

ANALIZA POGREŠAKA U STROJNIM PRIJEVODIMA SA ŠVEDSKOG NA HRVATSKI

Sandra Ljubas

Sažetak

Cilj ovog rada jest analizirati pogreške koje nastaju u strojno prevedenim tekstovima sa švedskog na hrvatski jezik s pomoću Google prevoditelja (GP) te time dati preliminarni uvid u situaciju s alatima za strojno prevođenje za švedsko-hrvatski jezični par. Analizirani materijal sastoji se od tehničkog, političkog, administrativnog i znanstvenog korpusa (koji su ukupno dugački 80 kartica), od ljudskih prijevoda tih tekstova te naposljetku od strojnih prijevoda. Zabilježene pogreške raspodijeljene su prema Simeoninoj klasifikaciji pogrešaka (2008) u osam kategorija: neprevedene, izostavljene i dodane riječi, leksičke, stilističke, pravopisne i morfosintaktičke pogreške, pogreške u redu riječi te u dvije dodatne kategorije: pogreške u dosljednosti i nejasne rečenice. Rezultati pokazuju da se u prijevodima GP-a na hrvatski jezik najčešće pojavljuju morfosintaktičke pogreške, ali zapravo su leksičke one koje najviše utječu na (ne)razumljivost teksta. Također je utvrđeno da GP osobito često griješi u prijevodu švedskog formalnog subjekta „det“, dvočlanih švedskih glagola, povratnih zamjenica, dvostrukih negacija, kratica, složenica itd.

1. Uvod

Neovisno o tome radi li se o profesionalnim prevoditeljima ili o ljudima koji u svom svakodnevnom životu nailaze na jezične barijere, većina se ljudi danas već susrela s nekom vrstom strojnoga prevođenja ili su se i sami već koristili nekim alatima za strojno prevođenje. Najčešće takvi prijevodi ne odgovaraju očekivanjima korisnika pa je samo strojno prevođenje došlo na poprilično loš glas. Manjkavi prijevodni izbori javno dostupnih sustava za strojno prevođenje često služe kao izvor zabave i smijeha, a uvriježeno je mišljenje da oni ne mogu poslužiti kao pomoć pri prevođenju. Međutim, kako korisnik u većini javno dostupnih

strojnoprevoditeljskih sustava ne može, primjerice, naznačiti o kojoj se vrsti teksta radi ili pak na raspolaganje staviti određenu terminologiju, tako ti besplatni sustavi tek daju naslutiti što se sve može postići sa sustavima za profesionalnu uporabu (Muegge 2009: 11). Računala još ne mogu zamijeniti čovjeka, ali valja prilagoditi svoja očekivanja i razumjeti što ti sustavi mogu, a što ne mogu učiniti, kako mogu utjecati na rad profesionalnih prevoditelja te koje mogućnosti nude „običnim“ ljudima u svakodnevnom životu prije nego što se u potpunosti odbaci njihova korisnost. Korisnici se, ovisno o situaciji, često mogu zadovoljiti i prijevodima niže jezične kvalitete, a koji su im ipak uštedjeli novac i vrijeme, a postoje i druga područja uporabe strojnih prijevoda, o kojima će riječ biti u potpoglavlju 3.1.

2. Cilj

Brojni su se znanstvenici već bavili istraživanjima strojnoga prevođenja s velikih svjetskih jezika i na njih (npr. Andersson i Johansson 2010, Ranta 2010, Lindqvist 2012, Tidqvist 2015), najčešće uključujući engleski i neki drugi jezik, no manji je broj onih istraživanja koja uključuju hrvatski jezik (npr. Simeon 2008, Brkić, Kučić i Seljan 2011, Pavlović 2017). Ovo je istraživanje pilot projekt kada govorimo o strojnom prevođenju sa švedskog na hrvatski jezik te sam njime željela pružiti uvid u situaciju sa strojnim prevođenjem u toj jezičnoj kombinaciji te procijeniti do koje su mjere takvi prijevodi razumljivi. Godine 2003. jezični alati u vidu jedno- i višejezičnih općih rječnika, korpusa, enciklopedija, sustava za strojno prevođenje i sl. za hrvatski jezik još nisu postojali „u svima prihvatljivom, javno pristupačnom obliku kao brza i besplatna usluga“ (Tadić 2003: 47, 48). U ovom ću istraživanju provjeriti je li Hrvatska u međuvremenu napredovala s digitalizacijom u smislu javno dostupnih sustava za strojno prevođenje te vidjeti kako danas funkcionira strojno prevođenje tekstova na hrvatski jezik. Cilj je provjeriti koliko se pogrešaka javlja u prijevodima Google prevoditelja sa švedskog na hrvatski jezik te proučiti prirodu tih pogrešaka. Istraživanje je provedeno početkom studenog 2016. godine na tada aktualnoj verziji GP-a. Odabrala sam GP u vjeri da korisnici često posežu upravo za tim sustavom za strojno prevođenje zbog njegove dostupnosti, prisutnosti i lakoće uporabe, ali i zato što Google predano radi na poboljšanju svoga alata za prevođenje, što je krajnji cilj svih istraživanja iz ovog područja. I

nakon ovog su istraživanja ponovno ažurirani algoritmi GP-a te više ne govorimo o statističkom, već o neuronskom sustavu (v. poglavlje 3.1.). Još uvijek je otvoreno pitanje koliko ta promjena utječe na kvalitetu izlaznih podataka, osobito kada govorimo o jezičnim kombinacijama koje ne uključuju engleski jezik.

Pogreške koje zabilježim u svom istraživanju rasporedit ću u određene kategorije te pokušati objasniti zašto neke od njih nastaju, kako bi se eventualno mogle spriječiti u budućim prijevodima. Vjerujem da razumijevanje pogrešaka pomaže s razumijevanjem izlaznih podataka (strojnog *outputa*) te da nova saznanja koja se dobivaju analiziranjem pogrešaka pomažu s pronalaskom „jednostavnih i praktičnih rješenja za one koji revidiraju i uređuju tekstove nakon što su prošli kroz sustav“¹ (Ahrenberg i Merkel 2006b), a time i s daljnjom nadogradnjom sustava za strojno prevođenje. Isto se tako nadam da će ovo istraživanje pomoći u sagledavanju ukupne slike i trenutne situacije s hrvatskim jezikom i strojnoprevoditeljskim sustavima.

3. Ključni pojmovi i prethodna istraživanja

U ovom ću poglavlju definirati temeljne pojmove te dati pregled najvažnijih podataka iz prijašnjih istraživanja koja se tiču strojnoga prevođenja.

3.1 Svrhe i postupci strojnoga prevođenja

Strojno je prevođenje izvorno bilo zamišljeno kao FAHQMT (*fully automatic high-quality machine translation*), odnosno željelo se stvoriti sustav za strojno prevođenje koji bi u potpunosti bez ljudske pomoći kvalitetno prevodio (Bennett i sur. 1986, prema Tomicic 2001: 4). Danas se kao rezultat strojnoga prevođenja prihvaćaju i „grubi“ prijevodi koji se potom dorađuju, a nadaleko je rašireno i prevođenje koje je tek strojno potpomognuto.

Postoje različite svrhe za koje prijevodi mogu poslužiti, a o tim svrhama ovisi i potrebna razina jezične kvalitete. Neki se strojni prijevodi objavljuju bez redakture pa je jezična kvaliteta iznimno bitna (Sågvall Hein 2008). Drugi se prepravljaju i dorađuju nakon što su prošli kroz sustav, a prije nego što će ih se objaviti (Ahrenberg i Merkel 1997: 96). No ne objavljuju se svi prijevodi. Katkad je

¹ Moj prijevod. Samostalno sam sa švedskog i engleskog prevela i sve ostale navode u ovom radu.

čitatelju ili naručitelju dovoljan „grub“ prijevod kako bi otprilike razumio o čemu se u nekom tekstu radi (1997: 97). Takav tekst ne mora odgovarati visokim jezičnim standardima da bi ispunio potrebe naručitelja pa je ljudski prijevod u tom slučaju zapravo prespor, preskup, pa čak i previsoke kvalitete. „Grubi“ prijevodi često su dostatni korisnicima na internetu, ali i, primjerice, u Europskoj komisiji (EAMT 2008) čiji su zahtjevi preveliki da bi im ljudski prevoditelji mogli dovoljno brzo udovoljiti. U takvim je institucijama upravo nedostatak vremena najveći problem pa sustavi za strojno prevođenje zaista mogu olakšati posao ljudskim prevoditeljima (Lindqvist 2012: 1). Usto se sustavi za strojno prevođenje rabe kao podrška u komunikaciji pri prijevodima *e-mailova* ili razgovora na *chatu* (Ahrenberg i Merkel 1997: 97). U takvim slučajevima ljudsko prevođenje nije ni moguće jer su prijevodi potrebni u realnom vremenu, dok je sama kvaliteta manje važna. Naposljetku postoji i mogućnost višejezične potrage za informacijama na internetu odnosno mogućnost da tražilica ponudi sve relevantne rezultate koji se tiču neke upisane riječi, neovisno o jeziku upita (Sågvall Hein 2008). U današnje vrijeme „valja razmišljati i o alatima koji bi korisnicima omogućili pretraživanje takva repozitorija znanja pohranjenog u tekstovima na raznim jezicima“, za što su „potrebni elementi gruboga strojnog prevođenja“ (Tadić 2003: 72).

Postoje različiti postupci strojnoga prevođenja. Izravno je prevođenje najstariji, najjednostavniji i „najprimitivniji postupak“ (Hedden 2010). Sustav se koristi leksikonom i prevodi jednu po jednu riječ ili frazu, a zatim uz pomoć određenih pravila rješava prijevodne probleme (poput reda riječi, deklinacija, prijedloga, glagolskih vremena itd.). Rezultati ovise o kvaliteti rabljenog leksikona (Sågvall Hein 2005: 25). Nasuprot tomu, transfer (prijenos) u cijelosti se temelji na analizi rečenice, što je i najveća razlika između njega i izravnoga prevođenja. Transfer je mnogo učinkovitiji od izravnog prevođenja kada se radi o identifikaciji gramatičkih službi unutar rečenice (ibid.). Ta metoda najviše nalikuje ljudskom procesu prevođenja (Hedden 2010), no u ovom je trenutku najučestalija metoda nazvana statističko 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 kraj druge također nalaziti jedni kraj drugih“ (Ahrenberg i Merkel 2006a). Ta predviđanja mogu biti vrlo točna, uz nužan preduvjet da arhiv obuhvaća velik broj postojećih

prijevoda. Što je više reprezentativnog materijala dostupno, to će i rezultati biti bolji. Još je jedna prednost statističkih sustava to što ih korisnici sami mogu „obučavati“ (Tidqvist 2015: 3). To znači da se sustavu može predložiti prijevod koji bolje odgovara danom kontekstu, što on pamti za sljedeće prijevode i tako polako popravlja svoje rezultate. Najveći broj javno dostupnih sustava za strojno prevođenje upravo su statistički, no postoji i nova paradigma u svijetu strojnog prevođenja, a krije se pod kraticom NMT (*neural machine translation*). Neuronsko strojno prevođenje novi je pristup koji se temelji na neuronskim mrežama i na principima dubokog učenja. Sustav s pomoću neuronske mreže pokušava razumjeti kontekst ulazne rečenice X i izračunati vjerojatnost da prijevod Y odgovara, a da mu pritom nisu potrebne vanjske lingvističke informacije (Castilho i sur. 2017: 110). Dosadašnja su istraživanja pokazala kako neuronski sustavi pružaju bolje rezultate kada su u pitanju fleksija i kongruencija, no da statistički sustavi pružaju bolje rezultate u slučaju dužih segmenata (2017: 111). Najveći je problem neuronskih sustava prevođenje rijetkih riječi koje nisu sadržane u njihovim rječnicima, a dolazi i do češćih izbacivanja riječi nego li je to bio slučaj kod statističkih sustava. Te bi probleme moglo riješiti ugrađivanje statističkih obilježja poput prijevodnih modela i n-gram jezičnih modela u neuronsku mrežu (He i sur. 2016: 151).

