

# Računalno potpomognuto prevođenje materijala za učenje programiranja

---

**Manojlović, Maja**

**Master's thesis / Diplomski rad**

**2018**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Rijeka, Faculty of Humanities and Social Sciences / Sveučilište u Rijeci, Filozofski fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:186:173747>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-11-27**



*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



Filozofski fakultet u Rijeci – Odsjek za anglistiku

Diplomski studij engleskog jezika i književnosti i informatike – nastavnički  
smjer

Maja Manojlović

# Računalno potpomognuto prevođenje materijala za učenje programiranja

Diplomski rad

Mentorica: doc. dr. sc. Marija Brkić Bakarić

Komentorica: dr. sc. Martina Holenko Dlab

Rijeka, 2018.

# SADRŽAJ

SAŽETAK.....	3
1. UVOD.....	4
2. RAZVOJ STROJNOG I RAČUNALNO POTPOMOĞNUTOG PREVOĐENJA .....	7
2.1. STROJNO PREVOĐENJE KROZ POVIJEST .....	8
2.2. ALATI ZA STROJNO PREVOĐENJE.....	9
2.3. USPOREDBA STROJNOG I RAČUNALNO POTPOMOĞNUTOG PREVOĐENJA .....	9
2.4. BUDUĆNOST RAČUNALNO POTPOMOĞNUTOG PREVOĐENJA.....	10
3. RESURSI PREVODITELJA .....	12
4. ODABRANI ALATI .....	13
4.1. GOOGLE TRANSLATOR TOOLKIT .....	13
4.2. SMARTCAT .....	21
5. PREVOĐENJE POMOĆU GOOGLE TRANSLATOR TOOLKITA .....	34
5.1. PROJEKTI „SRETAN ROĐENDAN“ I „ISPRIČAJ PRIČU“ .....	35
5.2. BILJEŠKE PROJEKATA „SRETAN ROĐENDAN“ I „ISPRIČAJ PRIČU“ .....	40
6. PREVOĐENJE POMOĆU SMARTCATA .....	43
6.1. PROJEKT „ISTJERIVAČI DUHOVA“ .....	45
6.2. PROJEKT „MAČKE!“ .....	49
6.3. PROJEKT „SMEĆE IZ SVEMIRA“ .....	50
6.4. BILJEŠKE PROJEKTA „ISTJERIVAČI DUHOVA“ .....	53
6.5. BILJEŠKE PROJEKTA „SMEĆE IZ SVEMIRA“ .....	56
6.6. USPOREDBA PRIJEVODA.....	57
7. USPOREDBA ODABRANIH ALATA .....	62
8. ZAKLJUČAK .....	64
LITERATURA.....	65
POPIS SLIKA .....	66
POPIS TABLICA.....	68
POPIS GRAFIKONA .....	69

## SAŽETAK

Računalno potpomognuto prevođenje postalo je neizostavan dio prevoditeljske prakse. Svakodnevnim razvojem prevoditeljskih alata poboljšava se kvaliteta i brzina procesa prevođenja. Ovaj diplomski rad uspoređuje dva alata za strojno potpomognuto prevođenje i njihovo manipuliranje tekstem. Sustavi korišteni u svrhu ovog rada su Google Translator Toolkit i Smartcat. Za prevođenje su korišteni materijali Code Cluba za učenje programiranja.

U radu se prvo opisuje razvoj strojnog i strojno potpomognutog prevođenja. Zatim se ukratko navode i definiraju najčešće korišteni resursi prevoditelja, a potom se detaljno opisuju značajke odabranih alata. Konačno, opisuje se postupak prevođenja pomoću odabranih alata, prikazuje statistika prijevoda te uspoređuju rezultati alata.

Ključne riječi: strojno prevođenje, strojno potpomognuto prevođenje, prijevodna memorija, rječnik, automatsko pretprevođenje, *CAT* alat, Google Translator Toolkit, Smartcat

# 1. UVOD

Tema ovog diplomskog rada je analiza alata za strojno potpomognuto prevođenje kroz njihovu praktičnu primjenu na materijalima za učenje programiranja. Rad se sastoji od teorijskog dijela u kojem se navodi povijesni razvoj strojnog i strojno potpomognutog prevođenja, resursi prevoditelja te značajke Google Translator Toolkita i Smartcata. Uz to, u radu je implementiran i praktični dio koji se fokusira na analizu i zapažanja o procesu prevođenja pomoću Google Translator Toolkita i Smartcata.

Razvoj informacijske tehnologije i moderne komunikacije uvjetovao je potrebu za automatizacijom procesa prevođenja, a zbog mogućnosti brzog i jednostavnog pristupa sve većem broju informacija računala su postala primarna sredstva informiranja i komunikacije. U moru informacija rasla je i potreba za razumijevanjem istih te za prevladavanjem jezične barijere. U novonastaloj okolini željnoj sporazumijevanja nastojalo se očuvati neovisnost i kulturološki identitet te je profesija prevoditelja sve više jačala i razvijala se. Ipak, na tradicionalno prevođenje gledalo se kao na neekonomičan i vremenski zahtjevan posao te se težilo pronalasku novčano prihvatljivijeg načina prevođenja. Osim na profitabilnosti naglasak je i dalje bio na kvaliteti. Tada započinje daljnji tehnološki razvoj na području prevoditeljskih alata. Iako je strojno prevođenje započelo svoj razvoj kao posljedica potrebe za pronalaskom brzog načina jezičnog sporazumijevanja njegovo usavršavanje nije teklo bez prepreka. Kompleksnost procesa prevođenja nije se mogla pretočiti u alat koji bi samostalno producirao kvalitetne prijevode u svakoj situaciji.

Nakon bezuspješnog pokušaja kreiranja savršenog strojnog prevoditelja, promijenila se perspektiva gledanja na prevođenje te ga se počelo promatrati kao dinamički proces koji objedinjuje različite lingvističke aspekte. Poticao se razvoj alata koji će služiti kao pomoć prevoditeljima, umjesto alata koji će u potpunosti zamijeniti prevoditelje. Polazna misao bila je kako alati za strojno prevođenje ne moraju biti sposobni za samostalan rad, već da moraju biti dovoljno kvalitetni kako bi pomogli prevoditelju u radu. Tako je fokus sa strojnog prevođenja prešao na strojno potpomognuto prevođenje, a započeo je i ubrzan razvoj mnogobrojnih alata i programa namijenjenih lakšem radu prevoditelja, poput terminoloških baza podataka, dvojezičnih konkordancija, prijevodnih memorija i brojnih drugih. Suprotno strojnom prevođenju, kojem je cilj bilo samostalno prevođenje, svrha strojno potpomognutog prevođenja i njegovih alata je prvenstveno pomoć ljudskom prevoditelju u procesu prevođenja. Proces prevođenja nedvojbeno je lakši otkada postoje alati poput prijevodnih

memorija koji ubrzavaju rad prevoditelja i osiguravaju kvalitetu i konzistentnost. Unatoč strahu od mogućeg gubitka posla i opadanja kvalitete prijevoda, od ključne je važnosti da prevoditelji idu u korak s razvojem informacijske tehnologije te da iskoriste sav potencijal koji im pruža. Nužno je da percipiraju važnost primjene dostupnih prevodilačkih alata u svojoj praksi kako bi povećali produktivnost i kvalitetu rada.

U radu sa Google Translator Toolkitom korišten je korpus od četiri dokumenta koji pokrivaju cjelinu o HTML i CSS jezicima za izradu i uređivanje web stranica. Materijali korišteni za prevođenje preuzeti su sa web stranice Code Cluba<sup>1</sup>. Code Club je svjetska mreža klubova koje vode volonteri, a cilj im je približiti programiranje djeci u dobi od 9 do 13 godina. Prevodili su se dokumenti u .doc formatu naziva „*Sretan rođendan*“ („*Happy birthday*“) i „*Ispričaj priču*“ („*Tell a story*“) te bilješke spomenutih projekata u .html formatu. Projekti sadrže zadatke i izazove koje djeca moraju riješiti u svrhu upoznavanja s HTML i CSS programskim jezicima, a bilješke su namijenjene volonterima kluba te sadrže pojašnjenja zadataka, korištene resurse i ciljeve i ishode učenja. Korpus korišten u analizi rada sa Smartcatom sastoji se od pet dokumenata i pokrivaju cjelinu o učenju programiranja u Scratchu. Tri su dokumenta projekti naziva „*Istjerivači duhova*“ („*Ghostbusters*“), „*Mačke!*“ („*Cats!*“) i „*Smeće iz svemira*“ („*Space Junk*“), a uz njih su prevedene i bilješke projekata za voditelje kluba „*Istjerivači duhova*“ i „*Smeće iz svemira*“. Sa projektima „*Istjerivači duhova*“ i „*Smeće iz svemira*“ radilo se u .doc formatu, a sa projektom „*Mačke!*“ u .pdf formatu. Bilješke projekata preuzete su u .html formatu. U dogovoru sa jezičnim koordinatorom Code Cluba neki od prijevoda materijala korišteni u ovom radu dostavljeni su u njihovu direkciju te objavljeni na službenoj web stranici Code Cluba.

Cilj diplomskog rada bio je isprobati rad s dva različita alata za strojno potpomognuto prevođenje te koristiti različite funkcionalnosti koje nude u svrhu što kvalitetnije analize dobivenih podataka. Uz to, težilo se usporediti način na koji alati barataju različitim dokumentima i njihovim sadržajem. Prikazana je statistika vezana uz različite aspekte projekata i izvršena je usporedba dvaju alata imajući na umu različite mogućnosti koje pružaju. Pretpostavka je bila da će se rad u Smartcatu pokazati superiornijim te da će dobiveni prijevodi biti kvalitetniji i brže dovršeni. Temelj pretpostavke je činjenica da je Smartcat mnogo kompleksniji alat te u konačnici pruža više mogućnosti od Google Translator Toolkita. U analizi se u obzir uzimalo vrijeme potrebno za prevođenje pojedinih materijala, greške koje su se javljale, konačni rezultati prevođenja te osobno zadovoljstvo alatima.

