Prvi od istaknutih imenuju sadržaj upotrebom imenica i pridjeva kao najzastupljenijih vrsta riječi, a ako i dođe do upotrebe glagolskih oblika i riječi u tom kontekstu, oni su pretvoreni u glagolske pridjeve i odnosne rečenice. Informativni naslovi prenose sadržaj, pa je, za razliku od prethodnog, zastupljenija upotreba glagola, glagolskih oblika, glagolskih vremena, upitnih riječi i sl. Reklamni naslovi nude određeni sadržaj, stoga upotreba sugestivnih riječi i sugestivnih sredstava nije neprimjerena, već je čak i poželjna, a neke od jezičnih formi kojima se sugestivnost može postići svakako su imperativi, upitnici, uskličnici, crtice itd. Ustaljene izraze koji se pojavljuju u novinarsko-publicističkom stilu i čine ga prepoznatljivim Silić (2006) naziva žurnalizmima, a u ovom konkretnom stilu izvor im može biti u politici, sociologiji, ekonomiji, pravu i drugim srodnim područjima. Još jedna od značajki koja se može vezati za ovaj jezični funkcionalni stil svakako je upotreba internacionalnih riječi, koja je postala činjenica i snaga istaknutog stila.

Ovaj stil svakako treba poštivati jezičnu normu i normu ostalih stilova, a jezikom se koristiti racionalno, ekonomično i učinkovito.

4.2. Književnoumjetnički stil

Najindividualniji funkcionalni stil svakako je književnoumjetnički jer dopušta najveću slobodu, odnosno ima najblaži odnos prema normi.

> *Književnoumjetnički stil jezične činjenice (inačice) o kojima je riječ osmišljava i preosmišljava. On ne bira između njih kao postojećih, (s)tvarnih, danih, nego između njih kao mogućih. Književnik je dakle izravno 'poslušan' jezičnim, lingvističkim, normama, tj. normama jezika kao sustava, a ne društveno-jezičnima, sociolingvističkim, normama, tj. normama jezika kao standarda.* (Silić: 2006, str. 100)

Silić (2006) razlikuje jezik kao sustav i jezik kao standard, navodi ih kao dva ostvarenja jezične djelatnosti koja slijede drugačija pravila i zakonitosti, s jedne strane jezična, s druge strane i jezična i izvanjezična. Ovaj funkcionalni stil hrvatskog jezika, imajući to na umu, otklanja se od jezika kao standarda i priklanja se više jeziku kao sustavu. Gutić (2009) obilježja književnoumjetničkog stila promatra kroz želju književnika za stvaranjem izvornog i neponovljivog umjetničkog djela, što podrazumijeva veći stupanj slobode u izražavanju. Neovisnost i naglašenost već spomenute slobode vidljiva je u Tablici 3.

KNJIŽEVNOUMJETNIČKI STIL	**STANDARDNI JEZIK**
individualan	kolektivan
neimitativan	imitativan
nenormiran	normiran
uključuje lokalizme, dijalektizme, arhaizme, barbarizme	isključuje lokalizme, dijalektizme, arhaizme, barbarizme
neograničen izbor leksičkih i sintaktičkih jedinica	ograničen izbor leksičkih i sintaktičkih jedinica

Tablica 3. Razlike između književnoumjetničkog stila i jezika kao standarda (Barić i dr.: 1999 prema Gutić: 2009, str. 51)

Kako objašnjava Gutić (2009), ovaj funkcionalni stil karakteriziraju slikovitost, ritmičnost i bogatstvo riječi te je način izražavanja kojem pribjegavaju pjesnici, pripovjedači, dramatičari, putopisci i dr. Iako se prema Siliću (2006) često ističe kako književnici ne stvaraju u skladu s gramatikom i pravopisom, autor se ne slaže s tom konstatacijom:

Nije književniku do toga da narušava pravopisno-pravogovorno-gramatičko-leksička pravila (standardnog) jezika, nego do toga da svojim jezikom kaže ono što namjerava kazati. (Silić: 2006, str. 104)

Mihaljević (2002) kao obilježje književnoumjetničkog stila na leksičkoj razini navodi upotrebu poetizama, posuđenica i sinonima, dok se na tvorbenoj razini istaknuti funkcionalni stil služi tvorbenim mogućnostima za stvaranje novih riječi.

Jasno je da ovaj stil uzima od drugih stilova, pritom pazeći na originalnost i drugačiju primjenu onoga što je uzeto kako bi osmislio nešto svoje i originalno, u skladu s lingvističkim zakonitostima.

4.3. Usporedba odabranih jezičnih stilova

Ako se oslanja na upotrebu ekspresivnih jezičnih obilježja, književnoumjetnički stil neznatno je bogatiji od novinarsko-publicističkog, međutim razlika se krije u samom načinu upotrebe.

U novinarskome se tekstu smisao jezičnog sredstva nalazi u tekstu, a u književnoumjetničkome u podtekstu. Smisao je u novinarskome stilu kazan neposredno, a u književnoumjetničkome posredno ('između redaka'). (Silić: 2006, str. 77)

Još jedna od razlika između tih dvaju stilova koja se ističe jest poštivanje odnosno nepoštivanje norme koja je logična s obzirom na zadaće koje stilovi imaju.

Jezična se pravila u književnoumjetničkome stilu ne distanciraju od izvanjezične stvarnosti onako kako se distanciraju jezična pravila u drugim funkcionalnim stilovima. (Silić: 2006 str. 106)

Teško je ta dva stila i uspoređivati s obzirom na raznovrsnost novinarsko-publicističkog stila koja u nekim svojim formama primjenjuje razinu slobode koja je bliska književnoumjetničkom stilu, dok je u drugim ostvarenjima ta ista sloboda zatrta.

5. Vrednovanje odabranog alata za strojno prevođenje i analiza prijevoda

Istraživački dio ovog rada sastoji se od analize odabranih prijevoda tekstova s turskog na hrvatski jezik upotrebom *online* alata Google Prevoditelja. Glavni je cilj istraživanja prikupljanje informacija o vrstama pogrešaka koje se mogu očekivati prilikom prevođenja s turskog na hrvatski jezik s pomoću Google Prevoditelja te stjecanje uvida u trenutačno stanje toga alata za strojno prevođenje na konkretnom jezičnom paru. U sklopu tako definiranog općeg cilja pokušat će se ispitati ima li povezanosti između tipova teksta, ovisno o pripadnosti određenom funkcionalnom stilu, i kategorije pogrešaka u izlaznim podacima strojnog prijevoda.

S obzirom na cilj ovo se konkretno istraživanje, prema Lamza Posavec (2004), može svrstati u deskriptivna ili opisna istraživanja koja omogućavaju shvaćanje i opisivanje bitnih karakteristika određene pojave ili procesa i njihovu primjenu u praksi. Informacije dobivene ovim istraživanjem tako mogu poslužiti unapređenju sustava za strojno prevođenje u ovom konkretnom jezičnom paru, čime rad dobiva i pragmatičnu svrhu. Brojna istraživanja vezana za vrednovanje strojnog prevođenja, osobito jezika koji imaju širu primjenu, provedena su na svjetskoj razini, međutim sve je više istraživanja na nacionalnoj razini, koja uključuju i hrvatski jezik, a neka od istaknutijih vidljiva su u radovima Simeon (2008), Seljan i dr. (2011), Brkić i dr. (2011 i 2013), Dunđer (2015), Pavlović (2017) i Ljubas (2017 i 2018). Sva istaknuta istraživanja provedena su gotovo isključivo na jezičnom paru engleski i hrvatski, izuzev Ljubas (2017 i 2018), čije obje analize uključuju jezični par švedski-hrvatski.