3.2 Google prevoditelj

Google tražilica na internetu je predstavljena 1998. godine s ciljem da se unutar nje „organiziraju sve informacije svijeta i učini ih se dostupnima svima na slobodno korištenje“ (Andersson i Johansson 2010: 5). Kako bi taj cilj i postigao, Google je razvio brojne besplatne usluge, a Prevoditelj je jedna od onih koja je postala veoma poznata, kojom se koriste ljudi diljem svijeta te koju tvrtka neprestano razvija i poboljšava. Google se općenito trudi da uvijek ostane konkurentan na tržištu te aktualiziran (ibid.) pa zato i kvaliteta GP-a s godinama raste. Početkom studenog 2016. godine, kada je ovo istraživanje provedeno, Google se još služio statističkim modelom, no nedugo zatim predstavljen je neuronski model (Wu i sur. 2016).

U načelu ne postoje velike razlike između načina rada Google prevoditelja i Google tražilice: koriste se sličnom tehnikom, statistikom o jeziku (Ranta

2010: 50). Rezultati doduše jesu proizvodi „nagađanja“ i „pretpostavljanja“, no radi se o veoma dobrim procjenama koje se temelje na materijalu od više milijuna riječi (ibid). Točan način funkcioniranja GP-ovog statističkog modela poslovna je tajna, no smatra se da se služi nekim modificiranim statističkim algoritmom koji provodi izračune o tome kolika je vjerojatnost da neku određenu riječ slijedi neka druga, ali pritom pazi i na najnižu granicu (ne)vjerojatnosti točnosti nekog prijevodnog rješenja (Andersson i Johansson 2010: 15). Najveća je Googleova prednost to što je internet sam po sebi veoma obuhvatan korpus – i to najveći na svijetu (Koletnik Korošec 2011: 1).

GP danas pokriva 103 jezika. Kako je moguće kreirati toliko programa da se uspijeva prevoditi između svih tih jezika? Matematika nalaže da bi za to bilo potrebno 10 506 različitih programa², no ako se u većini slučajeva prevodi preko engleskoga kao međujezika, broj se potrebnih programa drastično smanjuje³. Kako ne postoji dovoljna količina paralelnih tekstova između svih svjetskih jezika, nego ih je uvijek najviše između engleskog i drugih jezika, međujezik je najbolja tehnika ako se želi postići da sustav prevodi između velikog broja jezika (Ranta 2010: 55). Razumljivo je zašto se Google odlučio na to rješenje – GP nije usmjeren na neko određeno područje, već njegova najveća prednost leži upravo u tome da pokriva više tisuća pojmova, čak i ako nije uvijek najprecizniji (2010: 58).

3.3 Izlazni podaci strojnih prijevoda i njihovo vrednovanje

Kako stroj olakšava prevođenje čovjeku, tako i čovjek katkad može olakšati stroju da izlazni podaci budu bolji. Prethodno uređivanje (*pre-editing*) izvornika odnosi se na jezično dorađivanje teksta prije prevođenja, što omogućuje stroju da bolje razumije tekst, dok *post-editing* označava naknadnu redakтуру prijevoda (Hedden 2010). Prevoditelj katkad može razumjeti što je autor teksta mislio čak i kada tekst nije najbolje sročeno pa u skladu s tim može napraviti prijevod koji će biti bolji od izvornika. Značenje neke riječi ili fraze osim toga može ovisiti i o raznim drugim čimbenicima, pa tako čak i izvantekstni čimbenici mogu razjasniti

² Formula $n(n - 1)$, $n = 103$

³ Ako se uvijek prevodi preko engleskog, ostaje ih samo 204; formula $2(n - 1)$, $n = 103$.

određene dvoznačnosti (Ranta 2010: 54). Kako strojni prevoditelj nema takve kognitivne sposobnosti, potrebno mu je pomoći s prijevodom tako da se uklone sve gramatičke i pravopisne pogreške te sve dvoznačnosti u tekstu (Sågvall Hein 2008). Još više na kvalitetu rezultata utječu pogreške u interpunkciji, nepotrebna velika slova i razbijanja rečenica u novi red pa valja ukloniti i njih (Andersson i Johansson 2010: 15). *Post-editing* pak dokazuje kako pitanje kvalitete prijevoda nije pitanje isključivo količine pogrešaka, već da je važna i priroda tih pogrešaka te koliko je vremena potrebno kako bi ih se ispravilo. Strojno je prevođenje, naravno, brže od ljudskog (već se uz pristup internetu 24 sata na dan mogu dobiti besplatni prijevodi), no ako se prijevod objavljuje, svakako pri vrednovanju kvalitete treba u obzir uzeti i vrijeme potrebno za redakciju (Hedden 2010).

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 Tomicic 2001: 8, 9). Velik je problem i pragmatičnost – program ne može prilagoditi prijevod čitatelju niti razmišljati o ciljnoj skupini dok prevodi, a osim toga je aspekt pragmatičnosti i vrlo teško istraživati (Tidqvist 2015: 12).

Izlazni podaci uvijek ovise o vrsti tekstova koji se prevode, no činjenica je da se u strojnim prijevodima javlja mnogo pogrešaka unatoč tome što ih se pokušalo popraviti i što se mnogo istraživalo u tom području (Lindqvist 2012: 1). Strojni prijevodi često sadrže sljedeće pogreške: neprevedene i izostavljene riječi, pogrešne leksičke odabire, pogrešnu interpunkciju, deklinaciju ili konjugaciju (Sågvall Hein 2008). Neki od kriterija kvalitete i vrednovanja strojnih prijevoda jesu: razumljivost (koja se procjenjuje subjektivno), vjernost izvorniku (koja se vidi u usporedbi izvornika s prijevodom), korisnost i prihvatljivost (koja ovisi o individualnim zahtjevima korisnika) te lingvistička obilježja (broj (ne)točnih ekvivalenta na razini riječi ili rečenice) (Butler 1992, prema Tomicic 2001: 12).

Razumljivost je veoma subjektivna kategorija te se „rezultat može kretati od prevedene rečenice koja izgleda kao sasvim ispravna rečenica ciljnog jezika do potpune besmislice čije značenje korisnik ni uz najveći napor i ulog mašte ne može rekonstruirati“ (Simeon 2008: 27, 28). Vjernost izvorniku odnosi se na zadržavanje izvornoga značenja rečenice. Ovdje je potrebno u obzir uzeti veće količine teksta – na razini riječi ili neke druge jedinice možemo pomisliti da je prijevod loš (npr. ako uočimo neku gramatičku pogrešku), no u cjelini prijevod i dalje može biti prihvatljiv (2008: 29). Stoga je priroda pogreške veoma važna dimenzija vrednovanja strojnih prijevoda.

Ovo se istraživanje bavi upravo prirodom pogrešnih rješenja koje je odabrani sustav za strojno prevođenje, Google prevoditelj, ponudio. Pri analizi pogrešaka važne su i količina i priroda pogrešaka. Valja uzeti u obzir da procjena prirode pogrešaka ne može u cijelosti biti objektivna, već da uvijek ovisi o osjećaju za jezik osobe koja vrednuje prijevod. Upitno je koliko je osoba stroga pri procjenjivanju pogrešaka, koliko dosljedno procjenjuje pogreške kada pregledava veliku količinu tekstova i sl. (2008: 30, 31). Ukoliko je uspješno odrađen kvalitativni dio određivanja vrste pogrešaka, utoliko će se relevantnije informacije moći izvući iz kvantitativnog dijela koji svjedoči o količini pogrešaka. Zato sam pokušala umanjiti moguća odstupanja čvrsto definirajući kategorije u koje ću smjestiti zabilježene pogreške. Također sam prilagodila postojeću klasifikaciju pogrešaka kako bi bolje odgovarala odabranom jezičnom paru. U nastavku prvo opisujem svoje hipoteze (poglavlje 4.), a zatim i proces koji je doveo do određivanja kategorija u dorađenoj klasifikaciji pogrešaka (poglavlje 5).

4. Hipoteze

Kako je ovo prva poznata analiza pogrešaka provedena za švedsko-hrvatski jezični par, teško je bilo predvidjeti kakvi će biti rezultati pa stoga i formulirati konkretne hipoteze. Osmislila sam pet hipoteza koje se temelje na mojim općenitim znanjima o GP-u te na istraživanjima koja su učinjena s drugim jezičnim parovima (Simeon 2008, Lindqvist 2012, Tidqvist 2015 i druga). Glavni je cilj ovog istraživanja bio dobiti uvid u situaciju sa strojnim prevođenjem sa švedskog na hrvatski, a pritom sam provjerila sljedeće hipoteze:

1. GP zadržava strukturu izvornih rečenica (ne spaja ih i ne razdvaja), a time zadržava i isti broj rečenica kao u izvorniku. Iako je prilično uobičajeno da čovjek spaja ili razdvaja neke rečenice, pretpostavila sam da bi GP mogao sadržavati algoritam zbog kojeg bi točku i veliko slovo koje slijedi shvaćao kao kraj i početak nove rečenice te da bi taj obrazac ponavljao u prijevodima.
2. GP ne griješi u ortografiji i nema zatipke u prijevodima sa švedskog na hrvatski jezik. Statistički se sustavi temelje na korpusima koji obuhvaćaju već objavljene, uređene prijevode koji bi trebali biti očišćeni od pogrešaka ovakvog tipa, zato ih nisam očekivala ni u strojnim prijevodima GP-a.
3. U prijevodima GP-a sa švedskog na hrvatski pojavljuju se engleske riječi. Međujezik povećava broj koraka u prevođenju i time izgleda za nove i nepredviđene pogreške. Smatrala sam mogućim da bi se u prijevodu našle riječi koje su „zapele“ na engleskom jeziku u tom međukoraku.
4. GP griješi u poretku riječi i sintaktičkoj strukturi rečenica u prijevodima sa švedskog na hrvatski jezik. Iako hrvatski jezik ima mnogo slobodniju sintaksu od švedske, bilo bi pogrešno ako bi GP uvijek preslikavao švedsku sintaksu u hrvatsku rečenicu (primjerice, da se finitni glagol uvijek nađe na drugom mjestu, kao što je to slučaj u švedskom).
5. GP griješi pri uporabi zareza u prijevodima sa švedskog na hrvatski jezik. U hrvatskom jeziku postoji više konkretnih pravila o uporabi zareza nego što ona postoje u švedskom koji se iznimno rijetko koristi zarezima pa valja provjeriti hoće li se GP uspjeti snaći s pravilnom uporabom zareza u hrvatskom jeziku.