---

<sup>1</sup> Stranice Code Cluba: <https://www.codeclub.org.uk/> (dohvaćeno: srpanj 2018.)

U prvom poglavlju navode se neke od povijesnih činjenica vezanih uz strojno i strojno potpomognuto prevođenje, kao i definicije tih termina te razlike, odnosno sličnosti, među njima. Zatim se, u drugom poglavlju, pobliže definiraju značajke i mogućnosti već spomenutih alata, njihove prednosti i nedostaci te općenito princip rada. U poglavljima koja slijede naglasak je na praktičnom radu u alatima te analizi i prikazu statistike pojedinih materijala. Konačno, u zaključku se navode osnovna zapažanja rada.

## **2. RAZVOJ STROJNOG I RAČUNALNO POTPOMOĞNUTOG PREVOĐENJA**

Napredak u razvoju informacijske tehnologije u kombinaciji sa zahtjevima moderne komunikacije doveli su do sve većeg razvoja automatizacije procesa prevođenja. Globalizacijom su se razbile kulturološke, ekonomske i lingvističke barijere, a potreba za komunikacijom i međusobnim razumijevanjem postala je sve veća. Trenutačna dostupnost informacija uvjetovala je sve češće oslanjanje na računala kao sredstvo informiranja i komunikacije. Nove tehnologije također pružaju veću fleksibilnost od papirnatih medija u smislu razmjene svakojakih proizvoda i usluga. Brže su, prikladnije su i jednostavnije za korištenje, a uz to su i rentabilnije. Sve to potaklo je potrebu zemalja diljem svijeta za suradnjom u različitim sferama života, kao što su ekologija, ekonomija i obrazovanje.

U suvremenom komunikacijskom svijetu računala i mobiteli postali su primaran medij za prenošenje informacija. Time je utjecaj na prirodu i proces prevođenja sve više jačao, a različiti aspekti modernog života uvjetovali su potrebu za bržim i efikasnijim metodama samog procesa prevođenja. Nepoznavanje jezika često se poistovjećuje sa ograničenim pristupom informacijama (Delpech, 2014.). Kako bi se te barijere umanjile i omogućio slobodan protok informacija, morali su se prvo razriješiti novonastali lingvistički problemi. Došlo je do sukoba očuvanja materinjeg jezika i pripadnosti i potrebe za razumijevanjem drugih kultura i nacionalnosti. Profesija prevoditelja jačala je kao posljedica potrebe zadržavanja neovisnosti i kulturološkog identiteta (Kučiš, 2010.). Jezična raznolikost nije smjela postati prepreka međusobnom razumijevanju.

Različiti aspekti modernog života doveli su do potrebe za efikasnijim metodama prevođenja (Craciunescu & Gerding-Salas, 2004.). Često se prevođenje izbjegavalo pod izgovorom da rezultira prevelikim troškovima, a uz to je i dugotrajno. Prema nekim istraživanjima zaključeno je kako prosječan prevoditelj nije u mogućnosti kvalitetno obraditi više od 4 do 6 kartica teksta dnevno, što je otprilike 2500 riječi. Ekonomska potreba za pronalaženjem financijski prihvatljivijih rješenja rezultirala je prepuštanjem prevođenja neadekvatnim individualcima u svrhu financijskog profita. Nastojanje da se pronade ekonomičnije rješenje međunarodne komunikacije, a da se pritom prevodilačka profesija ne degradira, dovelo je do daljnjeg tehnološkog napretka na području razvoja prevoditeljskih alata koji će zadovoljiti potrebe prevoditelja za trenutno dostupnim informacijama i pristupom



ekstenzivnim bazama podataka. Tako su danas elektronički rječnici, prijevodne memorije i terminološke baze podataka neizostavan dio prevoditeljske prakse (Kučiš, 2010.).

## 2.1. STROJNO PREVOĐENJE KROZ POVIJEST

Petr Smirnov-Troyanskii je 1930-ih godina razvio ideju o sustavu automatiziranog prevođenja koji se odvija kroz tri faze. U prvoj fazi lingvist koji poznaje izvorni jezik analizira riječi i pretvara ih u njihove osnovne oblike u skladu sa sintaktičkim pravilima. Zatim stroj organizira spomenute osnovne oblike u rečenice na ciljnom jeziku. Konačno, drugi lingvist, koji poznaje ciljni jezik, analizira i ispravlja taj produkt. Iako je Smirnov-Troyanskii imao inovativnu ideju, ovaj princip prošao je relativno neprimijećeno sve do kasnih 1950-ih godina. Za vrijeme Hladnog rata, koji je svoj vrhunac dostigao 50-ih godina 20. stoljeća, došlo je do sve većih nadmetanja u sferi komunikacije između SAD-a i sovjetskog Saveza. Tada se prevođenje sve više ispreplitalo sa novim tehnologijama te se nastojalo pronaći što brža i efikasnija rješenja cjelokupnog procesa. U to doba postalo je evidentno koliko ljudski prijevod iziskuje vremena i truda te koliko je neučinkovit i spor kada je u pitanju luksuz poput vremena. Potreba za osmišljavanjem stroja koji bi uklonio ove nedostatke bila je sve veća. Nakon toga, razvoj računala doveo je i do sve hrabrijih pokušaja njegovog korištenja u svrhu prijevoda prirodnih jezika.

Prve inačice programa za strojno prevođenje uporište su pronašle u detaljnim dvojezičnim rječnicima koji su nudili nekolicinu ekvivalentnih riječi u ciljnom jeziku za svaku riječ u izvornom jeziku. Također su se temeljile na nizu pravila o poretku riječi u rečenici. Ipak, kompleksnost procesa prevođenja zahtijevala je daljnji napredak u poboljšanju programa, koji su išli u korak s naprecima u lingvistici. Unatoč velikom trudu, rezultati nisu bili obećavajući što je dovelo do formiranja ALPAC<sup>2</sup> odbora čija je svrha bila evaluirati programe za strojno prevođenje. Odbor je donio zaključak da je strojno prevođenje spor i neefikasan proces koji je dvostruko skuplji od ljudskog prevođenja te da je investiranje u njegovo istraživanje beskorisno. S druge strane, odbor se složio da treba poticati razvoj alata koji će pomagati prevoditeljima u prevođenju, poput rječnika, baza podataka i slično, umjesto sav fokus usmjerivati na razvoj programa koji će samostalno kvalitetno prevoditi.

1990-ih godina došlo je do ključnih napredaka i promjene strategije pristupa strojnom prevođenju. Na prevođenje se prestalo gledati kao na statički proces temeljen na fiksnim

---

<sup>2</sup> Engl. *Automatic Language Processing Committee*

gramatičkim pravilima i počelo ga percipirati kao dinamički korpus koji se konstantno mijenja i prilagođava ovisno o svojim korisnicima i socijalnim i kulturološkim obilježjima. U obzir se uzela činjenica kako programi za strojno prevođenje ne moraju nužno biti savršeni i adekvatni za samostalan rad, već da samo moraju biti dovoljno dobri i kvalitetni kako bi pomogli prevoditelju u procesu prevođenja. Tako se sve više prakticiralo strojno potpomognuto prevođenje – proces koji uključuje korištenje specifičnih alata i tehnologija prilagodljivih potrebama prevoditelja. Počeo je razvoj mnogobrojnih alata i programa namijenjenih lakšem radu prevoditelja, poput terminoloških baza podataka, dvojezičnih konkordancija, prijevodnih memorija i brojnih drugih (Craciunescu & Gerding-Salas, 2004.).

## **2.2. ALATI ZA STROJNO PREVOĐENJE**

Strojno potpomognuto prevođenje odlikuje se raznim programima, no najvažniji i najkorisniji svakom prevoditelju su rječnici, konkordancije, prijevodne memorije i dvojezični *online* tekstovi.

Rječnici omogućuju prevoditeljima brz pristup potrebnim pojmovima i terminima, pružaju njihove definicije i prijeводе te u nekim slučajevima i kontekst i primjere rečenica. Iako ih je koristiti, pristupačni su, štede vrijeme i moguće je koristiti više od jednog rječnika simultano. Prijevodne memorije su svojevrsne baze podataka u koje se spremaju segmenti izvornog jezika u kombinaciji sa segmentima ciljnog jezika, odnosno dvojezični parovi segmenata. Kada prevoditelj naiđe na sličan ili isti segment u svom radu, program mu automatski ponudi prijedlog prijevoda iz prijevodne memorije. Na taj način olakšava se i ubrzava proces prevođenja jer prevoditelj može samo odabrati predloženo, umjesto da ručno upisuje prijevod. Dvojezični *online* tekstovi pružaju rješenja za fiksne fraze, dok konkordancije nude popis svih pojavljivanja niza slova unutar definiranog korpusa i na taj način definiraju određene uzorke koji se ponavljaju (Craciunescu & Gerding-Salas, 2004.).