Pojmovno određenje predmeta samog istraživanja pokriveno je u teorijskom dijelu rada, izuzev konkretnijeg pristupa samom pojmu vrednovanja strojnog prevođenja. Kako ističe Simeon (2008), iako općeprihvaćena metoda evaluacije strojnog prevođenja još uvijek nije usustavljena, vrednovanje kvalitete prijevoda općenito, bez obzira na to je li prijevod ljudski ili strojni, kompleksan je proces, ali potreban za osiguravanje kontinuiranog unapređenja sustava i alata za prevođenje. Ljubas (2020) podupire istaknutu tezu objašnjavajući kako su se gotovo paralelno s razvojem teorije prevođenja i sustava za strojno prevođenje polako razvijale i metodologije vrednovanja tih sustava. Prema Brkiću i dr. (2011), metode vrednovanja mogu biti automatske ili ljudske, od kojih se prva temelji na imitaciji druge. Prema istom autoru, ljudska evaluacija može se smatrati „zlatnim standardom“ vrednovanja, no s obzirom na činjenicu da jedna metoda imitira ponašanje druge, autor obje smatra subjektivnima. Uzimajući u obzir faktore poput uštede vremena i novca, autor ipak prednost daje automatskim sustavima

za vrednovanje strojnog prevođenja u odnosu na ljudske. Kad je riječ o strategijama vrednovanja u području strojnog prevođenja, ističu se dvije prema Trujillo (1999), kako navodi Simeon (2008) u svom radu: *black box* i *glass box* vrednovanje. Najvažnija je razlika između istaknutih kategorija u činjenici da je *black box* prilagođeniji korisnicima i prevoditeljima, dok je *glass box* relevantniji istraživačima i tvorcima sustava za prevođenje. Kad se radi o dimenzijama samog vrednovanja, Arnold i dr. (2002) navode tri: jasnoću, odnosno razumljivost prijevoda, vjernost prijevoda originalu i naposljetku prirodu pogrešaka koje sustav generira.

Za potrebe ovog istraživanja koristit će se treća dimenzija, kojom će se u analizi sustavno prebrojiti generirane prijevodne pogreške te se klasificirati i opisati njihove kategorije. Prema Simeon (2008), analiza pogrešaka ne podrazumijeva samo kvantifikaciju već i njihovu lingvističku obradbu. Hutchins (1992) drži da je u većini slučajeva vrednovanja strojnog prevođenja brojenje pogrešaka najkorisnije i najpraktičnije rješenje za dobivanje informacija o kvaliteti strojnog prijevoda. Analizom pogrešaka moguće je dobiti podatke o količini posla koji je potreban za korekciju sirovih izlaznih podataka strojnog prijevoda. Sam postupak analize podrazumijeva ljudskog vrednovatelja zaduženog za brojenje svakog dodavanja ili brisanja riječi, svake zamjene riječi nekom drugom riječi i svakog slučaja primjene odgovarajuće fraze ciljnog jezika te naposljetku izračun postotka potrebnih ispravaka na razini čitavog teksta. Autor dodaje kako puko brojenje pogrešaka ipak nije dovoljno, već postoji potreba za dodatnom klasifikacijom utvrđenih pogrešaka prema relativnoj težini i neizostavno na lingvističkoj razini. Kao nedostatak ovog tipa analize ističe nedostatak objektivnosti jer procjena i određivanje granica pogrešaka u cijelosti ovisi o onome tko vrednuje prijevod.

5.1. Metodologija

Istraživački dio ovog rada oslanja se na vrednovanje strojnoprevoditeljskog sustava za jezični par turski-hrvatski s pomoću prethodno definirane analize pogrešaka. Ovo istraživanje bit će podvrgnuto kvalitativnoj analizi podataka prikupljenih kvantitativnom metodom, točnije generirane pogreške bit će pobrojene i kategorizirane prema unaprijed određenim kriterijima, a u nedostatku eventualne postojeće kategorizacije prema novim kriterijima koji proizađu iz analizirane građe. Ono što Lamza Posavec (2004) vidi kao prednost kvantitativnih metoda u odnosu na kvalitativne karakterizira se u većoj preciznosti prikazanih rezultata, mogućnostima generalizacije zaključaka i otkrivanja skrivenih odnosa unutar istraživane pojave.

Analiza je provedena na dva stilski različita teksta: književnoumjetnički stil predstavlja odabrano poglavlje knjige *Ruhi Mücerret,* autora Murata Menteşa,[20] a novinarsko-publicistički stil članak naslova *Neobična borba Švedske s bolesti Covid-19,* autora Ömera Faruka Aydemira, objavljen 28. travnja 2020. na službenim stranicama državne novinske agencije *Anadolu Agency*[21]. Kad je riječ o tematici, odabrani roman može se svrstati u kategoriju pustolovnog romana s ratnom tematikom, dok se članak bavi aktualnom temom epidemije koronavirusa.

Originalni tekstovi na turskom jeziku sadrže približno 12 kartica teksta[22], jednako kao i strojni prijevodi; konkretan broj znakova s razmacima po tekstu i pripadajućem prijevodu prikazan je u Tablici 4., s dodatnom referentnom točkom konvencionalnog prijevoda[23].

	Književnoumjetnički tekst		Novinarsko-publicistički tekst	
	Znakovi s razmacima	Pojavnice	Znakovi s razmacima	Pojavnice
Original	8448	1067	9057	1111
Strojni prijevod	8488	1411	8613	1299
Konvencionalni prijevod	9169	1517	8534	1336

Tablica 4. Prikaz broja znakova s razmacima i pojavnica u odabranim tekstovima za analizu prema vrsti i jezičnom stilu

Istraživačka pitanja glase: Pojavljuju li se u ovom jezičnom paru pogreške koje nisu očekivane u strojnom prevođenju? Utječe li jezični stil na veći broj pogrešaka u strojnom prevođenju? Opća hipoteza ovog istraživanja jest da tip teksta utječe na broj i vrstu pogrešaka u prijevodu. Kako je cilj istraživanja prikupljanje informacija, konkretnije hipoteze nisu unaprijed postavljene.

Klasifikacija pogrešaka prema kojoj će biti raspodijeljene zabilježene pogreške preuzeta je od Simeon (2008):

[20] *Suvremeni turski autor često uspoređivan s Tarantinom i Palahniukom, u svom trećem romanu donosi priču o Ruhiju Mudžeretu, posljednjem živućem veteranu turskog Rata za nezavisnost s početka 20. stoljeća.* (https://katalog.kgz.hr/pagesResults/bibliografskiZapis.aspx?selectedId=995001832&AspxAutoDetectCookieSupport=1)

[21] Preuzeto s https://www.aa.com.tr/tr/analiz/isvecin-kovid-19la-sira-disi-mucadelesi/1821173.

[22] 1 kartica teksta = 1500 znakova s razmacima

[23] Autorica prijevoda magistra je turkologije i anglistike koja radi kao *freelance* prevoditelj i izvorni je govornik hrvatskog jezika.

- pravopisne pogreške
- pogreške u redu riječi
- morfosintaktičke pogreške
- stilske pogreške
- leksičke pogreške (pogrešno odabran prijevodni ekvivalent)
- neprevedene riječi
- izostavljene riječi
- umetnute riječi.

Pretpostavka je da će odabrana klasifikacija biti prikladna za odabrani jezični par, a u analizi će biti pobliže objašnjeno što svaka kategorija podrazumijeva na primjerima iz samog istraživanja. Po potrebi će zadana klasifikacija biti revidirana po uzoru na Ljubas (2017), ako se postojeća pokaže neadekvatnom.