5. Materijal i metoda

Prvi je korak ovog istraživanja bio načiniti korpus tekstova koje je čovjek već preveo sa švedskog na hrvatski jezik kako bih na raspolaganju imala i referentne prijevode. S obzirom na relativni nedostatak javno dostupnih paralelnih korpusa u toj jezičnoj kombinaciji, odabrala sam tekstove iz diplomskih radova studenata Katedre za skandinavistiku pri Sveučilištu u Zagrebu. Studenti prevoditeljskog smjera u okviru svojih diplomskih radnji prevode različite vrste tekstova kako bi pokazali široka prevoditeljska znanja koja su usvojili tijekom studija, no za svoje sam istraživanje odabrala samo određenu vrstu tekstova. Izbjegavala sam

književno-umjetničke tekstove te tekstove pisane svakodnevnim jezikom jer su oni najteži sustavima za strojno prevođenje, a birala sam one tekstove koji su pisani više ili manje kontroliranim jezikom. Pronašla sam i odabrala jedan tehnički tekst (dugačak 7 kartica⁴), jedan politički tekst (18 kartica), četiri administrativna teksta (24 kartice) te tri znanstvena i jedan popularno-znanstveni tekst (31 kartica). Tih deset švedskih izvornika čine prvi korpus, a ljudski prijevodi tih tekstova drugi korpus koji je paralelan s prvim i koji mi je poslužio za usporedbu strojnih prijevoda s ljudskima. Svi izvornici i prijevodi dostupni su u pripadajućim diplomskim radovima na repozitoriju Filozofskog fakulteta u Zagrebu⁵. Treći se korpus sastoji od strojnih prijevoda GP-a generiranih od 4. do 8. studenog 2016., kad je snazi još bila statistička verzija alata.

Ovi su korpusi izvor svim informacijama prikupljenima unutar istraživanja. Nisam se bavila proučavanjem procesa strojnog prevođenja, već sam se usredotočila na konkretne prijevode. U sljedećem potpoglavlju (5.1) opisujem kako sam prilagodila postojeću klasifikaciju pogrešaka te kako sam definirala određene kategorije, a svoje sam postupke potkrijepila primjerima iz analiziranoga korpusa.

5.1 *Klasifikacija pogrešaka*

U analizi pogrešaka u tekstovima koji su bili prevedeni s ruskog na engleski i s engleskog na hrvatski Simeon (2008: 108, 109) klasificira pogreške u osam kategorija:

1. neprevedene riječi (riječi koje su zadržane na izvornom jeziku)
2. izostavljene riječi
3. umetnute riječi (riječi iz ciljnog jezika koje su neopravdano umetnute)
4. leksičke pogreške (riječi koje se semantički znatno razlikuju od riječi u izvornome tekstu)
5. pravopisne pogreške

⁴ Pri čemu je jedna kartica jednaka 1800 znakova s razmacima.

⁵ <http://darhiv.ffzg.unizg.hr/> (18. 4. 2017.)

6. morfosintaktičke pogreške (pogrešni oblici, nesročnost između subjekta i predikata, pogreške u članovima, pogrešno odabrane funkcionalne riječi itd.)
7. stilističke pogreške (nespretno sročene sintagme, prijevodni ekvivalenti koji su semantički bliski polaznoj riječi, ali mogu dovesti do nesporazuma ili zbuniti čitatelja)
8. pogreške u redu riječi.

Preliminarna je analiza pokazala da će ova klasifikacija pogrešaka, uz manje preinake i prethodno striktno definiranje svih kategorija (v. potpoglavlja 5.1.1 – 5.1.8) odgovarati potrebama mog istraživanja. Potkrijepila sam sve primjerima iz korpusa, pri čemu su **masnim slovima** označene pojedine jedinice koje ću komentirati, ne osvrćući se pritom na ostale pogreške u tim rečenicama. Na početku primjera u zagradi uvijek navodim relevantni segment izvornika, a po potrebi i ljudski prijevod ili značenje izvornika.

5.1.1 Neprevedene riječi

GP katkad ne prevede neke riječi pa se u prijevodima mogu naći švedske riječi, kao npr. „gångriktningspil” što znači „strelica za pokazivanje smjera kretanja”:

Primjer 1. (Installera reliefsymboler och **gångriktningspil**, enligt beskrivning på sid 4.)
Ugradite reljefne simbole i **gångriktningspil**, kao što je opisano na stranici 4.

Međutim, zbog problema međujezika u korpusu se često pojavljuju i riječi na engleskom pa sam i njih ubrajala unutar kategorije neprevedenih riječi:

Primjer 2. (**Kemikaliestrategi** för Giftfri miljö.) (značenje = *strategija o kemikalijama*)
Chemical strategije netoksični okoliša.

Ubrajala sam, dakle, sve riječi koje nisu prevedene na hrvatski, osim ako su za to postojali opravdani razlozi za kojima bi se povel i ljudski prevoditelji.

5.1.2 Izostavljene riječi

GP ponekad „zaboravi” neke riječi u prijevodima. Ne postoji razlog zašto bi GP izbacio riječ „betydligt” („značajno”) u sljedećoj rečenici:

Primjer 3. (Vissa genvarianter finner man **betydligt** oftare tillsammans än man kunde förvänta.)

Određene genske varijante se nalaze (**ø**) češće zajedno nego što se moglo očekivati.

Da se ovdje radi o ljudskom prijevodu smatrala bih da je prevoditelj slučajno pogriješio. No svakako prevoditelj nekad može i svjesno odabrati da izostavi neku riječ u prijevodu (ako će značenje biti dovoljno jasno izraženo i bez nje), a jednako se tako značenje rečenice može zadržati ako stroj učini isto:

Primjer 4. (Sambandet mellan **det kapitalistiska systemet** och rasismen syns tydligt i den globala arbetsdelningen och den imperialistiska exploateringen.)

(Ljudski prijevod: Odnos između **kapitalističkog sustava** i rasizma očit je u globalnoj podjeli rada i imperijalističkom iskorištavanju.)

Odnos između **kapitalizma** i rasizma je vidljivo u globalnoj podjeli rada i imperijalističke eksploatacije.

Vjerojatno bi i mnogi drugi prevoditelji odabrali prevesti „det kapitalistiska systemet“ kao „kapitalistički sustav“ što je najbližije izvorniku, no ne postoji samo jedan ispravan prijevod pa je tako i samo „kapitalizam“ legitiman prevoditeljski izbor. Iako je tu jedna riječ izostavljena, ovo se nije smatralo pogreškom.

5.1.3 Umetnute riječi

Katkad GP čini upravo suprotno i umeće riječi koje u izvorniku nisu postojale niti su bile potrebne, a rezultat su pogrešno sravnjenih fraza u korpusu:

Primjer 5. (När **ø** respekten för det nationella oberoendet undergrävs och det generella förbudet mot anfallskrig åsidosätts, då öppnas nya vägar för imperialistisk expansion.)

Kada je **nadžaćana** poštivanje nacionalnu neovisnost potkopalo i opće zabrane agresivnog rata, a zatim je otvorio nove putove za imperijalističke ekspanzije.

No, kao i u prošloj kategoriji, u nekim je situacijama opravdano ili čak poželjno umetnuti neke riječi. Različiti su razlozi za to. Katkad je zbog sintakse potrebno dodati neke riječi ili prevoditelj umeće dodatne informacije kako bi nešto bolje pojasnio ciljnoj publici. Tako i strojni prijevod može biti uspješan sve ako sustav i

umetne koju riječ te jedinica i dalje može biti adekvatno prevedena. To nisam ubrajala kao pogrešku:

Primjer 6. (Efter **o** beslut) (značenje izvornika = *Nakon odluke*)
nakon **donošnja** odluke

5.1.4 Leksičke i stilističke pogreške

Leksičke se pogreške „znatno razlikuju od riječi u izvornom tekstu“, a stilističke su pogreške „semantički bliske polaznoj riječi, ali mogu dovesti do nesporazuma ili zbuniti čitatelja“ (Simeon 2008: 108, 109). Prijevodi su katkad nezgrapni i ekvivalenti ne odgovaraju uvijek u potpunosti danom kontekstu, no upravo o tom stupnju neodgovaranja ovisi radi li se o leksičkim ili stilističkim pogreškama. U sljedećem su primjeru odabrani ekvivalenti relativno prihvatljivi, no usporedba s ljudskim prijevodom pokazala je kako je čovjek odabrao stil koji je mnogo primjereniji administrativnome tekstu, a stroj u tome nije uspio. Ovakve sam primjere promatrala kao stilističke pogreške:

Primjer 7. (Det här ska du skicka med din ansökan)
(Ljudski prijevod: Dokumenti koje je potrebno priložiti uz prijavu)
Ovo bi trebalo poslati uz svoju prijavu

Leksičke pak pogreške mnogo više utječu na nerazumljivost rečenica, koje zbog pogrešnog leksičkog odabira mogu postati u potpunosti neprihvatljive, kao u sljedećem primjeru:

Primjer 8. (Människans **arvs massa**, det så kallade **genomet**, innehåller 20000 till 25000 gener.)
Ljudski **genom**, tzv **genom** sadrži 20.000 do 25.000 gena.

5.1.5 Pravopisne pogreške

Pravopisne pogreške uključuju pogrešno napisane riječi, pisanje velikog i malog slova te pogrešnu interpunkciju. Kada je riječ o interpunkciji, zarezi su se pokazali kao najveći problem, unatoč tome što sam pri procjenjivanju pogrešaka unutar ove kategorije bila prilično „blaga“. Računala sam samo one rečenice gdje zarez po svim pravilima hrvatskog jezika nikako ne smije biti izostavljen i one gdje ga nikako ne smije biti, kao u sljedećem primjeru:

Primjer 9. (Vissa undantag finns främst vad gäller pensionsförmåner.)

Neke iznimke, uglavnom se odnose na mirovinu.