## **2.3. USPOREDBA STROJNOG I RAČUNALNO POTPOMOGNUTOG PREVOĐENJA**

Iako su pojam strojnog prevođenja<sup>3</sup> i računalno potpomognutog prevođenja<sup>4</sup> usko povezani i nerijetko se preklapaju, važno je napomenuti da između njih ipak postoje nijanse

---

<sup>3</sup> Engl. *MT = Machine Translation*

koje ih razlikuju. Prema (Kučiš, 2010.), ovo su dvije „autonomne i potpuno različite tehnologije koje su nastale kao posljedica različitih pristupa u procesu prevođenja.“ S obzirom da nisu iste tehnologije, nije neobično što ne luče iste rezultate i što se koriste u različitim kontekstima.

Dok je u srži strojnog prevođenja samostalno prevođenje teksta s izvornog jezika na ciljni jezik bez intervencije čovjeka, računalno potpomognuto prevođenje sa svojim alatima služi isključivo kao pomoć ljudskom prevoditelju u procesu prevođenja. Strojno prevođenje koristi se matematičkom sposobnosti računala da na temelju brzih kalkulacija u analizi strukture teksta ili rečenice rastavi tekst na prevodive osnovne elemente te na osnovu toga generira prijevod sa ekvivalentnom jezičnom i kulturološkom strukturom u ciljnom jeziku. Računalno potpomognuto prevođenje, tj. *CAT*, s druge strane nudi prevoditelju niz alata i resursa koji mu pomažu u realiziranju procesa prevođenja. Također, *CAT* donosi inovacije i u pristupu procesiranja izvornog i ciljnog teksta. Bilo koji dijelovi teksta mogu se izmijeniti u bilo kojem trenutku, a moguće je izraditi i više verzija s ciljem usporedbe i evaluacije. Sve to omogućuje prevoditelju fleksibilniji i efikasniji rad i, u konačnici, poboljšanje kvalitete prijevoda. Uz to, svaki prevoditelj može manipulirati radnim okruženjem i podesiti ga na način na koji njemu najbolje odgovara i u kojem se osjeća najproduktivnije (Kučiš, 2010.).

## **2.4. BUDUĆNOST RAČUNALNO POTPOMOGNUTOG PREVOĐENJA**

Ono što je bilo prisutno od početka razvoja strojnog i strojno potpomognutog prevođenja jest strah da će ovakvi alati pretvoriti prevoditelje u ništa drugo nego uređivače teksta i učiniti ih manje važnima od računalnih programa. Strah od mogućeg gubitka posla, ali i opadanja u kvaliteti prijevoda uzrokovao je da mnogi prevoditelji i dan danas na neki način odbijaju prihvatiti razvoj novih tehnologija usmjerenih ka strojno potpomognutom prevođenju. Unatoč tome, bitno je naglasiti da strojno prevođenje nije dovoljno razvijeno niti adekvatno da bi potpuno eliminiralo potrebu za ljudskim prijevodom. Ne postoji takav strojni prevoditelj koji bi producirao savršen prijevod s jednog jezika na drugi bez ikakve ljudske intervencije jer jezik kao takav značajno ovisi o kontekstu i dubljem razumijevanju. Čovjek je i dalje ključan faktor u realizaciji kvalitetnog i profesionalnog prijevoda. Kontekst kao takav nije uvijek moguće pružiti zbog čega je strojno prevođenje limitirano na situacije kada je kontekst nepotreban za produkciju kvalitetnog prijevoda.

---

<sup>4</sup> Engl. *CAT* = *Computer Assisted/Aided Translation*

Prevoditelji trebaju prepoznati prednosti razvoja tehnologije i naučiti kako iskoristiti njihov potencijal u svoju korist bez da se osjećaju ugroženima. Razvoj informacijske tehnologije otvara brojna vrata prevoditeljima i ako ih se maksimalno iskoristi mogu uvelike utjecati na boljitak produktivnosti i poboljšanje kvalitete prijevoda (Delpech, 2014.). Prevodilački alati suvremenog doba mijenjaju odnos između prevoditelja i teksta, točnije pružaju novo iskustvo i način prevođenja. Iako je računalno potpomognuto prevođenje neusporedivo s tradicionalnim prevođenjem uz papir i olovku, neupitna je činjenica da računalno potpomognuti programi omogućavaju prevoditelju fleksibilniji rad na tekstu i višedimenzionalan pristup prevođenju. Iako već uznapredovala, prevoditeljska praksa tek je na početku svog razvoja. Uz konstantan razvoj informacijske tehnologije i prevodilačkih alata nužno je percipirati važnost implementacije i primjene istih u procesu prevođenja. Njihov se potencijal treba prihvatiti i iskoristiti za povećanje produktivnosti, konkurentnosti i kvalitete rada te na njega gledati kao na sastavni dio prevoditeljske realnosti (Kučiš, 2010.).

### 3. RESURSI PREVODITELJA

U prethodnom poglavlju spomenuti su neki od resursa koje prevoditelji koriste u svom radu pa su tako navedene konkordancije, dvojezični *online* tekstovi, prijevodne memorije i rječnici. U ovom poglavlju naglasak će biti na dva najvažnija, odnosno najčešće korištena, resursa koja se koriste u tradicionalnom, ali i strojno potpomognutom prevođenju.

Ono što najviše odlikuje alate za strojno potpomognuto prevođenje je mogućnost korištenja prijevodnih memorija i rječnika. Prijevodna memorija je svojevrsna baza podataka u koju se spremaju prijevodi na kojima korisnik radi, a koji se zatim mogu koristiti za kasniju upotrebu, odnosno kao pomoć pri radu na drugim prijevodima. Točnije, prijevodna memorija pohranjuje prevedene segmente i njihove ekvivalente u izvornom jeziku. Segmenti mogu biti bilo što – rečenice, odlomci, naslovi, elementi liste i slično. Korištenje prijevodnih memorija osigurava konzistentnost u dokumentu, olakšava prevođenje tekstova koji sadrže specijaliziran vokabular te ubrzavaju sam proces prevođenja i tako štede vrijeme. Ipak, kao i svemu, i prijevodnim memorijama treba pristupati s oprezom te imati na umu da ponekad prijevod iz prijevodne memorije nije prikladan u danom kontekstu.

Još jedan nezaobilazni element strojno potpomognutog prevođenja su rječnici. Oni također osiguravaju konzistentnost, olakšavaju prevođenje i štede vrijeme prevoditelju. Sadrže popis termina prevedenih na jedan ili više jezika, njihove definicije, komentare i bilo kakve dodatne informacije koje prevoditelj smatra relevantnima, poput primjera upotrebe u rečenici. Korisni su za spremanje specifičnih termina, značenja kratica i akronima te fraza specifičnih za tvrtku, primjerice slogana, naslova i slično, kao i za popisivanje termina koji se ne smiju prevoditi, primjerice imena.

Iako u današnje vrijeme postoje brojni resursi implementirani u *CAT* alate, najučestalije se koriste upravo prijevodne memorije i rječnici. Jednostavno se koriste, a značajno olakšavaju rad prevoditeljima. Pomoću njih ubrzava se cjelokupan proces prevođenja i, što je najvažnije, osigurava konzistentnost i kvaliteta u dokumentu, a također je prednost što se mogu koristiti iznova na različitim projektima.

## 4. ODABRANI ALATI

### 4.1. GOOGLE TRANSLATOR TOOLKIT

Google Translator Toolkit je *online* uređivač teksta koji omogućava korisniku izmjenu i dijeljenje prijevoda generiranih Google Translate-om. Google Translate je dobro poznata *online* usluga koja u trenutku prevodi rečenice, riječi i fraze. Alat je razvijen od strane Google Inc. kompanije te je predstavljen 8. lipnja 2009. godine. Iako zamišljen kao paket za upravljanje procesima i naziva *Google Translation Center*, Toolkit je promijenio ime i postao još jedan prijevodni alat za osobnu upotrebu (Google Translator Toolkit, 2018.). Google je za ovaj alat strateški i pametno odabrao nekoliko najvažnijih i najboljih funkcija i obilježja *CAT* alata dostupnih na tržištu. Iako naglasak nije toliko na inovativnosti koliko na objedinjavanju različitih funkcija koje zasebno već postoje u drugima alatima za prevoditelje, upravljanje i koordiniranje grupnog rada u Toolkitu je na razini skupljih TM<sup>5</sup> brendova i sustava.

U doba kada prijevodne memorije i alati postaju sve kompleksniji i skuplji, s naglaskom na korištenje od strane eksperata i stručnjaka, Google Translator Toolkit naglašava jednostavnost cijelog koncepta strojno potpomognutog prevođenja. Uzevši to u obzir, alat je jednostavan za korištenje – sučelje je intuitivno, jednostavno i minimalističko sa nekoliko osnovnih opcija, a dokumentacija je sažeta i jasna. Započeti je lako – korisnik jednostavno prenosi dokument sa računala, unosi web adresu, kratak tekst ili čak naziv članka objavljenog na Wikipediji. Zatim kreira prijevodnu memoriju i rječnik te odabire da li ih želi dijeliti i s kim. Početne postavke automatski postavljaju memoriju na globalno dijeljenje. Nakon toga dokument je spreman za prevođenje. Izgled početne stranice Toolkita prikazan je na slici 1.

---

<sup>5</sup> Engl. *Translation Memory* = prijevodna memorija

## What would you like to get translated?

Add content to translate

## What languages would you like to translate into?

- |  |   |                                   |
|--|---|-----------------------------------|
| <input type="checkbox"/> Arabic                | <input type="checkbox"/> Chinese (Simplified) | <input type="checkbox"/> French   |
| <input type="checkbox"/> German                | <input type="checkbox"/> Italian              | <input type="checkbox"/> Japanese |
| <input type="checkbox"/> Portuguese (Portugal) | <input type="checkbox"/> Russian              | <input type="checkbox"/> Spanish  |

Type language name to pick more languages

▼ Tools

## Pick a translation memory (optional)

Specify the [translation memory](#) (TM) where you want to store your translated segments. Translations from this TM are preferred over alternate translations from other TMs.