5.2. Postupak analize

Na samom početku istraživanja odabrana su dva teksta na turskom jeziku, koja je zatim isti prevoditelj preveo na hrvatski jezik kako bi postojala referentna točka u analizi strojnog prijevoda. Za istraživanje su odabrana dva teksta različitih jezičnih stilova: jedan književnoumjetničkog stila, drugi novinarsko-publicističkog, koji svojim obilježjima, možemo pretpostaviti, stvaraju najviše poteškoća sustavima strojnog prevođenja. Spomenuta dva teksta na turskom jeziku, pojedinačno duljine približno šest kartica, čine prvi korpus analiziranih tekstova. Drugi korpus čine konvencionalni prijevodi odabranih tekstova, koji služe za usporedbu sa strojnim prijevodom, dok se treći korpus sastoji od strojnih prijevoda prvog korpusa s pomoću alata Google Prevoditelja. Bitno je istaknuti kako je strojni prijevod odabranih tekstova generiran u siječnju 2021. godine, kada spomenuti alat radi na principu neuronskog strojnog prevođenja u kombinaciji sa statističkim, te da su navedeni korpusi jedini izvor informacija prikupljenih unutar istraživanja.

Kako je već ranije navedeno, klasifikacija pogrešaka koja će biti korištena u ovom istraživanju preuzeta je od Simeon (2008). Uz klasifikaciju pogrešaka po tipu razlikovane su dodatno dvije kategorije pogrešaka prema specifičnoj težini, ponovo po uzoru na Simeon (2008): teže pogreške (kategorija I) i lakše pogreške (kategorija II).

Prvu kategoriju, koja obuhvaća teže pogreške, tj. one za koje se opravdano može pretpostaviti da znatno otežavaju razumijevanje teksta, čine: zadržavanje riječi iz izvornog jezika, izostavljene riječi, suvišne riječi i pogrešan leksički odabir.

U drugoj su kategoriji pogreške koje u manjoj mjeri utječu na razumijevanje teksta: pravopisne, morfosintaktičke i stilističke pogreške te pogrešan red riječi. (Simeon: 2008, str. 123)

Dodatno je potrebno detaljnije opisati sve odabrane kategorije kako bi se raspodjela evidentiranih pogrešaka mogla što točnije odraditi.

Kategorija I	Kategorija II
leksičke	pravopisne
neprevedene	red riječi
izostavljene	morfosintaktičke
umetnute	stilske

Tablica 5. Raspodjela vrsta pogrešaka po kategorijama težine

Iz Tablice 5. vidljiva je raspodjela prethodno klasificiranih pogrešaka u dvije kategorije prema težini, odnosno utjecaju na razumijevanje i jasnoću teksta. Leksičke pogreške, koje su opravdano svrstane u kategoriju I, svakako su po težini najznačajnije jer pogrešnim odabirom leksičke jedinice prijevodi u ciljnom jeziku mogu biti u potpunosti nerazumljivi. Stupanj razumljivosti, nažalost, i dalje je subjektivno utemeljen za potrebe ovog konkretnog istraživanja jer ovisi o vrednovatelju samog prijevoda. Analizu olakšava mogućnost konzultiranja s konvencionalnim prijevodom u slučaju nedoumice. Ostatak pogrešaka unutar te kategorije poprilično je jasan i odnosi se na pogreške strojnog prevoditelja koji ne prevede određenu riječ ili sintagmu u ciljni jezik, već ostavi riječ u izvornom jeziku, no može se dogoditi da ona bude prevedena na međujezik[24]. Izostavljene riječi obilježava postojanje izvornog oblika bez prijevodnog ekvivalenta u ciljnom jeziku – strojni prevoditelj „zaboravi“ prevesti riječ ili rečenicu. Prema opisu, može se zaključiti kako ta vrsta pogrešaka više odgovara vrednovanju konvencionalnog prijevoda, no s obzirom na njezino zadržavanje u novijim istraživanjima na području vrednovanja strojnog prevođenja može upućivati na pogreške u sravnjivanju korpusa. Antipod istaknutoj vrsti pogrešaka unutar iste kategorije čine umetnute riječi koje su ponekad

[24] U slučaju Google Prevoditelja međujezik je engleski jezik jer se u njemu ne treniraju izravno jezični parovi u kojima nema engleskoga.

potrebne za odgovarajući prijevodni ekvivalent. Jednako kao prethodna vrsta pogrešaka, može upućivati na treniran sustav koji prema korpusima iz kojih uči preuzima opisne prijevodne ekvivalente tamo gdje je potrebno.

Od pogrešaka iz kategorije II najjednostavnije je opisati pravopisne pogreške, a one se odnose na pogrešno napisane riječi, interpunkciju, mala i velika slova. Od predmetnog jezičnog para potencijalno se može očekivati pojava pravopisnih pogrešaka u primjerima zavisnosloženih rečenica, a vezano za pozicioniranje zareza ili u pisanju velikog i malog slova kod višečlanih sintagmi u nazivima institucija. Dodatno, što se iskustveno ističe, turski jezik ima više općih imenica koje su istovremeno vlastita imena, pa tu potencijalno može doći do pogrešnog tumačenja sustava i pogreške u odabiru. Svakako je bitno istaknuti da pravopisne pogreške u manjem obujmu utječu na samo značenje prijevoda, ali i da će u ovoj analizi biti uzete u obzir sve prepoznate pogreške, bez obzira na to utječu li na jasnoću prijevoda ili ne, kako bi se zabilježila svaka pogreška sustava.

Red riječi u rečenici kategorija je koju nije potrebno dodatno objašnjavati, no pogreške u tome ipak mogu utjecati na jasnoću ili čak i točnost prijevoda. S obzirom na različitu strukturu jezika i različit, čak i suprotan, linearni poredak riječi, u odabranom jezičnom paru očekuju se pogreške u toj kategoriji zbog potencijalnog pogrešnog prijevoda međuovisnosti jezičnih jedinica, pa samim time i pogrešnog pozicioniranja rečeničnih elemenata. Ta pogreška, može se pretpostaviti, uvelike ovisi o točnosti prijevoda morfosintaktičkih elemenata, duljini same rečenice, višestrukom međusobnom odnosu zavisnosti s većim brojem odredbenica i korištenju umetnutih rečenica u tekstu. I za taj tip pogreške vrijedi pokušaj bilježenja svake pogreške u tekstu bez obzira na to utječe li na značenje ili ne.

Morfosintaktičke pogreške najraznovrsnija su vrsta pogrešaka jer obuhvaćaju i morfološke i sintaktičke pogreške, a odnose se najčešće na pogrešne rečenične oblike, nesročnost, poteškoće u konjugaciji i deklinaciji. Ta vrsta pogrešaka, s obzirom na razlike u obilježjima jezičnog para koji je predmet ove analize, a koje su istaknute u 3. poglavlju, brojčano će svakako biti dominantnija od prethodnih. S obzirom na kategorizaciju i pripadnost lakšoj kategoriji II, ipak se očekuje razumljivost i jasnoća prijevodnih ekvivalenata.

Kad je riječ o posljednjoj vrsti pogrešaka iz druge kategorije, stilskoj, ona će uvelike ovisiti o konvencionalnom prijevodu kao referentnoj točki usporedbe i neizbježnom ljudskom faktoru. Definicija tih pogrešaka može se ipak ograničiti na semantički bliske prijevodne jezične inačice, koje potencijalno mogu dovesti do pogrešnog tumačenja kod čitatelja ili na nespretno prevedene

sintagme. U vezi s tom vrstom pogrešaka bit će najviše konzultacija s konvencionalnim prijevodom prilikom analize, što će osigurati veći stupanj objektivnosti u odnosu na ostale prepoznate pogreške. Dodatno, stilski obilježeni tekstovi svakako će doprinijeti količini pogrešaka koje sustav potencijalno generira.