Osim problema sa zarezima, učestalije su bile situacije kada su se bezrazložno u tekstu pojavljivali znakovi interpunkcije (npr. navodni znakovi) te su se javljali problemi s velikim i malim slovom, i to ne u nekim kompliciranijim imenima npr. institucija, već čak i na početku rečenice:

Primjer 10. (Invandrarnas delaktighet i det fackliga arbetet är ett viktigt inslag i demokratin på arbetsplatsen.)

sudjelovanje imigranata u sindikalnom radu je bitan element demokracije na radnom mjestu.

5.1.6 Morfosintaktičke pogreške

Ako promislimo o razlikama između švedskog i hrvatskog, postoji mnogo prostora za morfosintaktičke pogreške: hrvatski, primjerice, ima sedam padeža i kompliciraniju konjugaciju od švedskog, a u švedskom se javljaju dvočlani glagoli i mnogo češće pasivne konstrukcije. Stoga ova kategorija obuhvaća najveći broj različitih pogrešaka. U korpusu su najčešće zabilježene pogrešne deklinacije i konjugacije te problemi sa sročnošću, no smatra se da takve pogreške mnogo manje utječu na razumljivost teksta od, primjerice, leksičkih pogrešaka. Zato je bilo dostatno računati morfološke i sintaktičke pogreške unutar iste kategorije. Često se i unutar iste jedinice moglo zabilježiti više pogrešaka:

Primjer 11. (Projektet Odin **har präglats** av nytänkande och samarbete mellan forskare och företag inom olika områden och från flera länder.)

(Ljudski prijevod: Projekt Odin **obilježavaju** nove ideje i suradnja znanstvenika i poduzeća iz različitih područja i zemalja.)

Projekt Odin **je karakterizirala** inovacije i suradnje između znanstvenika i tvrtki u raznim područjima i iz nekoliko zemalja. (= pogrešan rod imenice te glagolski oblik)

5.1.7 Pogreške u redu riječi

Kao i u slučaju zarezima, nisam bila stroga pri procjeni što je pogreška u redu riječi u hrvatskom jeziku. Ipak, katkad je upravo poredak riječi pridonio tome da

rečenica postane nejasna. Pogreške u redu riječi općenito su prilično uobičajene u strojnim prijevodima, pa čak i u relativno dobro prevedenim rečenicama:

Primjer 12. (Om vistelsen i Sverige ska överstiga tre månader ska tillståndet vara klart före inresan.)

Ako boravak u Švedskoj će premašiti tri mjeseca, odobrenje će biti završena prije ulaska.

5.1.8 Ostale kategorije

Preliminarna je analiza također pokazala kako su u ovoj klasifikaciji potrebne dvije dodatne kategorije kako bi ona u potpunosti odgovarala potrebama mog istraživanja. Prvu sam kategoriju nazvala „dosljednost“ u prijevodu i ona se odnosi na prevođenje termina na dosljedan način tijekom cijelog teksta. Čak i ako je prijevod nekog termina pogrešan, važno je da je on preveden na isti način kroz cijeli tekst jer tako čitatelj mora samo provjeriti točan prijevod tog jednog termina i eventualno ga predložiti sustavu i tako poboljšati cijeli prijevod. Također, čak i ako za neki termin postoji više ispravnih ekvivalenata, bolje ih je prevoditi dosljedno. Iz ovog su isključene svakodnevne riječi koje se često iz stilističkih razloga prevode na različite načine. Primjer je nedosljednog prijevoda:

Primjer 13. (Migrationsverket)

Zavod za migracije (= naslov)

(Ska du besöka Migrationsverket i Norrköping eller Stockholm behöver du först boka tid.)

Ukoliko posjetite migracije u Norrköping ili Stockholmu morate prvo dogovoriti termin.

(Sen ska du besöka Migrationsverket för att få ett nytt uppehållstillståndskort.)

Potom biste trebali posjetiti imigracijske službe za dobivanje nove boravišne karticu.

Termin „Migrationsverket“ ponavljao se više puta u tekstu i najčešće je preveden ispravno kao „Zavod za migracije“, a druga su se dva prijevoda pojavila tu i tamo.

Računala sam svako njihovo pojavljivanje kao leksičku pogrešku, ali samo jednom unutar kategorije dosljednosti.

Unatoč mom trudu da kategoriziram svaku pogrešku na koju naiđem, zabilježila sam velik broj u cijelosti nekoherentnih i zbunjujućih rečenica (ili ponekad samo dijelova rečenica) koje su vrvjele od pogrešaka koje nisam uspijevala razlikovati. Kategoriju sam jednostavno nazvala „nejasne rečenice“.

Primjer 14. (Genom att offra sig själv fick han hemliga kunskafer om bland annat runor och genom att kasta sitt ena öga **i jätten Mimers brunn fick han del av alla Mimers kunskafer.**

(Značenje izvornika = *Žrtvujući samog sebe dobio je tajna saznanja o, između ostalog, runama, a bacivši jedno svoje oko u Mimirov zdenac dobio je dio od svih Mimirovih znanja.*)

Bio razapet, on je primio tajna znanja, uključujući runa i bacajući jedan od njegovih očiju **u veleslalomu Mimir je dobro, on je bio dio svega Mimer znanja.**

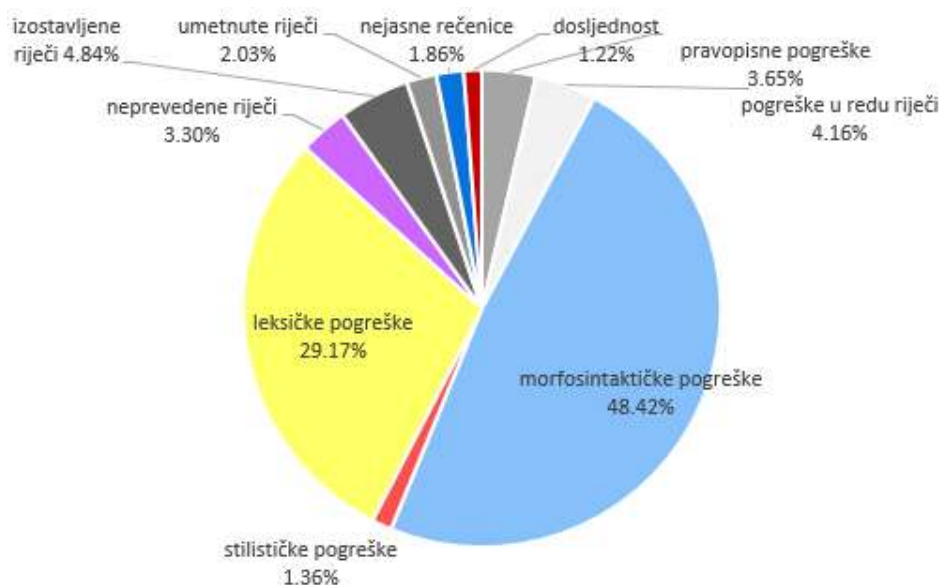
5.2 Metoda

Nakon što sam prikupila svih deset izvornika i odredila kategorije pogrešaka, prevela sam sve tekstove u sučelju GP-a te započela s analizom. Kako je određivanje pogrešaka prilično subjektivno i podložno mom shvaćanju pojedinih primjera, kvalitativni sam dio analize nadopunila kritičnim sagledavanjem izvornika (kako bih vidjela utječu li neka obilježja izvornika na pogreške u strojnim prijevodima) i ljudskih prijevoda (kako bih vidjela nailazi li i čovjek na slične poteškoće kao stroj). Također sam kvantitativnom analizom stekla uvid u najčešću vrstu pogrešaka kako ukupno tako i s obzirom na pojedine vrste tekstova. Valja napomenuti da se pogreške nisu određivale samo na jednoj razini, npr. na razini riječi, već da je to uvijek ovisilo o samoj prirodi pogreške. Jasno je da se kategorije neprevedenih, izostavljenih i umetnutih riječi manifestiraju na razini pojedinih riječi, dok leksička pogreška može biti i jedna riječ, ali i sintagma ili skup riječi. Pravopisne se pogreške manifestiraju na razini pojavnica, pogreške u redu riječi na razini rečenica, a kategorija dosljednosti čak na razini cijelog teksta. U sljedećem poglavlju (6.) donosim rezultate analize.

6. Rezultati

Gotovo polovica pogrešaka (48,42 %) u mom su korpusu morfosintaktičke prirode (v. sliku 1). Iako je zabilježeno 2271 takvih pogrešaka, treba se sjetiti da je logično da ih se pojavljuje razmjerno mnogo zbog toga što ta kategorija obuhvaća najviše različitih pogrešaka, ali i da takve pogreške ne predstavljaju osobito velik problem za čitatelja koji želi dobiti uvid u sadržaj teksta, već samo za onu osobu koja će taj tekst po potrebi redigirati. Možemo čak pretpostaviti da je potrebno manje vremena da bi se ispravio, primjerice, pogrešan padež, nego neka leksička pogreška. Njih je pak bilo 1368 (29,17 %) i to je druga najveća kategorija.

Sve ostale kategorije obuhvaćaju manje od 5 % pogrešaka: zabilježeno je 227 izostavljenih riječi (4,84 %), 195 pogrešaka u redu riječi (4,16 %), 171 pravopisnih pogrešaka (3,65 %), 155 neprevedenih riječi (3,30 %), 95 umetnutih riječi (2,03 %), 87 nejasnih rečenica (1,86 %), 64 stilističkih pogrešaka (1,36 %) te 57 pogrešaka u kategoriji dosljednosti (1,22 %).



Slika 1. Raspodjela različitih vrsta pogrešaka u korpusu.

6.1 Različite vrste izvornika

Kako se ukupni korpus sastojao od četiri različite vrste tekstova, u sljedećem sam koraku provjerila postoje li neke razlike u raspodjeli pogrešaka među njima (v. tablicu 1). U tablici je navedena vrsta izvornika, broj analiziranih kartica, ukupan

broj pogrešaka te broj pogrešaka po kartici, a u donjem dijelu tablice slijedi broj svih pojedinih vrsta pogrešaka izražen u postocima.

Tablica 1. Rezultati prema različitim vrstama tekstova.