By default, translated segments are stored in your "Personal TM" which is searchable only by you and users with whom you explicitly share access. If you would like to make the translated segments publicly searchable and shared with everyone, select "Global, shared TM."

Osobna prijevodna memorija (mannydrugimail@gmail.com) ▼

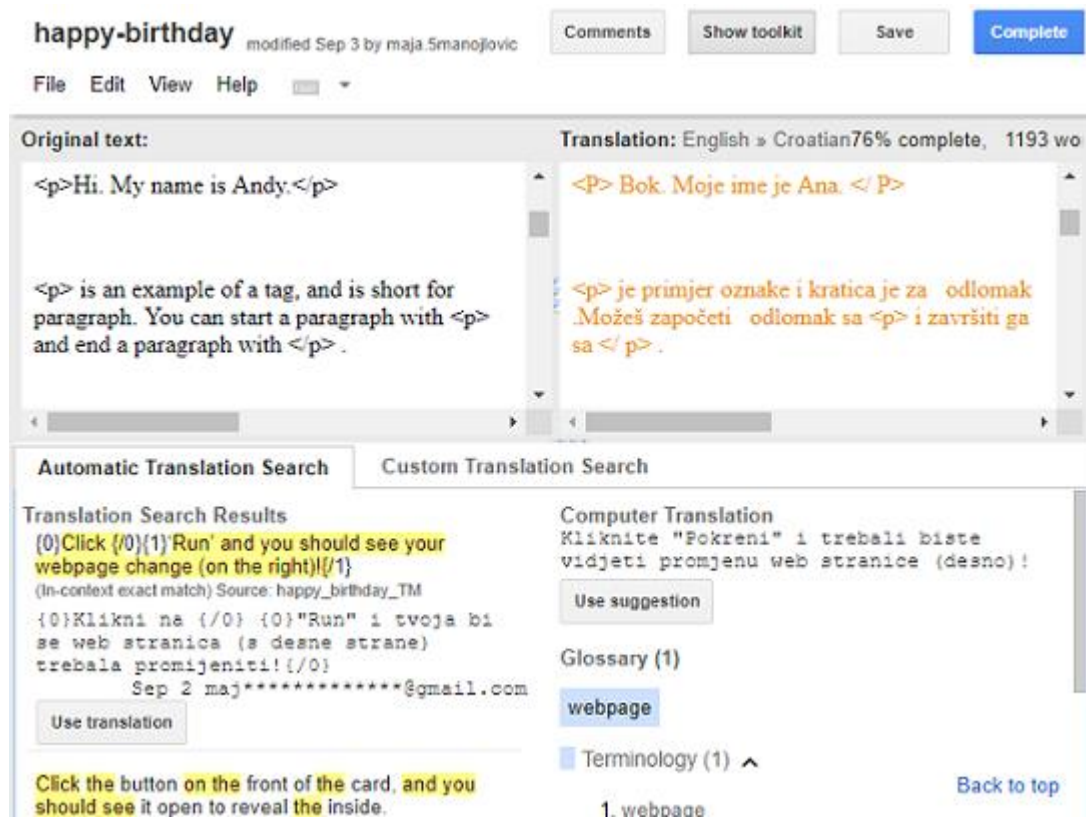
## Pick a glossary (optional)

Specify the [glossary](#) that you want translators to use as the primary reference for terminology specific to this translation.

-- None -- ▼

**Slika 1. Početna stranica Toolkita**

Ekran je, kao i kod većine prevoditeljskih alata, podijeljen na dva dijela – s lijeve strane nalazi se tekst na izvornom jeziku, dok je s desne strane prostor za prevođenje na ciljni jezik (slika 2). Tekst je podijeljen na segmente, a ovisno o postavkama korisnika u desnom se prozoru prikazuju prijedlozi prijevoda generirani Google Translate-om ili pak izvorni tekst.



Slika 2. Ekran alata

Organizacija suradnje s drugim korisnicima također ne predstavlja problem. Jednostavnim pozivom na projekt radu se mogu priključiti i drugi korisnici te simultano prevoditi ili surađivati po principu prevođenja i revizije (Garcia & Stevenson, 2009.). Primarni jezik korisničkog sučelja ovog alata bio je engleski, no od 2017. godine korisničko sučelje dostupno je na 37 jezika, među kojima je i hrvatski (Google Translator Toolkit, 2018.).

Iako jedan od jednostavnije koncipiranih alata, Google Translator Toolkit pruža razne mogućnosti svojim korisnicima. Glavna značajka ovog sustava jesu prevođenje i lokalizacija različitih vrsta sadržaja i dokumenata. Nadalje, korisnici se u svojim prijevodima mogu koristiti vlastitim, ali i dijeljenim, prijevodnim memorijama te rječnicima. Generirani prijevodi mogu se uređivati, mijenjati i usavršavati uz pomoć mnoštva alata za poboljšanje prijevoda. Još jedna od korisnih značajki jest i mogućnost dijeljenja prijevoda i tekstova te suradnje na istima. Na taj način se korisnici mogu povezati i pozvati jedni druge na rad na određenom projektu, ali i tražiti pomoć od profesionalnog prevoditelja ako naiđu na prepreke u radu. Pri završetku rada na projektu, dovršeni se dokumenti mogu pohraniti i spremili direktno u sustavu, ali i izvršiti izvoz, odnosno preuzeti na računalo, u originalnom formatu.



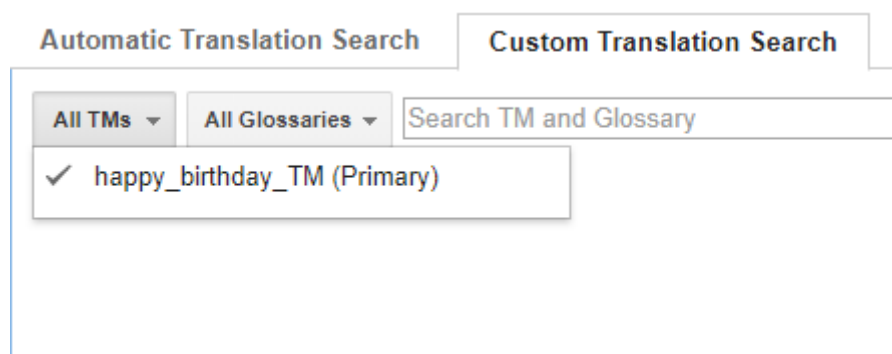
Na samim počecima, točnije 2009. godine, ovaj je sustav podržavao samo jedan izvorni jezik, engleski, te 47 ciljnih jezika. Danas se taj broj povećao na preko 100 000 jezičnih parova koji omogućavaju prevođenje sa 345 izvornih jezika na 345 ciljnih jezika. Ono što korisnik mora imati na umu je da dokument koji se prevodi može sadržavati samo jedan ciljni jezik. Ukoliko sadržava više jezika, onda je potrebno segmente na drugim jezicima odvojiti u zasebne dokumente. Na taj način će segmenti pisani istim jezikom biti grupirani u odvojenim dokumentima.

Toolkitu je jednostavno pristupiti te se na njemu može raditi preko većine poznatih web preglednika. Tako Toolkit podržava *Google Chrome*, *Firefox*, *Internet Explorer* i *Safari*. Korisnicima je također omogućen rad sa različitim formatima dokumenata no potrebno je pripaziti da je maksimalna veličina svakog dokumenta 1 MB te da je godišnje moguće prenijeti 1 GB dokumenata. Korisnici mogu međusobno dijeliti dokumente te na njih postavljati ograničenja. Uloge koje korisnici mogu imati na određenom dokumentu uključuju vlasnika, koji može dijeliti i obrisati dokument i urednika, kojem je dopušteno samo uređivati sadržaj dokumenta. Također, korisniku se može dodijeliti samo mogućnost komentiranja ili čitanja. Sljedeći formati podržani su u sustavu Google Translator Toolkita:

- *Microsoft Word* (.doc/.docx)
- obični tekstualni dokumenti (.txt)
- *Rich Text Format* dokumenti (.rtf)
- hipertekst (.html)
- *Android Resource* (.xml)
- *Application Resource Bundle* (.arb)
- *Chrome* ekstenzije (.json)
- *Java* aplikacije (.properties)
- *Apple iOS* aplikacije (.strings)
- *AdWords Editor Archive* (.aea), gdje je maksimalna veličina datoteke 25 KB
- *AdWords Editor Share* (.aes)
- *SubRip* - titlovi (.srt)
- *SubViewer* (.sub)

Toolkit nudi dvije vrste prijevodnih memorija koje korisnik sam odabire: osobnu ili privatnu i globalnu ili dijeljenu. Ukoliko se korisnik odluči koristiti privatnu prijevodnu memoriju na projektu na kojem radi, onda će se u nju spremati i pohranjivati samo