Nakon kreiranja korpusa, klasifikacije i kategorizacije pogrešaka, analiza se svodi na brojenje i sortiranje pogrešaka unutar odgovarajućih vrsta. Nedostatak ove analize, koji je postao jasan već na njezinu samom početku, jest nedostatak objektivnosti. Naime, određivanje pa naposljetku i klasifikacija pogrešaka prilično je subjektivna i podložna jezičnom znanju i shvaćanju vrednovatelja. Dodatno, ponavljanje i kvantifikacija pogrešaka bez obzira na razinu na kojoj se manifestira – riječ, sintagma, rečenica ili čak čitav tekst – može utjecati na zastupljenost određene kategorije.

5.3. Rezultati

S obzirom na postojeću, odabranu klasifikaciju i podatke iz analiziranih korpusa u tablicama 3. i 4. prikazani su dobiveni brojčani rezultati koji otkrivaju ukupno 414 prepoznatih pogrešaka na korpusu od dvanaest kartica teksta, što u prosjeku znači tridesetak pogrešaka po kartici teksta. Dodatno je prikazana razlika u brojkama s obzirom na pripadnost teksta jezičnim stilovima koji su analizirani. Zastupljenost pogrešaka u književnoumjetničkom tekstu s obzirom na kategorije gotovo je izjednačena, s laganom prednošću kategorije I u omjeru 51,82 % naspram 48,18 %. Unutar kategorija, na razini cijelog teksta i na razini pojedinačnih vrsta pogrešaka razlike su ipak znatnije. Očekivano prednjače leksičke pogreške s 35,22 % (87 pogrešaka), slijede morfosintaktičke s 25,10 % (62 pogreške), stilske sa 16,60 % (41 pogreška), neprevedene riječi s 9,72 % (24 pogreške) i 13 izostavljenih riječi (5,26 %) kao najistaknutije. U novinarsko-publicističkom tekstu zabilježeno je mnogo manje pogrešaka općenito, dok je razlika u zastupljenosti kategorija ipak znatnija, i to u korist kategorije II, odnosno *lakše* kategorije u omjeru 61,08 % prema 38,92 %. Na vrhu su brojčano ponovo leksičke pogreške s 31,74 % (53 pogreške), zatim stilske s 22,16 % (37 pogrešaka) i morfosintaktičke s 20,96 % (35 pogrešaka). U odnosu na prvi tekst istaknutiji je pogrešan red riječi u rečenicama s čak 10,78 % (18 pogrešaka), zajedno s pravopisnim pogreškama u postotku od 7,19 % (12 pogrešaka). Što se izostavljenih riječi tiče, podjednako su zastupljene u oba teksta; u novinarsko-publicističkom nešto više s 6,59 % (11 pogrešaka), dok neprevedenih i umetnutih riječi gotovo i nema.

Vrsta pogreške/teksta	Književnoumjetnički	Novinarsko-publicistički
Kategorija II	119	102
pravopisne	8	12
red riječi	8	18
morfosintaktičke	62	35
stilske	41	37
Kategorija I	128	65
leksičke	87	53
neprevedene	24	1
izostavljene	13	11
umetnute	4	0
Ukupan broj	247	167

Tablica 6. Broj pogrešaka za svaki funkcionalni stil po tipovima i kategorijama pogrešaka

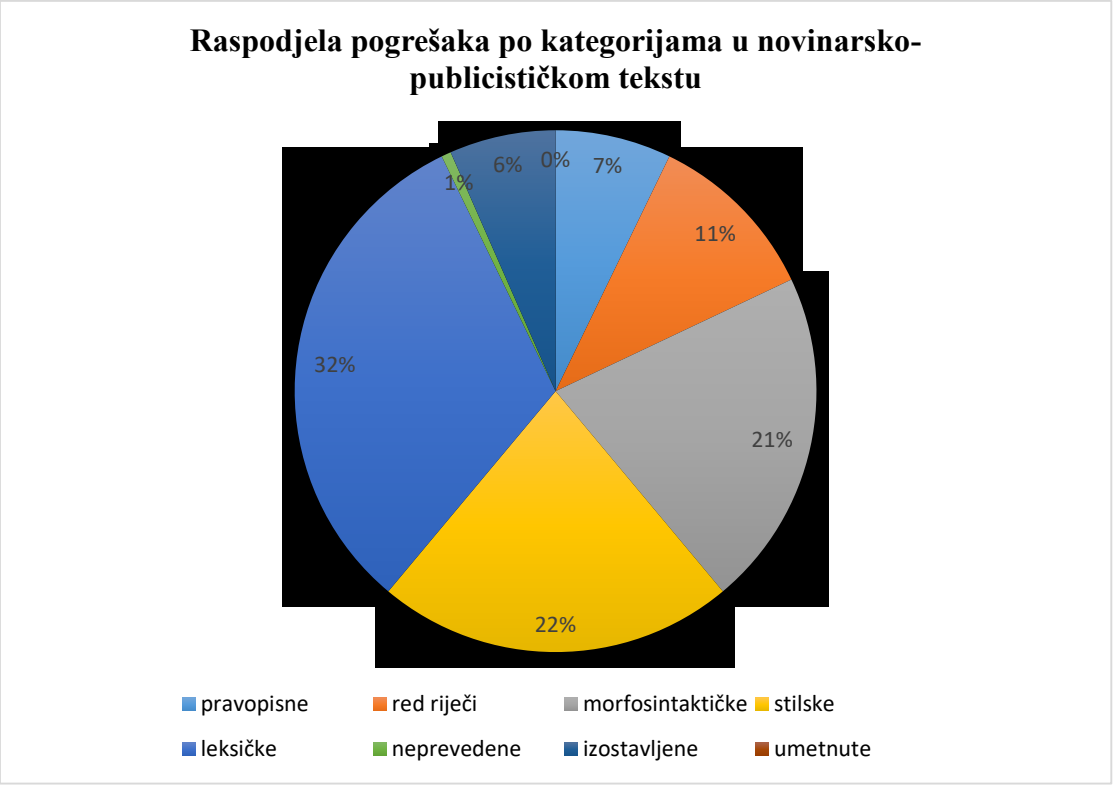
Vrsta pogreške/teksta	Književnoumjetnički	Novinarsko-publicistički
Kategorija II	48,18 %	61,08 %
pravopisne	3,24 %	7,19 %
red riječi	3,24 %	10,78 %
morfosintaktičke	25,10 %	20,96 %
stilske	16,60 %	22,16 %
Kategorija I	51,82 %	38,92 %
leksičke	35,22 %	31,74 %
neprevedene	9,72 %	0,60 %
izostavljene	5,26 %	6,59 %
umetnute	1,62 %	0,00 %

Tablica 7. Postotak pogrešaka za svaki funkcionalni stil po tipovima i kategorijama pogrešaka

Lakoća, koja je kategorično na strani drugog teksta, osjetna je i prilikom same analize i klasifikacije pogrešaka, što se dade zaključiti i iz ukupnog broja pogrešaka po analiziranom tekstu. Iako je brojčano i kategorički književnoumjetnički tekst teži u odnosu na novinarsko-

publicistički, najzastupljenije vrste pogrešaka u oba teksta ipak su očekivane s obzirom na odabrani jezični par: leksičke, morfosintaktičke i stilske prisutne su u gotovo 80 % analiziranog teksta. Ono što se ističe posebice kod novinarsko-publicističkog teksta, a nije toliko zastupljeno u književnoumjetničkom tekstu, jest velik broj pogrešaka u redu riječi ili rečeničnih dijelova. Vrsta pogreške u kojoj književnoumjetnički tekst ipak prednjači pred novinarsko-publicističkim, što je jasnije istaknuto u grafičkim prikazima, jesu neprevedene riječi. Razlog istaknutim diskrepancijama bit će potkrijepljen primjerima i obilježjima tekstova koji su na to utjecali. Analizom se ipak pokazalo točnim da jezični stil utječe na broj pogrešaka u prijevodu, a može se pretpostaviti da je razlog tome dominacija novinarsko-publicističkog stila u usporednim korpusima koji služe za treniranje sustava za strojno prevođenje. Dodatno se može zaključiti kako su već postojeće klasifikacije pogrešaka prikladne za primjenu i na ovom jezičnom paru, čime je već u ovoj fazi dobiven odgovor na postavljeno istraživačko pitanje.