Izvornici:	tehnički	politički	administrativni	znanstveni
Broj kartica:	7	18	23,9	30,8
Ukupan broj pogrešaka:	339	1021	1382	1735
Broj pogrešaka po kartici:	48,4	56,7	57,8	56,3
Pravopisne pogreške	2,9 %	2,9 %	4,1 %	4,2 %
Pogreške u redu riječi:	0,5 %	4,1 %	5 %	4,6 %
Morfosintaktičke pogreške:	39,2 %	53,5 %	40,7 %	59,2 %
Stilističke pogreške:	2,06 %	1,1 %	0,8 %	1,9 %
Leksičke pogreške:	29,7 %	28,9 %	32,3 %	30,2 %
Neprevedene riječi:	10,6 %	1 %	3,1 %	3,6 %
Izostavljene riječi:	1,1 %	4,7 %	4,9 %	6,1 %
Umetnute riječi:	1,1 %	2 %	1,8 %	2,5 %
Nejasne rečenice:	10,3 %	0,7 %	1,4 %	1,3 %
Dosljednost:	2 %	0,5 %	1,5 %	1,3 %

Pogled na brojke pokazuje da su pogreške razmjerno ravnomjerno raspoređene među različitim vrstama tekstova. Razlike u kategorijama pravopisnih i stilističkih pogrešaka te umetnutih riječi i dosljednosti manje su od 2 %, a čak je i u kategoriji leksičkih pogrešaka razlika malena (3,4 %). Veće se razlike zamjećuju u kategoriji morfosintaktičkih pogrešaka gdje je u znanstvenom korpusu zabilježeno njih 1028, što čini gotovo 60 % ukupnog udjela pogrešaka, a u administrativnom odnosno tehničkom korpusu zabilježeno ih je 563 odnosno 133, što je oko 40 %. Također je u tehničkom korpusu pronađeno manje pogrešaka u redu riječi (samo 2 odnosno 0,5 %) i manje izostavljenih riječi (samo 4 ili 1,1 %), no zato u njemu ima mnogo neprevedenih riječi (36 ili 10,6 %) i nejasnih rečenica (35 ili 10,3 %). U nastavku se objašnjavaju mogući razlozi za to (potpoglavlje 6.1.1), a potom slijede pojašnjenja ostalih rezultata, prema vrstama izvornika (potpoglavlja 6.1.2 – 6.1.4).

6.1.1 Tehnički korpus

Moja su očekivanja bila da će tehnički tekst zbog svoje prirode i kontroliranog jezika kojim je pisan biti najbolje preveden, no rezultati svjedoče o tome da GP nije imao dovoljno paralelnih tekstova na raspolaganju da bi mogao izvući dobre ekvivalente za tehničke termine koji su se u tom tekstu pojavljivali. GP je te termine najčešće ostavljao neprevedene ili bi ponudio u cijelosti nekoherentne

rečenice, što je rezultiralo time da ovaj korpus odskake od ostalih s udjelom pogrešaka u tim dvjema kategorijama (čak 10,6 % neprevedenih riječi i 10,3 % nejasnih rečenica, dok je prosjek ostalih tekstova 2,5 % odnosno 1,1 %). Tekst je zamišljen kao upute za ugrađivanje uređaja Prisma, no vjerujem da korisnik ne bi uspio s ugradnjom ako bi slijedio upute prevedene na ovakav način:

Primjer 15. (Eftersom ingen rörelse känns i tryckfronten, ...)

(Ljudski prijevod: Budući da prednja ploča ne daje nikakve podražaje...)

Budući da nema pomaka osjećaj pritisak front, ...

Zbog velikog je broja neprevedenih riječi i nejasnih rečenica ovdje zabilježen manji broj morfosintaktičkih pogrešaka nego u ostalim tekstovima („samo“ 39,2 %).

6.1.2 Politički korpus

Prijevod političkog teksta bio je značajno jasniji od većine ostalih prijevoda, što svjedoči o tome da GP lakše prevodi svakodnevnije tekstove koji se bave temama koje su aktualne u svim kulturama (poput feminizma, klasnog društva, patrijarhata, ekologije, rasizma, nacionalnih manjina itd.) jer uspijeva pronaći veći broj paralelnih tekstova koji mu pomažu u prijevodu. U tri presudne kategorije ovaj je tekst imao najmanji udio pogrešaka: leksičke pogreške (28,9 %), nejasne rečenice (0,7 %) i pogreške u dosljednosti (0,5 %), a čak je i stil odgovarao političkom programu. Ovaj je tekst bio dovoljno razumljiv da bi čitatelj mogao izvući osnovne informacije iz njega, no samo u tolikoj mjeri da procijeni je li mu tekst od dovoljne koristi da ga da prevesti na neki drugi način, a ne i da u potpunosti shvati sve misli. Slijede primjeri rečenica koje su razumljive, unatoč tome što sve sadrže određen broj pogrešaka:

Primjer 16. (I alla tider har människor drömt om och kämpat för en värld där ingen behöver gå hungrig, en värld utan förtryck och utan krig.)

U svim vremenima, ljudi su sanjali i borili za svijet u kojem nitko ne treba ići gladan, svijet bez pritiska i bez rata.

Primjer 17. (Vi menar att en hållbar ekologisk utveckling är intimt förknippad med en hållbar mänsklig utveckling.)

Vjerujemo da je održivi ekološki razvoj usko je vezan uz održivi razvoj društva.

Primjer 18. (Ett demokratikrav är att alla som tillhör en nationell minoritet ska ha möjlighet att medverka i beslut och ta tillvara sina intressen inom samtliga samhällsområden.)

Demokracija zahtijeva da svaka osoba koja pripada nacionalnoj manjini ima priliku sudjelovati u donošenju odluka i štititi svoje interese u svim područjima društva.

6.1.3 Administrativni korpus

Općenito se smatra da su administrativni tekstovi prikladni za strojno prevođenje. Uz dobre podatke za učenje mogu se postići rezultati koji će zahtijevati tek minimalnu redakturu. No kako GP nije specijaliziran ni za jedno područje, i ovdje su rezultati bili svakojaki. Zaista je zabilježen nešto manji udio morfosintaktičkih pogrešaka (40,7 %), ali zato najveći broj leksičkih pogrešaka (32,3 %). Upravo su ključni pojmovi bili pogrešno prevedeni, što je gotovo u potpunosti onemogućilo praćenje i razumijevanje prijevoda. Pri procjenjivanju (ne)razumljivosti prijevoda, valja gledati dublje od samih brojki jer se pokazalo da težina, a ne količina pogrešaka najviše utječe na to kakav je prijevod:

Primjer 19. (KORTFATTAD BESKRIVNING AV TÄVLINGENS SYFTE) (= naslov)

KRATKI OPIS DOGAĐAJA PREDMET organizatori

Primjer 20. (I vilket land ska momsens redovisas och betalas?) (= naslov)

U bilo kojoj zemlji, PDV se obračunava i plaća za?

Primjer 21. (Ditt tillstånd måste vara klart innan du reser till Sverige.)

Vaš uređaj mora biti završen prije putovanja u Švedsku.

Osim leksičkih odabira, osobito velik problem predstavljale su izostavljene riječi (kojih je zabilježeno 4,9 %). Najčešće su izostavljeni glagoli i veznici (čija je važnost za razumijevanje ključna) te riječi koje su sadržavale najvažnije semantičke informacije u tekstu, kao u sljedećem primjeru:

Primjer 22. (Utlänningar i Sverige **har** enligt grundlagen yttrandefrihet, religionsfrihet, förenings- och församlingsfrihet på samma sätt som svenska medborgare.)
(Značenje izvornika = *Strani državljani u Švedskoj prema Ustavu imaju pravo na slobodu izražavanja, vjerske slobodne, slobodu udruživanja i okupljanja jednako kao i švedski državljani.*)

Stranci u Švedskoj, **ø** prema Ustavu **ø** izražavanja, vjere, udruživanja i okupljanja na isti način kao i švedskih građana.

Još je jedan od indikatora lošeg strojnog prijevoda činjenica da je zabilježeno tek 0,8 % stilističkih pogrešaka. U načelu je teško uočiti stilističke pogreške u tekstovima u kojima je toliko riječi pogrešno prevedeno, a izabrani prijevodni ekvivalenti nimalo ne odgovaraju izvorniku.

Uzroke loših prijevoda treba tražiti i u izvornicima. Opet se pokazalo da je GP lakše prevodio one administrativne tekstove koji su se bavili učestalijim i aktualnijim temama, a lošije one koji su obilovali stručnijim terminima. Također su mnogi od ovih izvornika bili loše napisani i u njima sam zabilježila veći broj zatipaka i ostalih pogrešaka koje ljudski prevoditelj može lako zanemariti i svejedno dobro prevesti tekst, dok strojnom prevoditelju to predstavlja veće probleme. To su primjeri poput: „nägra” umjesto „några” ('neki'), „omrodet” umjesto „området” ('područje'), „maste” umjesto „måste” ('morati'), „nodvändigt” umjesto „nödvändigt” ('potreban'), „invandradre” umjesto „invandrare” ('useljnik'). Kada bi se ove pogreške prepravile, izlazni bi se podaci zasigurno popravili u tim rečenicama.

6.1.4 Znanstveni korpus

U znanstvenom je korpusu zabilježen najveći broj morfosintaktičkih pogrešaka (59,2 %), no generalno je GP ovdje pokazao i neke dobre maneuvre te je ponudio dobre prijevode na mjestima gdje su izvornici predstavljali određene poteškoće i za ljudske prevoditelje. Sljedeća je rečenica dobro prevedena što se vjerojatno može pripisati činjenici da postoji velik broj dobrih paralelnih tekstova ovog tipa u korpusu na kojemu je GP treniran:

Primjer 23. (Hållbar utveckling är en utveckling som tillfredsställer dagens behov utan att äventyra förutsättningarna för kommande generationer att tillfredsställa sina behov.)

Održivi razvoj je razvoj koji zadovoljava potrebe sadašnjosti bez ugrožavanja perspektive budućih generacija da zadovolje svoje potrebe.

I leksičkih je pogrešaka bilo mnogo (30,2 %), ali su katkad bila dovoljna samo jedan ili dva pogrešna leksička odabira da cijela rečenica postane nerazumljiva i neupotrebljiva, kao u sljedećem primjeru:

Primjer 24. (I dag kan forskarna **betrakta** det naturliga urvalet – **live**)

(Ljudski prijevod: Danas znanstvenici mogu **promatrati** proces prirodnog odabira – **uživo**.)

Danas, istraživači mogu **uzeti u obzir** prirodna selekcija - **živi**.

Od elemenata teksta koji predstavljaju poteškoće i ljudskim prevoditeljima izdvajam teške znanstvene termine, imena institucija i organizacija te osobito neologizme. U tim su me situacijama izlazni podaci znali i pozitivno iznenaditi. Ipak, problem je ovog korpusa to što je GP izostavio brojne riječi.

Najvećim se problemom pokazao odabrani popularno-znanstveni tekst. Radilo se o kratkom tekstu od 1,5 kartice koji je obilovao povijesnim imenima i koji je pisan nešto kreativnijim stilom, što se ipak pokazalo kao prevelik zalogaj za GP. Velika većina prijevoda nalikuje na nesuvislu nakupinu riječi, a ne na koherentan tekst:

Primjer 25. (Inte minst under 1100-talet härjades flera märlarorter av estniska och karelska sjöfarare, som sannolikt stod för brännandet av Sigtuna år 1187.)