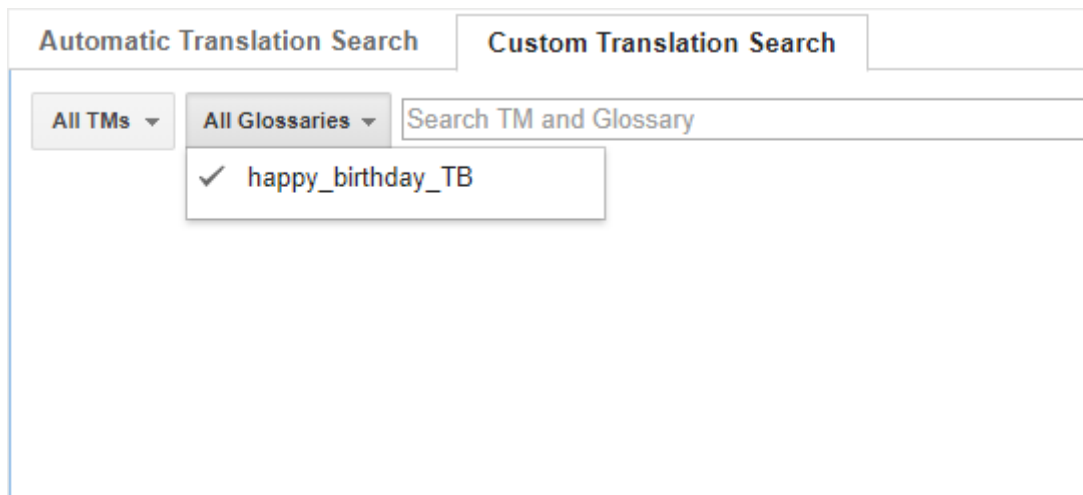
korisnikovi prijevodi. Nadalje, takva memorija može se koristiti za osobnu upotrebu, ali i podijeliti s drugim korisnicima. S druge strane, globalna memorija pohranjuje prijevode od svih korisnika Toolkita u svijetu i dostupna je svima u bilo kojem trenutku. Neovisno za koju prijevodnu memoriju se korisnik odluči, podržani format uvijek je *Translation Memory Exchange* (TMX), a maksimalna veličina 50 MB. Mogućnost odabira prijevodne memorije nudi se pri prevođenju svakog dokumenta što znači da korisnik na jednom dokumentu može koristiti osobnu prijevodnu memoriju, dok će se na drugom odlučiti za globalnu. Također, moguće je koristiti više prijevodnih memorija u isto vrijeme, na istom projektu. U izrađenu prijevodnu memoriju može se dodati već postojeća, a promjene se mogu vršiti i u toku prevođenja što korisnicima daje više fleksibilnosti i prostora za izmjene. Tijekom prevođenja korisnikovi se prijevodi automatski spremaju u odabranu prijevodnu memoriju i dostupni su za buduće korištenje. Princip po kojem prijevodna memorija radi je automatsko generiranje prijevoda na temelju spremljenih, prijašnjih, konstrukcija rečenica koje korisnik zatim mijenja i uređuje, ovisno o točnosti predloženog prijevoda. Na taj način je sam proces prevođenja uvelike olakšan jer korisniku štedi vrijeme nuđenjem prijedloga. U slučaju da se prijevodna memorija dijeli s drugim korisnicima, njima se dodjeljuje uloga vlasnika ili sudionika. Vlasnik može obrisati memoriju i manipulirati njome, dok sudionik samo sudjeluje u njenom stvaranju i korištenju. Također, ako više korisnika dijeli jedan dokument, prijevodna memorija dostupna je svakom od njih. Na slici 3 prikazana je prijevodna memorija dijeljenog dokumenta.



**Slika 3. Prikaz prijevodne memorije u prilagođenoj pretrazi prijevoda**

Uz prijevodne memorije, korisnici mogu kreirati i vlastite rječnike. U rječniku se nalazi popis riječi i fraza te njihovi prijevodi i definicije. Podržani format rječnika je CSU (UTF-8), a maksimalna veličina 1 MB. Razlog zbog kojeg se koristi UTF-8 je zato što

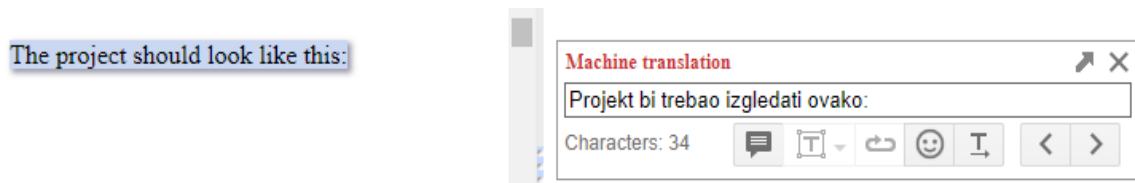
podržava velik broj jezika i posebnih znakova. Rječnik je moguće napraviti koristeći bilo kakve tablice, primjerice *Google Sheets*, no potrebno je poštivati određena pravila za izradu. Obavezno je navesti kôd za jezik te stupce s pojmovima. Kao što je slučaj sa prijevodnom memorijom, tako je i u postojeći rječnik moguće dodati novi te vršiti promjene u rječniku tijekom samog procesa prevođenja. Isto tako, rječnik se može dijeliti s drugim korisnicima pri čemu se dodjeljuju uloge vlasnika ili sudionika, a oni korisnici koji dijele isti dokument automatski dobivaju pristup njegovom rječniku. Doduše, rječnik se neće prikazivati u popisu rječnika, ali će biti integriran u projekt na kojem se radi te će korisnik moći pretraživati pojmove iz rječnika pomoću prilagođene pretrage prijevoda (slika 4).



**Slika 4. Prikaz rječnika u prilagođenoj pretrazi prijevoda**

Automatsko prevođenje koje vrši Toolkit sastoji se od tri, po potrebi četiri, jednostavna koraka. Najprije sustav dijeli odabrani dokument na segmente. Segment je obično jedna rečenica, zaglavlje ili natuknica. Segmente na kojima se radi moguće je spojiti ili razdvojiti, ali samo ako se nalaze u istom odlomku. Segment na kojem korisnik trenutno radi u izvornom jeziku ima plavu pozadinu. Segmenti u ciljnom jeziku mogu biti crvene, narančaste, ljubičaste, smeđe, plave ili zelene boje. Crvenom bojom označeni su strojno prevedeni segmenti (slika 5), a narančastom segmenti koje je korisnik izmijenio (slika 6). Ljubičasta boja označava netočno postavljenu oznaku oblikovanja, smeđa *high fuzzy* podudaranje sa prijevodnom memorijom, plava potpuno (stopostotno) podudaranje s prijevodnom memorijom, a zelena prijevod unutar konteksta, odnosno automatski preveden segment na temelju stopostotne podudarnosti sa prijevodnom memorijom i konteksta (A Step-

by-Step Guide on How to Use the Google Translator Toolkit, 2016.). Kada se na segment klikne, otvara se mali prozor u kojem se mogu vršiti izmjene, dodavati komentari, kopirati oznake iz izvornog jezika, kopirati izvorni segmenti i dodavati emotikoni. Na vrhu prozora navedeno je je li segment strojno preveden ili je uređivan, dok se na dnu nalazi broj znakova segmenta. Ukoliko se u izvornom jeziku nalazi termin iz rječnika, onda će on biti unutar pravokutnika isprekidanih linija.

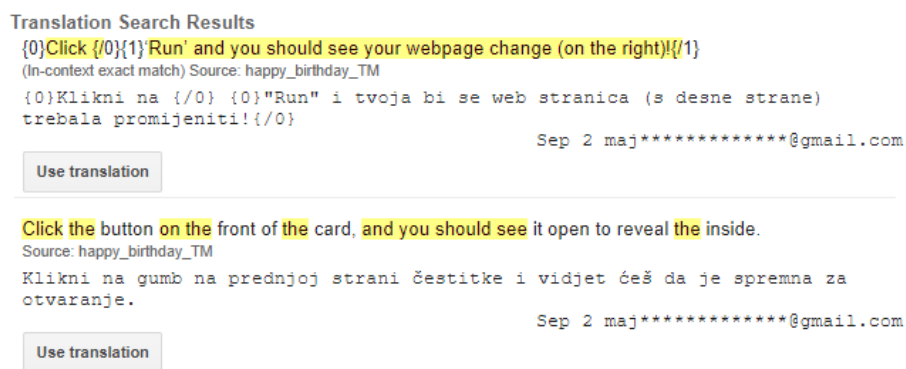


Slika 5. Strojno preveden segment



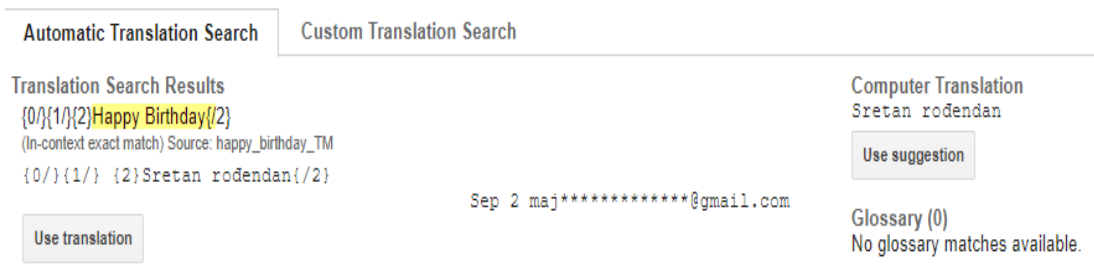
Slika 6. Izmijenjen segment

Nakon toga se za svaki segment pretražuju podudarnosti sa odabranom prijevodnom memorijom. Ako već postoje isti ili slični segmenti pohranjeni u prijevodnoj memoriji, Toolkit odabire onaj sa najvećim postotkom sličnosti i koristi ga kao prijedlog za prijevod segmenta. Izgled rezultata pretrage prijevodne memorije prikazani su na slici 7.