Slika 5. Raspodjela pogrešaka po kategorijama u književnoumjetničkom tekstu



Slika 6. Raspodjela pogrešaka po kategorijama u novinarsko-publicističkom tekstu

Istaknute brojke u daljnjoj će analizi biti potkrijepljene primjerima kako bismo se u jednoj mjeri dotakli i lingvističke obrade te kako bi se dobila jasnija slika stanja s jezičnim parom turskim i hrvatskim. Među pravopisnim pogreškama, posebice u književnoumjetničkom tekstu, ističu se vlastita imena koja su istovremeno i opće imenice u turskom jeziku, što je vidljivo niže u istaknutim primjerima:

Primjer 1.[25]

Janti *misafirin biçimli gövdesini mermilerle imzalıyorum...*

Mecima potpisujem oblikovano tijelo gosta ***Jantija****...*

Potpisujem fit tijelo svog ***zgodnog*** *gosta mecima...*

Primjer 2.

Masum *Cici'ye dönüp*

Obraćajući se ***nevinoj*** *Cici*

[25] Najprije originalni turski tekst, zatim strojni prijevod na hrvatski i naposljetku konvencionalni prijevod. Taj slijed korišten je u svakom sljedećem primjeru. Pogreška je uvijek istaknuta masnim slovima.

*Okrenuo sam se **Nevinom** Slatkišu*

Iz primjera je jasno da je na takav ishod utjecala i pozicija unutar rečenice, odnosno početak koji je uvijek pisan velikim slovom. Temeljem ostalih prijevoda vlastitih imena unutar teksta, pretpostavka je da sustav veliko slovo unutar rečenice prepoznaje kao vlastito ime i poštuje pravopisna pravila. Dodatno se dade zaključiti kako sustav nema dosljednosti u prijevodu jer je istaknuto vlastito ime u Primjeru 2. i ranije i kasnije spominjano u tekstu te prevedeno ispravno.

Kod reda riječi u rečenici unutar novinarsko-publicističkog teksta ističemo primjer u kojem je pozicija riječi unutar rečenice utjecala na značenje, odnosno u Primjeru 3. govori se o porastu u broju slučajeva, ali i o smrtnim ishodima, što nikako nije zanemariva obavijest.

Primjer 3.

*Ancak can kaybı ve vaka sayılarındaki **artışa** bakıldığında…*

*Međutim, s obzirom na gubitak života i **porast** broja slučajeva…*

*No, s obzirom na **porast** smrtnih ishoda i broja slučajeva…*

U sljedećem primjeru pozicija riječi u rečenici utječe na značenje tako da u strojnom prijevodu ističe mogućnost postojanja više autora istog djela, dok se u konvencionalnom vidi što je zapravo bila tema upitne rečenice.

Primjer 4.

*„Avni Bey, Yamyamın Damak Zevki'ni **de** siz mi yazmıştınız?"*

*„Gospodine Avni, jeste li **i** vi napisali Okus kanibala?"*

*„Gospodine Avni, jeste li vi napisali **i** Ukus ljudoždera?"*

Kao što je već istaknuto za morfosintaktičke pogreške, ta skupina svakako je najraznovrsnija, a može nositi poteškoće u sročnosti (rod, broj, padež, lice), oblicima glagola odnosno glagolskim vremenima, redoslijedu riječi i sl. te će neki od primjera biti istaknuti.

U primjerima 5. i 6. radi se o poteškoćama sa sročnosti, odnosno konkretnije, u prvom je samo istaknuta neadekvatna kategorija lica, dok je u drugom problem roda. Turski jezik ne poznaje kategoriju roda, no ipak ima mogućnost izraziti ga na leksičko-sintaktički način imenicama koje označavaju prirodni spol ili socijalni status, poput *Hanım* u istaknutom primjeru.

Primjer 5.

*„ O şeref **bendenize** ait.”*

*„Ta čast pripada **vama**.“*

*„**Meni** je čast.“*

Najčešći primjeri pogrešaka unutar ove vrste upravo je problem sročnosti uz sljedeće istaknute glagolske oblike.

Primjer 6.

*Bu adamla Zebercet **Hanım** vefat ettiği gün **rastlaşmış**, göz göze gelmiştik.*

*S tim smo čovjekom upoznali onog dana kada je Zebercet **Hanım preminuo** i suočili smo se licem u lice.*

*Sreo sam tog čovjeka na dan kada je **gospođa** Zeberdžet **preminula**, gledali smo se oči u oči.*

U sljedećem izdvojenom primjeru značenje nije narušeno, jednako kao i u nekim ranije spomenutima, no prijevod ne izražava radnju koju odabrani oblik odnosno glagolsko vrijeme nosi. U konkretnom slučaju Primjera 7. korišten je gerund *iken*, kojim se u turskom jeziku najčešće izražava radnja koja se događa istovremeno s korelativnim predikatom. Uz istaknuti primjer neadekvatno prevedenog gerunda *iken*, sukladno analizi ističu se u ovakvim situacijama najčešće upravo gerundi u turskom jeziku, dodatno glagolski oblici poput pluskvamperfekta, imperfekta i dr.

Primjer 7.

*Bastonum elimden **düşerken** ağzım bir karış açıldı.*

*Usta su mi se otvorila za centimetar **kad mi je** štap **pao** s ruke.*

***Dok mi je** štap **padao** iz ruke, usta su mi se širom otvorila.*

U Primjeru 8. vidljivo je zapravo više pogrešaka iz morfosintaktičke skupine, no ona koja se izdvaja promijenila je značenje i istaknula sasvim novu temu u rečenici.

Primjer 8.

Hastalığın olabildiğince yavaş yayılmasını *ve böylece sağlık sistemi ve sosyal düzenin çöküşünü engellemeyi* ***hedeflediklerini söyleyen*** *Tegnell ağır kısıtlamaların ve karantina uygulamalarının anlamsız olduğunu ve tarihsel bir dayanağı olmadığını iddia ediyor.*

Rekavši da imaju za cilj spriječiti širenje bolesti što je sporije moguće*, sprječavajući tako kolaps zdravstvenog sustava i društvenog poretka, Tegnell tvrdi da su teška ograničenja i karantenska praksa besmislena i da nemaju povijesnu osnovu.*

Tegnell kaže da im je cilj što sporije širenje bolesti *i sprječavanje kolapsa zdravstvenog sustava i društvenog poretka i tvrdi da su stroga ograničenja i provođenja karantena besmisleni te da nemaju povijesno uporište.*

Iako su stilske pogreške stvar subjektivnog dojma te stila prevoditelja i vrednovatelja, u Primjeru 9. u konvencionalnom prijevodu odabran je prijevodni ekvivalent koji stilski više odgovara novinarsko-publicističkom stilu, što Google Prevoditelj ipak nije uspio. U ovom slučaju nije ni strojni prijevodni ekvivalent nerazumljiv i prenosi isto značenje, no primjerenost kontekstu i stilu u konačnici je jednako bitna. Ova vrsta pogrešaka brojčano je zastupljena upravo zbog poštivanja stilova jednog i drugog teksta te je narušavanje stila promatrano više kao pogreška nego kao narušavanje samog značenja, što je bitno istaknuti. Važno je, u ovom primjeru, naglasiti i mogućnost pogrešaka u ljudskom prijevodu koja može rezultirati manjkavošću u metodi brojanja i usporedbe broja pogrešaka između strojnog i ljudskog prijevoda.