(Ljudski prijevod: Nije manje važno ni to da su u 12. stoljeću estonski i karelijski moreplovci opustošili više sela na Mälarenu, a vjerojatno su bili odgovorni i za požar u Sigtuni 1187. godine.)

Ne manje važno 1100s opustošio nekoliko Malar sorti estonskim i karelijanski pomoraca, koji će vjerojatno činili spaljivanje sigtuna godine 1187th

6.2 Razlike i sličnosti među različitim vrstama tekstova

Ako se vratimo na razlike između različitih vrsta tekstova u mom korpusu, primjećujem da ih zapravo ima jako malo, no svi su tekstovi odabrani jer su pisani manje ili više kontroliranim jezikom, jednoznačnim rečenicama te jer pokušavaju

prenijeti jasnu poruku zainteresiranim čitateljima. No rezultati su u tolikoj mjeri loši da je kategorizacija pogrešaka mogla biti samo vrlo gruba. Zbog loših se izlaznih podataka ne može precizno odrediti što su najveće poteškoće pri prevođenju sa švedskog na hrvatski niti što bi Google trebao promijeniti u svojim algoritmima, no u sljedećem poglavlju opisujem ono što sam ipak zamijetila.

7. Diskusija

Iz mojih se rezultata teško može iščitati poveznica između količine i prirode pogrešaka te razine razumljivosti teksta. Pogreške su mnogobrojne i teške, a izlazni podaci većinski nerazumljivi. Čini se da se ovo istraživanje ne uklapa u prijašnja istraživanja o strojnom prevođenju u Hrvatskoj i inozemstvu koja pokazuju da strojno prevođenje funkcionira sve bolje te da postoji velika potreba za strojno prevedenim tekstovima, što je uvelike uzrokovano odabirom manje „popularnog“ jezičnog para (v. poglavlje 8.). Ovo je istraživanje pokazalo da GP sa švedskog na hrvatski lakše prevodi svakodnevnije tekstove (a ne tehničke ili administrativne, kao što se očekivalo). Mnoge od tih pogrešaka rezultat su upravo razlika između švedskog i hrvatskog, ali često i činjenice da se GP koristi engleskim kao međujezikom. Također, GP ne može prilagoditi prijevod ciljnom čitatelju niti uzeti u obzir kontekst (kao što ljudski prevoditelj može), a javljaju se i poteškoće pri prevođenju kulturno specifičnih pojmova.

7.1 Različite vrste pogrešaka

U ovom ću potpoglavlju dati primjere pogrešnih prijevodnih ekvivalenata koji su se osobito često ponavljali i koje sam mogla grupirati prema kategorijama: morfosintaktičke pogreške (poglavlje 7.1.1), leksičke pogreške (7.1.2), pravopisne pogreške (7.1.3) i neprevedene riječi (7.1.4).

7.1.1 Morfosintaktičke pogreške

Kao što je već bilo rečeno, morfosintaktičke pogreške obuhvaćaju najveći broj različitih pogrešaka. Hrvatski je jezik upravo na morfološkoj razini mnogo kompliciraniji od švedskog pa ne čudi velik broj zabilježenih pogrešaka u toj kategoriji. U nastavku sam izdvojila dva morfosintaktička obilježja švedskog koja

GP teško prevodi na hrvatski te dva morfosintaktička obilježja hrvatskog koja GP teško sam proizvodi kada su potrebna u prijevodu. Važno je razmisliti o tome kako bi se ove pogreške mogle izbjeći uvođenjem određenih algoritama te tako unaprijediti GP i popraviti izlazne podatke u cjelini.

7.1.1.1 Švedski formalni subjekt „det“

Švedski se jezik često služi formalnim subjektom „det“ za uvođenje rečenice, što je gotovo nepoznat pojam u hrvatskom jeziku. Zato je „det“ pogrešno preveden u svim zabilježenim situacijama u korpusu, a najčešća je metoda GP-a bila da doslovno prevede tu konstrukciju, kao u sljedećem primjeru:

Primjer 26. (**Det tillverkas** fyra olika modeller av den akustiska Prisma TS övergångssignalen.)

To proizvedene četiri različita modela akustične Prisma TS prijelaznog signala.

7.1.1.2 Švedski dvočlani glagoli

Takozvani „partikelverb“, dvočlani glagoli, još su jedno tipično obilježje švedskog, a GP ponovno najčešće doslovno prevodi glagolsku česticu, što čitatelja može samo zbuniti:

Primjer 27. (Vi människor har under alla tider funderat över hur vi har **kommit till**.)
Mi ljudi imamo u svakom trenutku razmišljao o tome kako smo **došli do**.
(Značenje dvočlanog glagola = „nastati“, „postati“.)

7.1.1.3 Povratne zamjenice u hrvatskom

Iako i u švedskom i u hrvatskom postoje povratne zamjenice, one nisu raspodijeljene jednako među glagolima jednog i drugog jezika nego je zadatak GP-a da raspozna kada je koji oblik potreban u hrvatskom. U sljedećem je primjeru povratna zamjenica u hrvatskom bila potrebna da bi zamijenila recipročnu konstrukciju u švedskom, no kako zamjenica formalno ne postoji u švedskoj rečenici, tako je GP nije ni sam proizveo:

Primjer 28. (Nu har forskarna skapat en katalog med exempel på hur snabbt effekten av naturlig selektion faktiskt **kan ses**.)

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

7.1.1.4 Dvostruke negacije u hrvatskom

Dvostruke su negacije u švedskom nepoznate, a u hrvatskom uobičajene. Iako bi bilo moguće zamisliti razmjerno jednostavan algoritam kojim bi se GP naučio da prepozna situacije u kojima bi trebao sam složiti dvostruku negaciju u prijevodu, zasad se još nije uspjelo u tome te su ti prijevodi većinom još uvijek manjkavi, kao u sljedećem primjeru:

Primjer 29. (**Inget steg** framåt **får** avfärdas som alltför obetydligt.)

Bez korak naprijed **treba** odbaciti kao previše beznačajan.

7.1.2 Leksičke pogreške

Istraživanja oko vrednovanja strojnog prevođenja pokazala su da leksičke pogreške znatno više utječu na razumljivost teksta od, primjerice, morfosintaktičkih pogrešaka (v. poglavlje 3.). Stoga je osobito važno razmišljati o algoritmima koji bi poboljšali taj aspekt prijevoda, uz to što sustav uvijek mora imati pristup što većem digitalnom materijalu. Kao i u slučaju morfosintaktičkih pogrešaka, u nastavku sam grupirala tri najčešće vrste leksičkih pogrešaka.

7.1.2.1 Suprotni prijevodi

Najočiglednija se vrsta leksičkih pogrešaka uočava kada GP prevodi sasvim jednostavne, jasne i svakodnevne riječi njihovim antonimima ili pak nudi suprotne prijevode unutar iste rečenice. Takvi prijevodi čitatelja veoma zbunjuju, kao u sljedećem primjeru:

Primjer 30. (...eller, om den platsen **inte är** känd, den plats dit programkortet skickas för att användas där)

...ili, ako **je** lokacija **nije** poznata, mjesto gdje se šalje program kartica koja će se koristiti tamo.

Još je gori problem ako prijevod čitatelju ponudi informaciju za koju bi on mogao pomisliti da je ispravna jer ne očekuje da će jednostavne riječi i konstrukcije poput „sällan“ („rijetko“), „på 50 års sikt“ („u sljedećih 50 godina“), „yngre“ („mlađi“) biti pogrešno prevedene. U sljedećim je primjerima leksičke pogreške teško prozreti bez poznavanja švedskog jezika i uspoređivanja s originalom:

Primjer 31. (Tillgången på arbetstillfällena var dessutom så god att det **sällan** förelåg svårigheter att erbjuda invandrare arbete.)

Dostupnost poslova je dodatno toliko dobar da je **često** bilo teško ponuditi imigranata posao,

Primjer 32. (Sammantaget bedömer man att energianvändningen i bebyggelsen kan halveras **på 50 års sikt.**)

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

Primjer 33. (När den havsvik som skulle forma Mälaren under **yngre** järnålder.)

Kada je ulazni koja će oblikovati Jezero Malaren tijekom **kasnog** željeznog doba.

7.1.2.2 Kratice

Problem s kraticama leži u tome da ih GP uglavnom ne može identificirati kao elemente koji predstavljaju nešto drugo, čak i u sljedećem primjeru gdje je kratica vizualno odvojena zagradom te napisana neposredno iza ispisanog pojašnjenja za koje stoji. Kako se ta kratica slučajno preklapa s engleskom riječju za mahovinu, GP ju je tako i percipirao i preveo s engleskog na hrvatski:

Primjer 34. (...den s.k. Mini One Stop Shop (**MOSS**))

...takozvani Mini Stop Shop Jedan (**mahovina**).

Uz to što GP griješi s kraticama tako da im pronalazi pogrešne leksičke ekvivalente, često kratice uopće ne prevodi ili ih ne deklinira (iako primjere donosim na ovom mjestu, pogreške sam ubrojala u pripadajuće kategorije):

Primjer 35. (© 2004 Prisma Teknik **AB**)

© 2004 Prisma Teknik **AB** (= Ostatak može ostati nepreveden jer se radi o imenu firme, ali ne i kratica „AB“ koja znači „d.d.“)

Primjer 36. (Den allra största delen av Sveriges rymdsatsningar, ca 65 %, går till ESA.)
Velika većina švedskih prostora naporima, oko 65%, ide na **ESA**. (= U ovom bi slučaju kratica odgovarala jer i na hrvatskom stoji za „Europsku svemirsku agenciju“, no kratica se mora deklinirati.)

7.1.2.3 Složenice

Složenice su mnogo uobičajenije u švedskom nego u hrvatskom, stoga prevođenje složenica često zahtijeva neko drugo sintaktičko rješenje u prijevodu. GP rijetko odabire dobre ekvivalente za složenice, a najčešće pogriješi tako da prevede samo jedan dio složenice (primjer 37), prevodi komponente složenice kao da se radi o dvjema nepovezanim riječima (primjer 38) ili pogrešno interpretira spojnicu (primjer 39):

Primjer 37. (Genom att värna naturens och de ekologiska kretsloppen kan människorna underhålla och utveckla sina egna **livsbetingelser**.)
Do očuvanje prirode i ekološke ciklusi mogu ljudi održavati i razvijati vlastiti **život**.

Primjer 38. (**Ljuskaraktär**: Gult sken, ca. 590 nm våglängd)
Svjetlo Karakter: žuto svjetlo, ca. 590 nm valne duljine.

Primjer 39. (...ett samhälle befriat från **klass-, köns- och etniskt förtryck**)
... društva slobodnog od **klase, spola i etničke represije**.