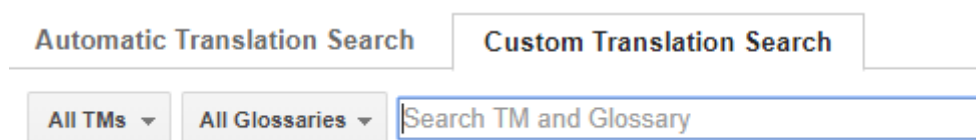
Slika 7. Rezultati pretrage prijevodne memorije

Ukoliko sustav nije pronašao podudarnost u prijevodnoj memoriji, onda se kao prijedlog automatski umeće strojni prijevod. Ako korisnik ne želi da mu se nudi strojni prijevod, u postavkama može podesiti da se umjesto njega prikazuje izvorni tekst. Korisničko sučelje sustava vrlo je jednostavno. Toolkit se nalazi na dnu stranice dokumenta koji se prevodi i podijeljen je na dvije kartice (slika 8).



**Slika 8. Izgled Toolkita**

Prva kartica je automatska pretraga prijevoda koja sadrži: rezultate pretraživanja, rječnik i strojni prijevod. U rezultatima pretraživanja prikazuju se podudarnosti iz prijevodne memorije i označava da li je prijevod identičan ili sličan. Rječnik prikazuje postojeće podudarnosti iz rječnika, ako je pridružen dokumentu, a strojni prijevod prikazuje prijevode segmenata generirane Google Translate-om. Druga kartica nosi naziv prilagođena pretraga prijevoda koja omogućava korisniku pretraživanje pojmova u vlastitoj prijevodnoj memoriji ili rječniku. Ako je projektu pridruženo više prijevodnih memorija i rječnika, korisnik u padajućoj listi može odabrati prijevodne memorije i rječnike nad kojima želi izvršiti pretragu (slika 9).



**Slika 9. Prilagođena pretraga prijevoda**

Mnogi dokumenti sadrže HTML oznake u izvornom tekstu. Kako bi se osigurala dosljednost izgledu originalnog dokumenta i jednostavnije korištenje, Toolkit takve oznake automatski zamjenjuje takozvanim *placeholderima*<sup>6</sup>, odnosno oznakama oblikovanja. Oni se numeriraju ovisno o njihovoj poziciji unutar segmenta te ih je moguće pomicati i dodavati nove. Kako bi korisnik bio upućen u kvalitetu prijevoda, GT alat nudi i prikaz prijevodne statistike za pojedini dokument. Statistika nudi prikaz prijevoda unutar konteksta, odnosno broja riječi koje su automatski prevedene iz prijevodne memorije koristeći stopostotno podudarne riječi i kontekst. Kao kontekst uzima se u obzir segment koji je prethodio segmentu koji se prevodi i segment koji slijedi nakon njega. Također je vidljiv i prikaz stopostotne podudarnosti iz prijevodne memorije, točnije riječi koje su automatski prevedene koristeći savršene podudarnosti iz prijevodne memorije, ali ne uzimajući u obzir kontekst. Osim toga, prikazan je i broj riječi koje su automatski prevedene koristeći skoro savršene podudarnosti iz prijevodne memorije, takozvane *high-fuzzy* podudarnosti. U statistiku je još uključen i broj riječi koje su se strojno prevele i broj riječi koje nije bilo moguće prevesti ni strojno ni uz pomoć prijevodne memorije.

Osim spomenutih značajki, Toolkit ima i nekoliko dodatnih opcija koje također mogu olakšati korisniku rad i korištenje sustava. Ako korisnik ima Standard 101/102-Key ili Natural Ps/2 tipkovnicu, na raspolaganju su mu razni prečaci kojima može olakšati navigiranje sustavom. Uz to, u bilo kojem trenutku vidljiv je prikaz dovršenosti projekta na kojem, ili kojima, korisnik radi. Konačno, u sustavu je omogućena transliteracija<sup>7</sup> što značajno utječe na prijevode u kojima se koriste arapski i indijski znakovi, primjerice Hindi, Punjabi i Bengali, odnosno olakšava njihovo prevođenje (Google Inc., 2018.).

## 4.2. SMARTCAT

Smartcat je *online* platforma koja omogućava povezivanje i suradnju prevoditelja i potencijalnih poslodavaca te omogućava prevoditeljima uređivanje prijevoda. Namijenjena je svima – prevoditeljskim tvrtkama i organizacijama, samostalnim prevoditeljima, urednicima i lektorima, odnosno *freelancerima*<sup>8</sup>, kao i raznim odjelima za lokalizaciju jezika unutar neke

---

<sup>6</sup> *Placeholderi* su varijable koje mogu zamijeniti određene dijelove teksta, npr. brojeve, imena, datume, slike i slično.

<sup>7</sup> Transliteracija je pretvaranje arapskih i indijskih znakova u latinično pismo.

<sup>8</sup> U slobodnom prijevodu, termin *freelancer* označava osobu koja izvršava određene poslove za neku fizičku osobu, poduzeće ili instituciju, a pri tome nije u radnom odnosu s tim poslodavcem (Freelanceri? Tko su freelanceri?, 2012.).

tvrtke. Zamišljen je od strane ABBYY kompanije 2012. godine kao *CAT* alat za korištenje unutar prevoditeljskih tvrtki. Godine 2016. postao je neovisna tvrtka i razvio se u platformu znanu danas. Osim provizije koja iznosi 5% od uplaćenog iznosa i koja se uplaćuje Smartcatu pri završetku rada na svakom projektu, usluga ne zahtjeva druge uplate, točnije besplatna je. Tri su vrste korisničkog računa u Smartcatu: *freelanceri*, prevoditeljske agencije i globalne tvrtke. U tablici 1 prikazane su karakteristike svakog računa (Smartcat, 2018.).

**Tablica 1. Obilježja korisničkih računa u Smartcatu**

	<b><i>FREELANCER</i></b>	<b>PREVODITELJSKA AGENCIJA</b>	<b>GLOBALNA TVRTKA</b>
Podrška 24/7	Da	Da	Da
Prevođenje vlastitih projekata/dokumenata	Da	Da	Da
Prijevodne memorije, rječnici, provjera kvalitete	Da	Da	Da
Strojno prevođenje	Da	Da	Da
Statističko strojno prevođenje	Da	Da	Da
Prikaz na tržištu <i>freelancera</i>	Da	Ne	Ne
Prikaz na LSP tržištu	Ne	Da (na zahtjev)	Ne
Različiti stadiji projekta	Ne	Da	Da
Više korisnika na projektu (prevoditelji, urednici, upravitelji, itd.)	Ne	Da	Da

Zapošljavanje <i>freelancera s tržišta</i>	Ne	Da (provizija)	Da (provizija)
Zapošljavanje drugih tvrtki s LSP tržišta	Ne	Na zahtjev (provizija)	Da (provizija)
Smartcat API	Ne	Da	Da

Brojne su mogućnosti dostupne korisnicima Smartcata. Ono za što se platforma najčešće koristi jest upravo prevođenje. Smartcat nudi besplatan *CAT* uređivač teksta, korištenje prijevodnih memorija i rječnika, provjeru kvalitete prijevoda, prikaz statistike i slično. U suštini, sve što je prevoditelju potrebno da kvalitetno obradi tekst.

Korisnici mogu pretraživati ponudu poslova te pratiti oglase, ali se i predbilježiti za dobivanje obavijesti o novim projektima te na taj način biti u toku s najnovijim ponudama. S druge strane, poslodavci koji su u potrazi za prevoditeljima, urednicima i lektorima mogu vršiti pretragu korisnika u sustavu i odabrati osobu ili osobe savršene za rad na određenom projektu ili zadatku, na temelju informacija o školovanju, certifikatima, području specijalizacije i prijašnjim poslovima. Filtriranje nudi opcije pretrage prema izvornom i ciljnom jeziku, tipu usluge (prevoditelj, urednik, lektor), materinskom jeziku, specijalizaciji, tarifi po riječi i slično, što pomaže osigurati kvalitetu odrađenog posla. Oni korisnici koji su preveli velik broj riječi u Smartcatu češće će se pojavljivati u rezultatima pretrage. Vidljivost također ovisi i o broju položenih testova te broju ocjena dobivenih od poslodavaca.

Sa više od 170 000 lingvista koji tečno govore različite jezike i koji su specijalizirani u različitim poljima, Smartcat je idealno mjesto za pronalazak kvalificiranih osoba za rad. Poslodavci mogu besplatno testirati prevoditelje i na temelju rezultata odlučiti da li su adekvatni za predviđeni projekt. Slanjem zahtjeva direktno Smartcatu, sa detaljima o broju traženih prevoditelja, cijeni, temi teksta te primjerom teksta, sustav pronalazi kandidate za posao te ih poziva da ispune test. Na temelju rezultata testa odabiru se najbolji kandidati i o tome obavještava poslodavca, koji zatim odlučuje kojeg kandidata želi zaposliti za rad na svom projektu. Kandidate se ne plaća za ispunjavanje testa, no dobiju značku koja označava da su položili test, što im može biti korisno u budućim poslovnim suradnjama.