Primjer 9.

Sürü bağışıklığı modeline göre toplumun ***büyük bir çoğunluğuna bulaşan Kovid-19****, insanların hastalığa karşı bağışıklık kazanmasını sağlayacak ve böylelikle salgın kademeli olarak atlatılacak.*

Prema modelu imuniteta stada, ***Kovid-19, koji zaražava veliku većinu društva****, omogućit će ljudima da postanu imuni na bolest i tako će se epidemija postupno prevladavati.*

Prema modelu imuniteta krda ***širenje bolesti Covid-19 na veliku većinu društva*** *omogućit će razvijanje imuniteta ljudi na tu bolest te će se epidemija na taj način postupno prebroditi.*

Sljedeći primjer bogat je leksičkim pogreškama koje mogu biti karakterizirane pogrešnim leksičkim odabirom, terminima ili nazivima, frazama i sl. Što se leksičkog odabira tiče, on se odnosi na istaknutu riječ *dazlak*, koja nosi značenje 'ćelav', dok je odabrani ekvivalent *skinhead* u potpunosti neprihvatljiv te nosi pejorativno značenje i pripadnost specifičnoj supkulturi.

Preostala dva primjera s početka i kraja rečenice fraze su koje strojni prevoditelj nikako nije adekvatno preveo te je rečenica u cijelosti nerazumljiva.

Primjer 10.

*Salondaki **iki dirhem bir çekirdek**, **dazlak** centilmeni **gözüm bir yerden ısırıyord**u.*

***Dva dirhama i sjeme** u dvorani, gospodin **skinhead me negdje grizao za oko**.*

Skockani, ćelavi gospodin u dnevnom boravku odnekud mi je bio poznat.

Uz istaknute leksičke pogreške problem strojnom prevoditelju stvaraju i termini odnosno nazivi institucija kao u sljedeća dva istaknuta primjera. Kao problem u tim slučajevima ponovo se ističe nemogućnost ostvarivanja dosljednosti u samom prijevodu.

Primjer 11.

*İsveç'te Kovid-19'la mücadele **Halk Sağlığı Kurumu** üzerinden yürütülüyor.*

*Borba protiv Kovid-19 u Švedskoj vodi se preko **Agencije za javno zdravstvo**.*

*U Švedskoj se borba protiv bolesti Covid-19 vodi putem **Zavoda za javno zdravstvo**.*

Primjer 12.

*Başkent Stockholm'de vaka sayılarının epey yüksek bir seviyeye gelmesi, hükümete ve **Halk Sağlığı Kurumuna** (Folkhälsomyndighet) yönelik birtakım eleştirilerin daha yüksek sesle dile getirilmesine sebep oluyor.*

*Zbog velikog broja slučajeva u glavnom gradu Stockholmu neke se kritike vlade i **Javne zdravstvene ustanove** (Folkhälsomyndighet) iznose glasnije.*

*Veliki porast broja slučajeva u glavnom gradu Stockholmu razlog je za brojne i sve glasnije kritike usmjerene prema vladi i **Zavodu za javno zdravstvo** (Folkhälsomyndighet).*

U primjerima neprevedenih riječi najčešće su u pitanju vlastita imena koja su istovremeno opće imenice i upitno je je li potrebno izjednačavati imena s ciljnim tekstom, no svakako je u tome potrebno zadržati dosljednost, što sljedeći primjeri neće potvrditi.

Primjer 13.

*„Masum **Cici**."*

*„Nevine **Cici**."*

*„Nevin **Slatkiš**.“*

Istaknuti primjer u sve tri inačice nosi značenje osobe, odnosno njezina imena i prezimena u kojem je strojni prevoditelj odlučio jedan dio prevesti, drugi ne. Dodatno se komplicira prijevod istog vlastitog imena u primjerima 14. i 15., gdje se ponovo ističe nedosljednost u prijevodu te se to ime prevodi vlastitim imenom Inocent[26], hrvatskom inačicom latinskog imena.

Primjer 14.

*„Öyle miii? Oh ne âlâ, şerefyab oldum **Masum** Bey evladım“ diyerek kandilli temenna çektim.*

*„Je li tako? O kako dobro, počašćen sam, gospodine **Inocent**.“*

*„Stvaaarno? Sjajno. Počašćen sam, gospodine **Nevin**“, rekao sam i mahnuo rukom.*

Primjer 15.

***Masum** Cici'ye, en yakın dostumu takdim ettim: „Bu bey Avni Vav, kendisi çok değerli bir psikopattır.“*

*Svog najboljeg prijatelja predstavio sam **Inocentu** Cici: „Ovaj učitelj Avni Vav vrlo je vrijedan psihopata.“*

***Nevinom** Slatkišu predstavio sam najbližeg prijatelja: „Ovo je gospodin Avni Vav, jako cijenjen psihopat.“*

U Primjeru 16. zabilježeno je pak umetanje međujezika ili *interlinguae*, odnosno u slučaju Google Prevoditelja istaknuto vlastito ime prevedeno je na engleski jezik.

Primjer 16.

*Figen **Negatif** eşikte, kısık sesle „Konuğumuz var“ dedi.*

*Figen **Negative** reče tihim glasom: „Imamo gosta“.*

Figen **Negatif** na pragu je tiho rekla, „Imamo gosta“.

Izostavljene riječi kao vrsta pogreške, bez obzira na određenu brojčanu zastupljenost u tekstovima, ipak ne utječu na samo značenje do razine razumijevanja konteksta. Iz Primjera 17. izostavljena sintagma nije neophodna za razumijevanje prevedenog teksta, međutim

[26]Inocent (lat. Innocentius), ime trinaestorice papa i jednoga protupape. Preuzeto s https://www.enciklopedija.hr/natuknica.aspx?id=27497. [pristupljeno 20. 2. 2021.]

klasificirana je kao pogreška jer dodatno u prijevodu stavlja naglasak na ono što će u ponovljenoj radnji biti drugačije učinjeno u odnosu na prvi put. U Primjeru 18. također nije neophodno uključiti ponovo izostavljeni dio u sam prijevod, no njegovim zadržavanjem pobliže se opisuju istaknuti naslovi u rečenici i jednoznačno se čitatelju pruža informacija da se tu radi specifično o filmovima, a ne, primjerice, o naslovima kazališnih djela.

Primjer 17.

*Yüksek müsaadenizle, usta bir dâhinin elini **bu defa** bilinçli bir şekilde sıkma fırsatını değerlendirmek isterim.*

Velika mi je čast upoznati vas, gospodine. Uz vaše veliko dopuštenje, želio bih iskoristiti priliku da svjesno stisnem ruku glavnom geniju.

*Velika mi je čast upoznati se s vama, gospodine. S vašim cijenjenim dopuštenjem, htio bih iskoristiti priliku da se opet rukujem s vama, **ovaj put** svjestan činjenice da je to ruka velikog stručnjaka.*

Primjer 18.