7.1.3 Pravopisne pogreške

U korpusu je zabilježeno 3,65 % pravopisnih pogrešaka. Izlazni se podaci i ovdje temelje na postojećim prijevodima, no trebalo bi poraditi na algoritmima koji bi naučili GP kako u krajnjoj verziji izbjeći pravopisne pogreške čak i kada ih preuzme. Dva su tipa takvih grešaka. Prve uključuju u potpunosti nemotivirano umetanje različitih interpunkcijskih znakova usred teksta te ispisivanje nasumičnih riječi velikim slovom, kao u sljedeća dva primjera:

Primjer 40. (Fattiga länders möjlighet **till utveckling** undergrävs genom att de högutbildade dras till de stora företagen i västvärlden.)
Potencijal siromašne zemlje **'za** razvoj potkopava visokoobrazovanih privučeni velikim tvrtkama na Zapadu.

Primjer 41. (På de nästföljande fyra mötena bedömdes de fem **bedömningsområdena** ett i taget (på första mötet bedömdes två områden) utifrån kraven i programmet och juryns synpunkter sammanställdes.)
U sljedeća četiri susreta procjenjuje na pet **područja OCJENA** jedan po jedan (prvi sastanak je ocijenjen dva područja) koji se temelji na zahtjevima programa i komentari žirija bili sastavio.

Iako takvi primjeri zbunjuju čitatelja, druga je vrsta još „opasnija“ jer nasumični pravopisni znakovi (najčešće zarezi) mogu promijeniti čitavo značenje rečenice:

Primjer 42. (Frågan om tillväxt måste kombineras med frågan om vad samhället ska **använda tillväxten** till och på vilket sätt tillväxten skapas.)
Pitanje rast mora biti u kombinaciji s pitanjem o tome što društvo treba **koristiti, rastu** i kako se stvara rast.

7.1.4 Neprevedene riječi

Često sam navodila kako sustav za strojno prevođenje mora imati pristup velikim arhivima prijevoda kako bi se smanjio broj leksičkih pogrešaka i neprevedenih riječi, no ovu sam posljednju kategoriju izdvojila upravo kako bih pokazala da ni sama količina dostupnih prijevoda ne može eliminirati takve pogreške. Kao što sam pokazala u potpoglavlju 7.1.2.1, GP katkad suprotno prevodi posve jednostavne riječi, a ponekad ih ne prevodi uopće, kao u sljedećem primjeru:

Primjer 43. (Klassernas sammansättning och **själv**uppfattning ändras med den ekonomiska utvecklingen.)
Sastav klase i **self**-percepcija mijenja s gospodarskog razvoja.

To se osobito jasno vidi na primjeru prevedenog kazala iz tehničkog teksta. Četiri se puta pojavljuje riječ „sid“ („stranica“) u izvorniku, svaki put u jednakom okruženju. Prvi i treći put GP je prevodi na odgovarajući način, drugi je put nije preveo uopće, a posljednji je put pridodao tu riječ prethodnoj jedinici kojoj „sid“ nikako ne pripada. Teško je razumjeti kako je moguće da je ta riječ ostala

neprevedena u drugom pojavljivanju ako je očito da GP u svojim korpusima može pronaći odgovarajući prijevod:

Primjer 44. (Generell beskrivning	sid 2
Sprängskiss	sid 2
Installation.....	sid 3
Kabelanslutning	sid 3)
Opći opis	stranica 2
Eksplozirala	sid 2
Instalacija	stranica 3
Kabelski priključak stranica	3

7.2 Ljudski prijevodi

Kako sam ciljano uz korpus švedskih izvornika prikupila i paralelni korpus s ljudskim prijevodima usporedno sam proučavala kako su se ljudski prevoditelji nosili s određenim poteškoćama. Zabilježila sam određene pogreške i u ljudskim prijevodima, no još češći bili su manevri kojima su ljudski prevoditelji katkad simplificirali prekomplikirane rečenice, spajali ili razdvajali rečenice kako bi dobili drukčiji ritam teksta ili na druge načine uzimali u obzir i mijenjali ton teksta. Konkretno pogreške najčešće su rezultat sitnijih propusta ili zabuna (koje se čovjeku događaju). Tako je u sljedećem primjeru prevoditelj zaboravio prevesti točku 6 i „ispravio“ pogrešku kada je također (zabunom) izostavio točku 8:

- Primjer 45. (6. Skruva fast ytterhöljet på stolpen.
7. Sätt tillbaka elektronikmodul och övriga delar i motsatt ordning, punkt 2 till 4.
 8. Anslut kabeln i stolpen enligt skiss nedan.
 9. Testa den akustiska funktionen och justera om så behövs, se sid 6.)
6. Vratite natrag elektronički modul i ostale dijelove prema odgovarajućem redoslijedu.
7. Spojite kabel za stup prema donjoj shemi.
 9. Provjerite zvučne funkcije i namjestite ih ponovno ako je potrebno (vidi str. 6).

Prevoditeljima se ponekad omakne te zaborave prevesti neku riječ, dio rečenice ili čak cijeli paragraf, a katkad pogrešno pročitaju ili shvate dio izvornika, kao što

vjerujem da se dogodilo u primjeru 45. U izvorniku piše da se mjerenja obično provode onda kada se očekuje da se količina klora *prestaje* povećavati u stratosferi („*sluta* att öka”), a ne onda kada se očekuje povećanje („öka”):

Primjer 46. (Mätningarna sker i en tid då halterna av klor förväntas **sluta att öka** i stratosfären.)

Mjerenja se obično provode onda kada se očekuje **povećanje** količine klora u stratosferi.

Ipak, najčešću ljudsku pogrešku čine zatipci, koje je teško zamijetiti čak i ako prevoditelj više puta naknadno iščita svoj prijevod:

Primjer 47. (Makt, inflytande och **materiell** levnadsstandard är idag ojämnt fördelade mellan etniskt segregerade bostadsområden.)

Moć, utjecaj i **finacijski** standard su danas neravnopravno raspodijeljeni između etnički odijeljenih dijelova gradova ili zemalja.

7.3 Hipoteze – potvrđene ili ne?

U ovom se potpoglavlju vraćam na svojih pet hipoteza:

1. Prva je hipoteza potvrđena: GP drži se strukture svih izvornih rečenica u korpusu osim onih u kojima se javlja kakav otežavajući čimbenik. To se, primjerice, odnosi na činjenicu da se u hrvatskom nakon godine piše točka, a u švedskom ne. GP bi katkad na hrvatskom dodao točku iza godine (što odgovara pravopisnom pravilu), ali bi onda s ostatkom materijala iste rečenice započeo novu rečenicu. U primjeru 48 je pak švedska rečenica završila godinom, a GP je točku koja označava kraj rečenice percipirao kao točku koja u hrvatskom slijedi iza godine te je unutar iste rečenice dodao rečenični materijal iz sljedeće:

Primjer 48. (Användningen av köpt energi inom sektorn minskar med minst 30 % till år 2025 jämfört med år **2000. Energianvändningen** är lägre år 2010 än år 1995.)

Korištenje kupljene energije u sektoru će se smanjiti za najmanje 30% do 2025. u odnosu na **2000. potrošnja energije** je niža u 2010. godini u odnosu na 1995.

- GP ne griješi u ortografiji pa je i druga hipoteza potvrđena. Kada se promatra način funkcioniranja statističkih sustava za strojno prevođenje, jasno je da je ova hipoteza morala biti potvrđena. Sustav izvlači dijelove iz postojećih i objavljenih prijevoda. Kada bi se zatipci javljali u izlaznim podacima to bi se jedino moglo objasniti time da je GP pronašao tek ograničeni broj paralelnih tekstova u kojima su se uvijek pojavljivali isti zatipci pa ih je GP također preuzeo. Na tragu toga objašnjavam i tri situacije u svom korpusu kada su se riječi „i u“ pojavile napisane skupa kao „iu“, za što ovdje navodim primjer:

Primjer 49. (Även **i fråga om** rätt till hälso- och sjukvård är utlänningar med tillstånd att stanna i Sverige jämställda med svenska medborgare.)

Čak **iu** slučaju prava na zdravstvenu zaštitu su stranci s dozvolom za boravak u Švedskoj izjednačen sa švedskim građana.

- Treća se hipoteza odnosila na utjecaj engleskog kao međujezika u prijevodima GP-a sa švedskog na hrvatski, odnosno provjerila sam pojavljuju li se engleske riječi u tim prijevodima. Hipoteza je potvrđena jer je gotovo polovica svih neprevedenih riječi u korpusu na engleskom jeziku (75 od 155 slučajeva):

Primjer 50. (Målgruppen för **Rymdkanalen** är i första hand ungdomar mellan 15-25 år.)
Ciljana publika **Space Channel** je prije svega mladi ljudi između 15-25 godina.

No, utjecaj engleskog pokazao se daleko širim od pukog pojavljivanja engleskih riječi u korpusu (koje čitatelj najčešće i razumije). Izdvajam primjer 51 u kojemu je GP bezličnu zamjenicu „man“ preveo na hrvatski kao „jedan“. Razlog pronalazim u tome što bi ekvivalent na engleskom glasio „one“ (što preuzima funkciju bezlične zamjenice) pa je GP, nakon odgovarajuće analize za prijevod na engleski, pogriješio u drugom koraku kada je promatrao samo engleski „one“ i na hrvatski to preveo denotativnim značenjem brojke, odnosno kao „jedan“:

Primjer 51. (Iakttar **man** livet runt sig, blir det uppenbart att arterna är anpassade till sin miljö.)

Jedan promatra život oko njega, postaje očito da su vrste prilagođen svojoj okolini.

GP također često percipira riječi kao engleske unatoč tome što je izvornik na švedskom (ako se radi o homografima), kao što se dogodilo s prijevodom kratice „MOSS“ (v. primjer 34) ili pak u sljedećem, smiješnom prijevodu:

Primjer 52. (**Love Dalén** och hans kolleger planerar nu en forskningsresa till regionen, för att ta reda på varför mammutarna dog ut även där.)

Ljubav Valley i njegovi kolege sada planiraju istraživački putovanje u regiji, kako bi saznali zašto su mamuti izumrli tamo.

4. Hipoteza koja glasi da GP griješi u redu riječi također je potvrđena. Zabilježeno je 195 takvih pogrešaka, unatoč tome što je hrvatska sintaksa mnogo slobodnija od švedske.
5. Hipoteza koja govori o tome da GP griješi pri uporabi zareza u prijevodima sa švedskog na hrvatski jezik također je potvrđena: 31 % svih pravopisnih pogrešaka odlazi na pogreške u zarezima. Zabilježene su 53 rečenice koje uključuju zareze na mjestima gdje se oni nikako ne bi smjeli pojaviti u hrvatskom jeziku te na mjestima u kojima zarezi nikako ne bi smjeli nedostajati.