Ako je zadatak opsežniji te na njemu radi više lingvista, poslodavac vrlo lako može nadgledati taj tim, točnije može dodavati nove članove i postavljati njihove tarife za



automatski izračun cijene, kao i popunjavati njihove profile sa dodatnim informacijama kako bi ih u budućnosti lakše pronašli i angažirali za rad na drugim projektima. Isto tako može ograničiti korisnička prava ovisno o ulozi koju korisnik ima na projektu. Lingvist može vidjeti i raditi samo na onim projektima koji su mu dodijeljeni, ne može mijenjati resurse projekta i svi termini koje unese u rječnik moraju biti odobreni od strane nadležnih korisnika. Upravitelj resursima, osim što ima ista prava kao i lingvist, može upravljati prijevodnim memorijama i rječnicima, odnosno dodavati i mijenjati termine te uvoziti i izvoziti dokumente. Upravitelju projekta dopušteno je sve što i lingvistu i upravitelju resursima, a uz to mu je omogućeno kreiranje, brisanje i izmjenjivanje svih projekata. Administrator ima najviše ovlasti te, uz sve već spomenute aktivnosti, može dodavati i uklanjati članove tima, kreirati grupe i upravljati isplatama. Osim ovih uloga koje sustav sam nudi, poslodavac može kreirati i vlastite uloge, nadjenuti im naziv te iz ponuđene liste odabrati ograničenja, odnosno mogućnosti, korisnika i razinu tih prava. Primjerice, ako je korisniku dodijeljeno pravo upravljanja rječnicima, moguće je odrediti točno kojim rječnicima, kojim jezikom u rječniku i slično.

Nadalje, olakšana je raspodjela zadataka, upravljanje projektima i komunikacija sa članovima, koja se vrši preko *chata* ili ostavljanjem komentara unutar dokumenta na kojem se radi. Osim što je vidljiv napredak cjelokupnog tima i projekta na kojem se radi, prikazan je i napredak individualnih članova tima. Projektu se može postaviti određeni vremenski rok, a ako se zadatak sastoji od nekoliko faza, primjerice prevođenja, uređivanja i lektoriranja, rok se može postaviti za svaki korak posebno.

Osim korisnika koji su već u timu, poslodavac može pronaći nove korisnike na platformi te ih pozvati u svoj tim. To može učiniti na dva načina: manualno ili automatski. Ukoliko se poziv vrši manualno, onda se korisnici koji su prihvatili poziv na projekt stavljaju na listu čekanja sve dok ih poslodavac ne odobri, odnosno odabere za rad. S druge strane, automatskim pozivom prednost ima onaj korisnik koji prvi prihvati poziv. Tada se njemu dodjeljuje zadatak i on se automatski pridružuje radu na projektu. Automatsko pozivanje veoma je korisno kada se ne radi o kompliciranim projektima, već o manjim dokumentima, te kada je poslodavac vremenski ograničen i ne može čekati individualne odgovore korisnika na pozivnice. Isto tako, ova je opcija savršena ako se prevoditelj i poslodavac nalaze u različitim vremenskim zonama ili kada poslodavac nije postavio određene uvjete koje prevoditelj mora ispuniti.

S obzirom da na projektima često radi više od jedne osobe, Smartcat je omogućio dijeljenje projekta na segmente. Poslodavac određuje na kojim će segmentima koji član tima

raditi te na taj način ograničava njihov pristup samo na određene segmente. To znači da svaki prevoditelj može uređivati samo one segmente koji su mu dodijeljeni. U suprotnom, svi članovi tima mogu uređivati bilo koji segment, čak i kada je on potvrđen, odnosno odobren, od strane prevoditelja.

Pri završetku posla, nijedna strana ne mora se brinuti o papirologiji i isplatama. U sustavu se sve automatizira na temelju unesenih podataka o tarifama i osobnim podacima poslodavaca i osoba koje izvršavaju zadatke. Poslodavci mogu sigurno i brzo isplatiti prevoditelje, a prevoditelji mogu primiti uplatu željenim putem i imati brz pristup povijesti uplata i stanju svog računa. Proces je organiziran i štedi vrijeme te se sve odvija po šabloni. Primjerice, tvrtke mogu platiti svim prevoditeljima istovremeno sa jednom transakcijom, a *freelanceri* mogu pratiti svoje platno stanje i primiti uplate na bankovni račun, preko Pay Pal-a ili *e-wallets* aplikacije, a mogu i kreirati i slati račune direktno sa Smartcata. Također je automatiziran i prijenos svih potrebnih dokumenata za rad između poslodavca i prevoditelja.

Smartcat omogućava rad na 255 različitih jezika od kojih svaki posjeduje odgovarajući kôd koji je moguće koristiti za API zahtjeve, uvoz TMX dokumenata, stvaranje prijevodnih memorija u Excelu i uvoz i izvoz XLIFF dokumenata. Formati dokumenata koje sustav podržava su sljedeći:

- *Microsoft Office* (.doc/.docx, .xls/.xlsx, .ppt/.pptx, .pps/.ppsx, .pot/.potx)
- *Open Office* (.odt, .odp)
- obični tekstualni dokumenti (.txt)
- *Rich Text Format* dokumenti (.rtf)
- hipertekst (.html, .xhtml, .php)
- dvojezični formati dokumenata (.xlf/.xliff/.sdlxiff/.mqxliff/.sdlxiff, .po, .ttx)
- dokumenti za stolno izdavaštvo (.pdf, .mif, .idml)
- dokumenti za tehničko pisanje (dita .xml, help + manual .xml)
- dokumenti za lokalizaciju (.csv, .tsv, .xml, Android .xml, .resx, .dtd, .json, .tjson, .yaml, .imc, .inx, .mif, .strings, .properties)
- *SubRip* – titlovi (.srt)
- slike (.jpg/.jpeg, .tif/.tiff, .bmp, .png, .gif, .djvu/.djv, .dcx, .pcx, .jp2, .jpc, .jfif, .jb2)
- paketi (.ttx, .sdlppx/.sdlrpx, .zip, .wsxz)

Kao i u drugim CAT alatima, i u Smartcatu se korisnik može koristiti prijevodnom memorijom i rječnicima. U prijevodnoj memoriji pohranjuju se prijevodi tekstova s jednog

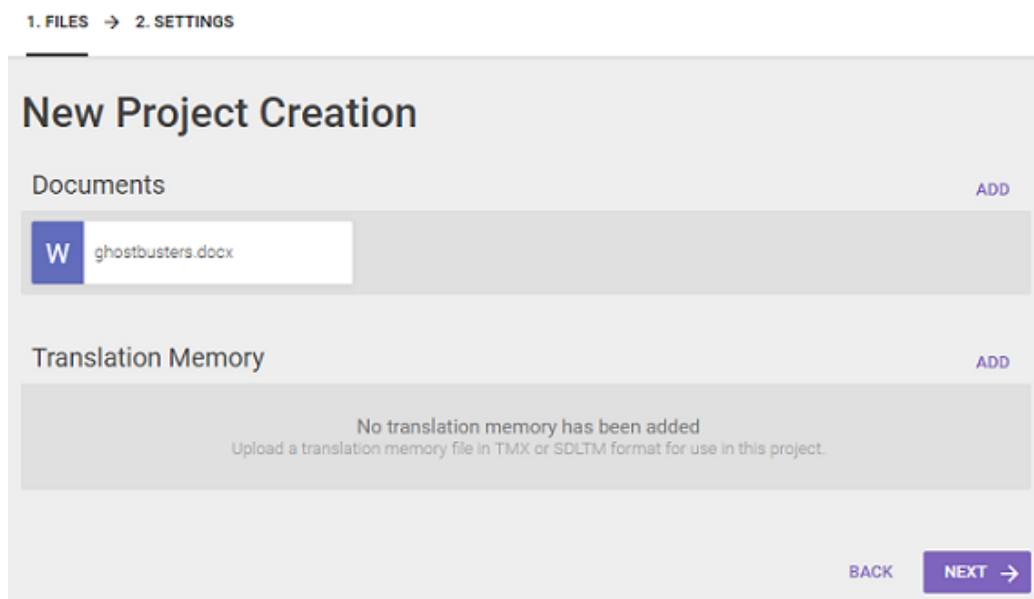
izvornog jezika na jedan ili više različitih ciljnih jezika. Kada korisnik izradi projekt, za njega se automatski stvara prijevodna memorija, osim ako već postojeća nije odabrana za taj projekt. Također, ako se precizira klijent za odabranu prijevodnu memoriju, sustav automatski pruža i relevantne informacije iz prijašnjih prijevodnih memorija korištenih u radu s tim klijentom. Točnije, novi projekt povezuje se sa prijevodnim memorijama starih projekata te je na taj način mnogo više potpunih i sličnih podudaranja dostupno na korištenje. U toku prevođenja, segmenti se spremaju u prijevodnu memoriju te koriste u daljnjem i budućem radu. Sustav uspoređuje segment koji se trenutno prevodi sa već prevedenim segmentima koji su spremljeni u prijevodnoj memoriji i nudi prijedloge ako je podudarnost 75% ili više. Ovaj postotak automatski je postavljen, ali ga korisnik može podesiti po svojoj želji i može biti različit za svaku prijevodnu memoriju. Smartcat dopušta manipulaciju prijevodnim memorijama pa tako korisnik može za nove projekte koristiti stare prijevodne memorije, mijenjati naziv, jezike, odabrati klijenta, temu ili projektnu grupu, izvoziti i uvoziti prijevodnu memoriju, ali ju i isključiti. Također se može vršiti provjera konteksta na način da se u obzir uzimaju prethodni segment i segment koji slijedi nakon segmenta koji se prevodi što uvjetuje točnije prijevode i veću podudarnost s prijedlozima.

Kao i kod prijevodnih memorija, Smartcat korisniku omogućava izradu novog rječnika i preciziranje jezika, klijenta i naziva, korištenje već postojećeg, prenošenje rječnika i isključivanje rječnika. Ako korisnik navede klijenta, sustav povezuje informacije iz prijašnjih rječnika korištenih u radu s tim klijentom sa novim rječnikom.