*„Marslıların Bedduası ve Astronot Niyazi **filmlerinde** mükemmeldiniz.“*

„Bio si savršen u Marsovskom prokletstvu i astronautu Niyaziju.“

*„Bili ste savršeni **u filmovima** Kletva Marsovaca i Astronaut Nijazi.“*

U nedostatku prikladne kategorije i vrste, nekolicina rečenica izdvojena je kao nesvrstana zbog nejasnoće i nemogućnosti isticanja specifičnih pogrešaka te će one biti navedene u sljedeća dva primjera:

Primjer 19.

Misafirperverliğimizi gösterelim, hal hatır soralım, izzet ikramda bulunalım, akabinde kanıyla tavanı boyarız.

Pokažimo svoje gostoprimstvo, pitajmo kako, liječimo dostojanstvo, a zatim obojimo strop krvlju.

Bit ću gostoljubiv, pitati ga kako je, sjajno ga ugostiti, a zatim obojiti strop njegovom krvlju.

Primjer 20.

Sanmam, elinin hamurunu kanla ıslatmaktan zevk alacağı kanaatindeyim

Mislim da neće uživati u namakanju paste svoje ruke krvlju

Sumnjam, mislim da će htjeti zakrvaviti svoje nesposobne ruke...

Ako usporedimo lakoću analize i jednostavnost prijevoda, svakako prednost ima novinarsko-publicistički tekst, u kojemu je, s obzirom na duljine izvornih rečenica te brojne odnose i suodnose rečeničnih dijelova u strojnom prijevodu, postupak prevođenja relativno uspješno proveden. Aktualnost teme odabranog članka potencijalno je mogla otežati posao odabranom sustavu za strojno prevođenje, ako uzmemo u obzir činjenicu da je taj isti sustav vjerojatno treniran na starijim paralelnim tekstovima u kojima ta tema nije bila prisutna, no u ovom slučaju to se nije dogodilo.

Kod književnoumjetničkog teksta razvrstavanje pogrešaka u odabrane kategorije bilo je izazovno i zahtjevno, no isto tako i očekivano. S obzirom na slobodnu formu pisanja s jedne strane, kreiranje *vlastitog* neprepoznatljivog jezika i udaljavanje od jezične norme, veći broj pogrešaka i njihova zahtjevna analiza bili su neizbježni. Dodatno, određeni stil pisanja prevoditelja konvencionalnog prijevoda s jedne i vrednovatelja prijevoda s druge strane mogu dovesti do različita tumačenja i interpretacije istog teksta.

Jedno od ograničenja provedene analize svakako je nedostatak objektivnosti, jednako kao i nemogućnost primjene ususstavljene klasifikacije pogrešaka. Unaprijed odabrane vrste pogrešaka prilikom analize usmjerile su vrednovatelja na traženje već klasificiranih pogrešaka umjesto sagledavanja prijevoda u cijelosti te potencijalno kreiranja prikladnije klasifikacije i kategorizacije. Na ovom jezičnom paru nedvojbeno bi bila zanimljivija i informativnija detaljnija raščlamba unutar vrsta pogrešaka, posebice unutar morfosintaktičke vrste. Analizom se pokazala očekivana relacija između jezičnih stilova i broja pogrešaka pa bi bilo zanimljivo vidjeti uspješnost strojnog prevoditelja u tekstovima koji su jednostavnije forme i jezika. Možemo pretpostaviti da bi prijevod bio jasniji, a broj pogrešaka manji.

Strojnom prevoditelju još uvijek nedostaju dosljednost u prijevodu terminologije, uzimanje konteksta u obzir, zadržavanje vjernosti izvornom tekstu i neizostavno vođenje računa o namjeni samog prijevoda u vidu korištenja i publike.

6. Zaključak

Strojno prevođenje, koje je tema ovog rada, postalo je sve zanimljivije i na znanstvenom području i u ostalim društvenim aspektima. Potreba za prijevodima koji su jeftini, brzi i pružaju inicijalnu informaciju o tekstu i kontekstu na bilo kojem jeziku neprestano se povećava. Istovremeno s potrebama za kvalitetnim alatima za strojno prevođenje raste i broj istraživanja i stjecanje znanja u svrhu njihova konstantnog unapređenja. Jasno je da je, kako bi postojeći sustavi bili bolji, itekako potrebno poznavati svaki aspekt trenutačnog stanja i kontinuirano prikupljati nove informacije.

Ovom radu cilj je upravo prikupljanje informacija o stanju alata za strojno prevođenje Google Prevoditelja na primjeru jezičnog para turski i hrvatski. Analiza je pokazala napredak i kontinuirano usavršavanje tog alata s obzirom na princip rada te traženje potencijalnih rješenja za postojeće probleme, ali i dodatno usavršavanje sustava. Podaci prikupljeni ovim istraživanjem mogu ukazati na eventualna mjesta za napredak ili promjenu pristupa u korištenju samog alata. Potrebno je poznavanje prednosti i ograničenja analiziranog sustava, ali i načina rada koji je pokriven u teorijskom dijelu, dok je istraživački dio na odabranom nereprezentativnom uzorku ipak pružio određene zaključke.

Za konkretniji doprinos društvenim vrijednostima i konkretnije zaključke istraživanje bi trebalo provesti na većem opsegu teksta, s većom raznolikošću jezičnih stilova i većim brojem vrednovatelja samih sustava kako bi se postigla viša razina objektivnosti. Analiza ipak pokazuje permanentnu borbu s istim obilježjima prirodnog jezika, bez obzira na jezični par koji je predmet analize. Za predmetni jezični par ipak je specifična veća ili manja zastupljenost pogrešaka unutar jezičnog para koja ovisi o sličnostima i strukturnim razlikama između turskoga i hrvatskoga. Daljnja istraživanja svakako trebaju biti usmjerena na otkrivanje prednosti i nedostataka trenutačnog principa rada Google Prevoditelja u odnosu na prijašnji kako bi se mogla potvrditi ispravnost smjera u istraživanjima i implementaciji koji može dovesti do poboljšanja tih prijeko potrebnih resursa.

7. Literatura

Antunović, G. i Pavlović, N., (2019), Redaktura strojnih prijevoda – sve važniji prevoditeljski zadatak, Jezik i um, str. 147-167.

Arnold et al. (2002), Machine Translation: an Introductory Guide, London, Blackwells-NCC.

Baker, M., (1992), In other words: a coursebook on translation, Routledge, New York.

Bar-Hillel, Y., (1953), Some Linguistic Problems Connected with Machine Translation. Philosophy of Science, 20(3), str. 217-225. Preuzeto s http://www.jstor.org/stable/185499 [pristupljeno 26. 1. 2021.]

Bowker, L., (2002), Computer-aided translation technology: a practical introduction, University of Ottawa Press.

Brkić, M., Seljan, S. i Matetić, M., (2011), Machine Translation Evaluation for Croatian-English and English-Croatian Language Pairs. U: Sharp, B., Zock, M., Carl, M. & Jakobsen, A. (ur.) Proceedings of the 8th International NLPCS Workshop: Human-Machine Interaction in Translation. Copenhagen, Copenhagen Business School, str. 93-104.

Brkić, M., Seljan, S. i Vičić, T., (2013), Automatic and Human Evaluation on English-Croatian Legislative Test Set, Lecture Notes in Computer Science - LNCS, 7816 (1), 311-317.

Čaušević, E., (1996), Gramatika suvremenog turskog jezika, Hrvatska sveučilišna naklada, Zagreb.