Budući da se ovdje radi o pilot istraživanju, teško je bilo formulirati konkretnije hipoteze od ovih koje su sada potvrđene. Većina je prijašnjih istraživanja uključivala prijevode s engleskog ili na engleski jezik, bilo da se radi o istraživanjima iz Hrvatske ili o međunarodnima. Kako se moji rezultati u velikoj mjeri razlikuju od rezultata prijašnjih istraživanja, zaključujem da oni uvelike ovise o konkretnom jezičnom paru. Moja analiza dokazuje kako je situacija s hrvatskim i strojnoprevoditeljskim sustavima mnogo ozbiljnija nego što bi se možda na prvu pomislilo. U zaključku se bavim mogućim razlozima za to.

8. Zaključak

Kada je riječ o strojnom prevođenju ili općenito o uporabi strojeva u raznim područjima, ljudi se često boje da oni predstavljaju veliku opasnost za njihovu egzistenciju jer će ih zamijeniti. No teško je zamisliti da bi ovakvi rezultati kakve je pokazalo moje istraživanje mogli predstavljati neku ozbiljnu prijetnju profesionalnim prevoditeljima. Sustavi za strojno prevođenje nisu nikad ni trebali predstavljati „suparnika ljudskome prevođenju, već su ili pomagala za porast učinkovitosti u tehnici ljudskoga prevođenja ili sustavi za prevođenje onoga što ljudi ionako ne bi prevodili“ (Tadić 2003: 36). Međutim, s porastom utjecaja interneta postoji druga prijetnja koja sve više raste. Internet je danas važno oruđe

u poslovnom svijetu, znanosti, obrazovanju, ali i načinu kreiranja slobodnog vremena. Neprestano raste potreba za brzo dostupnim informacijama. Istraživanja o distribuciji jezika na internetu pokazala su da „engleski jezik dominira sa 84 % na web-stranicama“ (Dovedan, Seljan i Vučković 2002: 289). Ako se za hrvatski jezik ne razvije dovoljno jezičnih tehnologija kojima bi se korisnici iz Hrvatske mogli koristiti na mrežnom prostoru (gdje se u današnjem modernom i globaliziranom vremenu toliko toga događa), postoji opasnost da će se oni okrenuti prema engleskom te da bi hrvatski naposljetku mogao izumrijeti u mrežnom prostoru, što je jednako velika katastrofa za jedan jezik, kao i da nestane u stvarnoj uporabi.

Ovdje želim istaknuti sličnost situacije u kojoj se hrvatski jezik nalazi s istraživanjem maorskog jezika na kojemu je Manuirirangi (2011: 1) opisao koliko manji jezici pate od nedostatka resursa koji su potrebni da bi se razvile jezične tehnologije. U maorskom, ali i hrvatskom, između ostalog se uočava nedostatak:

1. ljudi (koji su voljni koristiti se modernim tehnologijama, koji imaju vremena i znanja da ih razvijaju, pa čak i koji imaju dovoljno povjerenja u jezične tehnologije i koji uviđaju prednosti u njihovom razvijanju u ovom elektronskom svijetu)
2. materijala (preduvjet za razvijanje jezičnih tehnologija jest postojanje većih količina digitalnog materijala)
3. povjerenja (u manjim jezičnim skupinama ljudi se više brinu oko privatnosti i nematerijalne imovine)
4. financijskih sredstava (razvijanje jezičnih tehnologija košta i to je velik problem u manjim jezičnim skupinama i zemljama) (ibid.).

U kontekstu sustava za strojno prevođenje, veseli me vidjeti da Google predano radi na ažuriranju verzija svog Prevoditelja. Ipak, rezultati prijevoda sa švedskog na hrvatski jezik pokazali su da se ne nalazimo na razini gdje bi riječ bila o sitnicama koje treba prepraviti, već da je potreban ozbiljan rad na tom i drugim sustavima kako bismo u budućnosti mogli očekivati bolje rezultate. Valja osvijestiti i da jezične tehnologije zauzimaju važnu ulogu u jačanju demokracije i pomažu građanima da se uključe u suodlučivanje o važnim odlukama (Sågvall Hein 2005:

23). Radi se o pravu svakog čovjeka na informaciju na vlastitom jeziku te da se njime može koristiti na internetu.

Zaključno, voljela bih da se nastave istraživanja o jezičnim tehnologijama i sustavima za strojno prevođenje u Hrvatskoj te da odgovorne službe uvide kako je potrebno ulagati više novca u njihov razvoj. Potrebno je usustaviti konkretne prijedloge za poboljšanje sustava za strojno prevođenje te osuvremeniti ovo istraživanje informacijama o novim ažuriranjima GP-a. Daljnja istraživanja moraju biti usmjerena prema osvjetljavanju prednosti i eventualnih nedostataka prelaska na neuronski sustav, a mogla bi i još konkretnije rasvijetliti situaciju s hrvatskim i strojnim prevođenjem, a tako dovesti i do poboljšanja ukupne situacije. Važno je da se u to uključe nadležne institucije te da se istraživanja u području jezičnih tehnologija sustavno potiču, a ne da sva ostanu na razini individualnih napora, pa bi se onda mogli očekivati i pozitivni rezultati u budućnosti.

Literatura

Primarna literatura

<http://darhiv.ffzg.unizg.hr/>. Pristup: 18. 4. 2017.

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

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

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

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

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

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

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

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

Sekundarna literatura

Ahrenberg, Lars i Merkel, Magnus. 1997. „Språkliga effekter av översättningssystem”. U: Josephson, Olle (ur.), *Ord&Stil* 28. Uppsala: Hallgren & Fallgren. 96-115.

Ahrenberg, Lars i Merkel, Magnus. 2006a. „Maskinöversättning – Automatisk översättning”. URL: <http://sprakteknologi.se/vad-aer-sprakteknologi/maskinoeversaettning/automatisk>. Pristup: 2. 12. 2016.

Ahrenberg, Lars i Merkel, Magnus. 2006b. „Maskinöversättning - Introduktion”. URL: <http://sprakteknologi.se/vad-aer-sprakteknologi/maskinoeversaettning/introduktion>. Pristup: 2. 12. 2016.

Andersson, Håkan i Johansson, Elin. 2010. *Automatisk översättning. En analys av Google Translate*. Diplomski rad. Stockholm: KTH Skolan för datavetenskap och kommunikation.

Bennet, Paul A. i sur. 1986. *Multilingual Aspects of Information Technology*. Biddles Ltd, Guildford and King's Lynn: Biddles Ltd.

Brkić, Marija, Kučiš, Vlasta i Seljan, Sanja. 2011. „Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs”. U: Billenness, Clive, Hemera, Annette, Mateljan, Vladimir, Banek Zorica, Mihaela, Stančić, Hrvoje i Seljan, Sanja (ur.), *INFuture 2011: „Information Sciences and e-Society*. Zagreb: Odsjek za informacijske znanosti, Filozofski fakultet, Sveučilište u Zagrebu. 331-344. URL: <http://infoz.ffzg.hr/INFuture/2011/papers/6-02%20Seljan,%20Brkic,%20Kucis,%20Evaluation%20of%20Free%20Online%20MTs.pdf>. Pristup: 1. 4. 2016.

Butler, Christopher S. 1992. *Computers and Written Texts*. Guildford: Biddles Ltd.

Castilho, Sheila, Moorkens, Joss, Gaspari, Federico i sur. 2017. „Is Neural Machine Translation the New State of the Art?”. *The Prague Bulletin of Mathematical Linguistics* 108(1). 109-120. URL: <https://www.degruyter.com/downloadpdf/j/pralin.2017.108.issue-1/pralin-2017-0013/pralin-2017-0013.pdf>. Pristup: 12. 10. 2017.

- Dovedan, Zdravko, Seljan, Sanja i Vučković, Kristina. 2002. „Strojno prevođenje kao pomoć u procesu komunikacije“. *Informatologia* 35 (4). 283-291.
- EAMT: European Association for Machine Translation. 2008. „Machine Translation“. URL: <http://www.eamt.org/>. Pristup: 2. 12. 2016.
- He, Wei i sur. 2016. „Improved Neural Machine Translation with SMT Features“. U: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI-16*. 151-157.
- Hedden, Thomas D. 2010. *Machine Translation: A Brief Introduction*. URL: http://www.hedden.org/intro_mt.html. Pristup: 1. 12. 2016.
- Koletnik Korošec, Melita. 2011. „The Internet, Google Translate and Google Translator Toolkit – Nuisance or Necessity in Translator Training?“ *Tralogy Symposium on Translation Careers and Technologies*. Pariz: 3. – 4. 3. 2011.
- Lindqvist, Amanda. 2012. *Maskinöversättning – bra eller dåligt? En översättningsvetenskaplig studie om typiska fel och begriplighet hos maskinöversatta facktexter*. Diplomski rad. Göteborg: Sveučilište u Göteborgu.
- Manuirirangi, Hori. 2011. *Minority Languages & Translation Technologies Case Study: Te Reo Māori & Google Translator Toolkit*. URL: <http://www.mt-archive.info/>. Pristup: 6. 12. 2016.
- Muegge, Uwe. 2009. „Dags att avliva myterna om maskinöversättning“. U: Orsén, Kerstin (ur.), *Facköversättaren* 4/09. Bohus: Sveriges facköversättarförening. 11-13.
- Pavlović, Nataša. 2017. „Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka“. U: Stolac, Diana i Vlastelić, Anastazija (ur.), *Jezik kao predmet proučavanja i jezik kao predmet poučavanja*. Zagreb: Srednja Europa. 279-295.
- Ranta, Arne. 2010. „När kan man lita på maskinöversättning“. U: Andersen, Sabine Kirchmeier, Breivik, Torbjørg, Domeij, Rickard, Halskov, Jakob, Langgård, Per i Nørstebø, Sjur (ur.), *Språkteknologi för ökad tillgänglighet*. Linköping: Linköpings University Electronic Press. 49-60.
- Simeon, Ivana. 2008. *Vrednovanje strojnoga prevođenja*. Doktorska disertacija. Zagreb: Filozofski fakultet.
- Sågvall Hein, Anna. 2005. „Datorn behöver statistik och grammatik“. *Språkvård* 1: 23-32.
- Sågvall Hein, Anna. 2008. „Maskinöversättning“. URL: <http://stp.lingfil.uu.se/~evapet/Undervisning/MT08/MT%2008%20F1.pdf>. Pristup: 2. 12. 2016.

Tadić, Marko. 2003. *Jezične tehnologije i hrvatski jezik*. Zagreb: Ex Libris.

Tidqvist, Sofia. 2015. *Maskinöversättning – funkar det? En komparativ studie av fel i maskinöversättningar från spanska till svenska med verktygen Google Översätt och Bing Översättare*. Diplomska rad. Göteborg: Sveučilište u Göteborgu.

Tomicic, Suzana. 2001. *Maskinöversättning*. Diplomska rad. Linköping: Sveučilište u Linköpingu.

Wu, Yonghui i sur. 2016. *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. URL:

<https://arxiv.org/pdf/1609.08144v2.pdf>. Pristup: 29. 12. 2017.