Čaušević, E., (2018), Ustroj, sintaksa i semantika infinitnih glagolskih oblika u turskom jeziku. Turski i hrvatski jezik u usporedbi i kontrastiranju, Ibis grafika, Zagreb.

Čaušević, E., (2019), Poteškoće s kojima se mogu suočiti izvorni govornici turskoga tijekom učenja i usvajanja hrvatskoga jezika, Savjetovanje za lektore hrvatskoga kao inoga jezika 4., U: Banković-Mandić, I., Čilaš Mikulić, M. & Matovac, D. (ur.), FF-press, Zagreb, str. 73-85. https://www.academia.edu/40281752/Pote%C5%A1ko%C4%87e_s_kojima_se_mogu_suo%C4%8Diti_izvorni_govornici_turskoga_tijekom_u%C4%8Denja_i_usvajanja_hrvatskoga_jezika_Difficulties_that_Turkish_speakers_have_while_learning_Croatian_ [pristupljeno 27. 1. 2021.]

Dovedan, Z., Seljan, S., Vučković, K., (2002), Strojno prevođenje kao pomoć u procesu komunikacije, Informatologia, 35(4), str. 283-291.

Dunđer, I., (2015.) Sustav za statističko strojno prevođenje i računalna adaptacija domene, doktorska disertacija, Sveučilište u Zagrebu, Filozofski fakultet.

Finka, B. i László, B., (1962), Strojno prevođenje i naši neposredni zadaci, Jezik, 10 (4), 117-121. Preuzeto s https://hrcak.srce.hr/226114 [pristupljeno 26. 1. 2021.]

Gutić, T., (2009)., Književnoumjetnički stil, Hrvatistika, 3(3), str. 49-55. Preuzeto s https://hrcak.srce.hr/70012 [pristupljeno 4. 2. 2021.]

Hutchins, J., (1996), Computer-based Translation Systems and Tools, Elra Newsletter, 1.4, str. 6-9.

Hutchins, J., (2003), Has machine translation improved: some historical comparisons, MT Summit IX: proceedings of the Ninth Machine Translation Summit, New Orleans, USA, September 23-27, str. 181-188. Preuzeto s http://www.hutchinsweb.me.uk/MTS-2003.pdf [pristupljeno 26. 1. 2021.]

Hutchins, W. J., (1995), Machine Translation: a brief history, Concise history of the language sciences: from the Sumerians to the cognitivists, Edited by E. F. K. Koerner and R. E. Asher, Oxford: Pergamon Press, str 431-445. http://www.hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf [pristupljeno 21. 1. 2021.]

Hutchins, W. J., (2002.), The State of Machine Translation in Europe and Future Prospects. http://hutchinsweb.me.uk/HLT-2002.pdf [pristupljeno 20. 1. 2021.]

Hutchins, W., J., Somers, H.,L., (1992), An Introduction to Machine Translation, Academic Press, London.

Jakobson, R., (1959), On linguistic aspects of translation, On translation, edited by Reuben, A. B., Harvard University Press, Cambidge, Massachusetts, str. 232-239. https://www.academia.edu/26570349/Jakobson_Roman_1959_On_Linguistic_Aspects_of_Translation [pristupljeno 28. 1. 2021.]

Josić, Lj., (2010), Književnoumjetnički stil – funkcija standardnoga jezika, jezik sui generis ili »nadstil«?, Studia lexicographica, 4 (2(7)), str. 27-48. Preuzeto s https://hrcak.srce.hr/110388 [pristupljeno 4. 2. 2021.]

Kapović, M., (2010.), Čiji je jezik, Algoritam, Zagreb.

Katnić-Bakaršić, M., (2001), Stilistika, Naučna i univerzitetska knjiga, Sarajevo.

Koehn, P., (2010), Statistical Machine Translation. Cambridge University Press, New York.

Kojčinović, R., (2014). Leksička obilježja jezika u novinskim tekstovima i naslovima, Hrvatistika, 7. (7.), str. 39-50. Preuzeto s https://hrcak.srce.hr/134918 [pristupljeno 1. 2. 2021.]

Kučiš, V., (2010), Prevodilački alati u funkciji kvalitete prijevoda, Informatologia, 43(1), str. 19-33

Lamza Posavec, V. (2004), Metode društvenih istraživanja, Zagreb: Hrvatski studiji Sveučilišta u Zagrebu

Locke, W.N., Booth, A.D. eds., (1955) Machine translation of languages, Fourteen essays, John Wiley & Sons, New York (co-published with The Technology Press).

Lujić, B. (2007). Lingvističke teorije prevođenja i novi hrvatski prijevod Biblije. Bogoslovska smotra, 77 (1), 59-102. Preuzeto s https://hrcak.srce.hr/23491. [pristupljeno 13. 1. 2021.]

Ljubas, S., (2017), Analiza pogrešaka u strojnim prijevodima sa švedskog na hrvatski, Hieronymus – časopis za istraživanja prevođenja i terminologije, 4, str. 28-64.

Ljubas, S., (2018), Prijelaz sa statističkog na neuronski model: usporedba strojnih prijevoda sa švedskog na hrvatski jezik, Hieronymus – časopis za istraživanja prevođenja i terminologije, 4, str. 72-91.

Ljubas, S., (2020), Utjecaj višejezičnosti vrednovatelja na ljudsku procjenu kvalitete strojnih prijevoda, Jezikoslovlje, 21 (2), str. 207-235.

Matasović, R., (2008.), Poredbeno povijesna gramatika hrvatskog jezika, Matica hrvatska, Zagreb.

Mihaljević, M., (2002). Funkcionalni stilovi hrvatskoga (standardnog) jezika (S posebnim obzirom na znanstveno-popularni i personalni podstil). Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje, 28 (1), str. 325-343. Preuzeto s https://hrcak.srce.hr/68820 [pristupljeno 4. 2. 2021.]

Mihaljević, M., (2003), Kako se na hrvatskome kaže www?: kroatistički pogled na svijet računala, Hrvatska sveučilišna naklada, Zagreb.

Pavlović, N., (2017), Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka, u: Stolac, D. & Vlastelić, A. (ur.)Jezik kao predmet proučavanja i jezik kao predmet poučavanja.

Petrzelka, J., (2011), Statistical Machine Translation; how to maximize the quality of translation, LAP Lambert Academic Publishing, Saarbrucken.

Raguž, D., (1997), Praktična hrvatska gramatika, Medicinska naklada, Zagreb.

Seljan, S., Brkić, M., Kučiš, V., (2011), Evaluation of Free Online Machine Translation for Croatian-English and English-Croatian Language Pairs. 3rd International Conference "The Future of Information Sciences: INFuture2011 – Information Sciences and e-Society", 9-11 November 2011, Zagreb.

Silić, J., (2006), Funkcionalni stilovi hrvatskog jezika, Disput, Zagreb.

Simeon, I., (2002), Paralelni korpusi i višejezični rječnici, Filologija, 38-39, str. 209-205.

Simeon, I., (2008), Vrednovanje strojnoga prevođenja, Doktorska disertacija, Sveučilište u Zagrebu, Filozofski fakultet.

Šimić, J., Uglarik, D., Vuk, D., (2010)., Mogućnosti i ograničenja strojnog prevođenja, Praktični menadžment, 1 (1), str. 81-85. Preuzeto s https://hrcak.srce.hr/67847 [pristupljeno 23. 1. 2021.]

Tadić, M., (2003), Jezične tehnologije i hrvatski jezik, Ex Libris, Zagreb.

Težak, S., Babić, S., (1994), Gramatika hrvatskog jezika: priručnik za osnovno jezično obrazovanje, Školska knjiga, Zagreb.

8. Prilozi

9. Materijali za analizu