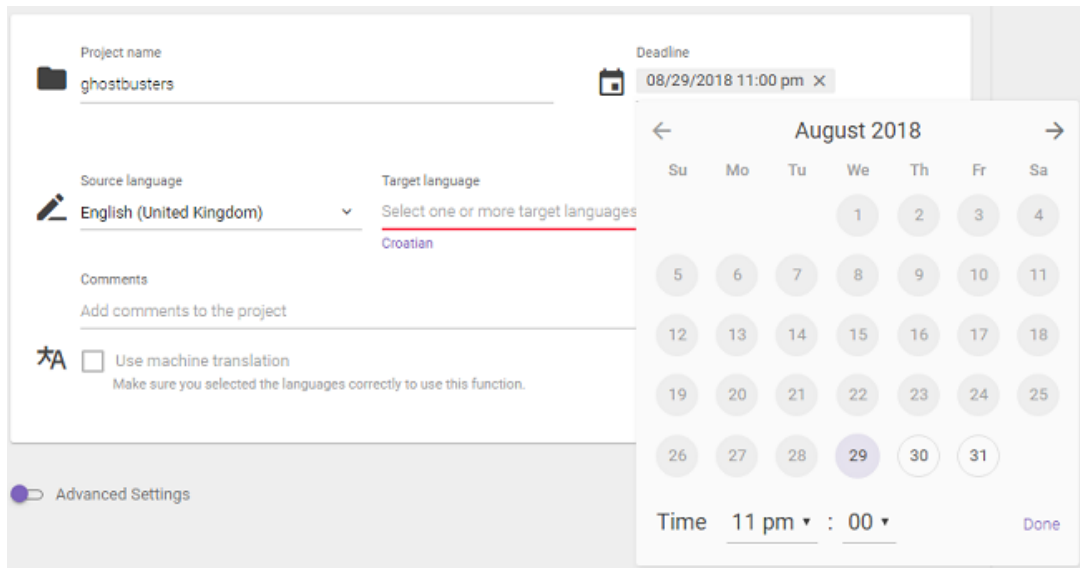
Uz prijevodne memorije i rječnike, Smartcat se odlikuje još jednim veoma korisnim svojstvom, a to je mogućnost kontrole kvalitete prijevoda. Kvaliteta prijevoda osigurava konzistentnost originalnog teksta i prijevoda. Primjerice, ako se u izvornom tekstu na kraju segmenta nalazi točka, a u prijevodu zarez, Smartcat će o toj pogrešci obavijestiti korisnika prikazujući mu ikonu uskličnika u narančastom trokutu pokraj segmenta u kojem se pogreška javila. Kada korisnik prođe kursorom preko ikone ili otvori karticu *QA Check* vidjet će popis i objašnjenja grešaka. Dakle, kontrola kvalitete vrši automatsku provjeru pravopisa, interpunkcije, terminologije, formatiranja i konzistentnosti sa prijevodnom memorijom. Popis svih potencijalnih grešaka u projektu moguće je preuzeti i na računalo u obliku izvješća gdje su greške grupirane prema tipu: pravopis, oznake (koje postoje da bi se zadržalo originalno oblikovanje), terminologija, prijevodna memorija, datumi i brojevi, interpunkcija i ostalo. Pri prenošenju dokumenta za prijevod, moguće je postaviti uvjete za provjeru kvalitete vezane uz terminologiju, prijevodnu memoriju i datume i brojeve. Primjerice, korisnik može odabrati da

se prikazuju upozorenja kada se prijevod ne podudara s prijevodom u rječniku, kada je pronađeno više stopostotnih podudaranja za jedan segment ili kada se neki broj ne podudara s originalnim tekstom. Isto tako, korisnik može naznačiti da se neka upozorenja ignoriraju ako procjeni da nisu relevantna, ali i naglasiti da su neka upozorenja ključna. U posljednjem slučaju sustav neće dopustiti da se segment označi kao preveden, odnosno potvrdi, sve dok se pogreška ne ispravi. Ukoliko segment nije potvrđen, projekt se ne može označiti kao dovršen. Važno je naglasiti i da sustav prepoznaje morfološke oblike riječi, što znači da se različite verzije iste riječi koja se nalazi u rječniku neće registrirati kao pogreške (npr. drugačije glagolsko vrijeme, padež ili broj).

Rad u Smartcatu započinje se u nekoliko jednostavnih koraka. Na početku korisnik u sustav prenosi dokument i prijevodnu memoriju u TMX formatu (ako postoji), kao što je prikazano na slici 10, a zatim odabere izvorni i ciljni jezik i, ako želi, rok dovršetka projekta (slika 11).

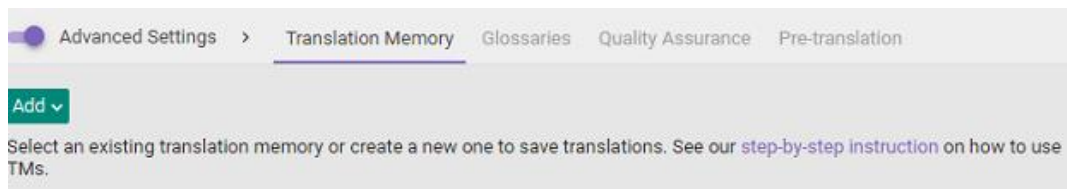


**Slika 10. Kreiranje novog projekta**



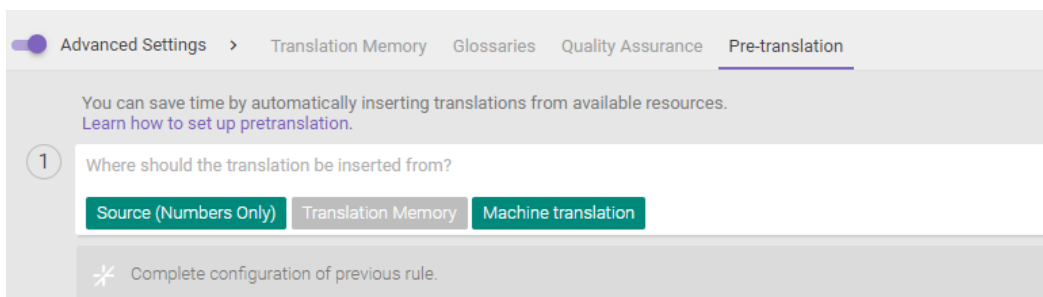
**Slika 11. Pripremanje projekta za rad**

Između ostalog, u naprednim postavkama moguće je postaviti razne funkcije poput postavljanja prikaza prijedloga strojnih prijevoda i podešavanja prijevodne memorije, rječnika i provjere kvalitete prijevoda (slika 12).



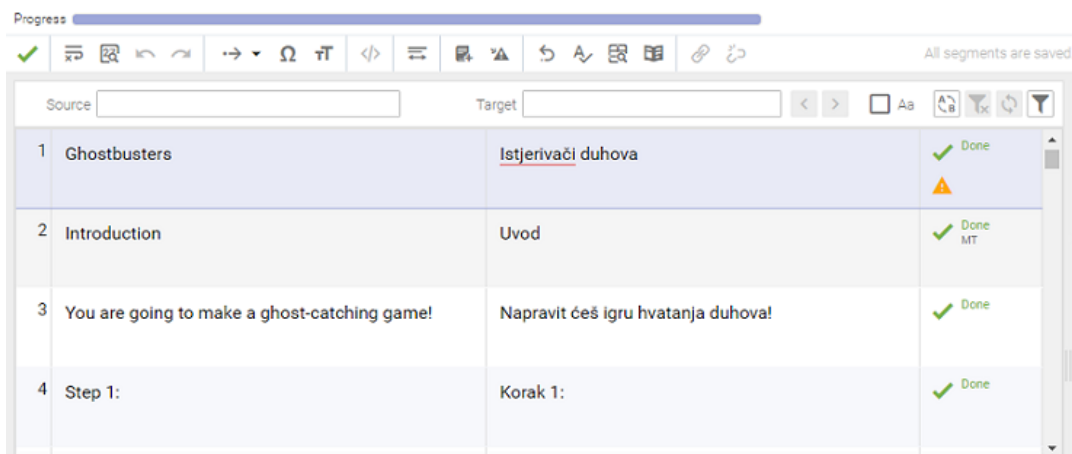
**Slika 12. Napredne postavke**

Također se može uključiti opcija automatskog prevođenja kojom sustav automatski umeće podudarnosti iz prijevodne memorije i strojni prijevod u tekst što uvelike ubrzava cjelokupan proces prevođenja jer prevoditelj ne mora ručno kopirati navedene podudarnosti ako se slaže s njima. Ova opcija omogućava i postavljanje automatskog umetanja brojeva u tekst, umetanja podudarnosti iz određene prijevodne memorije ako je zadovoljen određeni minimalni postotak podudarnosti i minimalna duljina segmenta, ali i postavljanje automatskog unošenja strojnog prijevoda za one segmente za koje nisu pronađene nikakve podudarnosti u prijevodnoj memoriji (slika 13).



**Slika 13. Podešavanje automatskog prevođenja**

U samom uređivaču teksta lako se snalaziti. Ekran je podijeljen na dva dijela – lijevi i desni prozor (slika 14).



**Slika 14. Ekran alata**

U lijevom prozoru nalazi se tekst na izvornom jeziku koji je podijeljen na segmente. U desnom prozoru je prostor za prijevod segmenata na ciljni jezik, koji se može vršiti ručnim upisivanjem ili pak odabirom ponuđenih prijevoda iz *CAT* kartice. *CAT* kartica sadrži sve prijevode odabranog segmenta koji se nalaze u resursima projekta koji su priloženi - podudarnosti iz prijevodne memorije, termini iz rječnika i strojni prijevod. Izgled *CAT* kartice prikazan je na slici 15.

CAT			
1	Ghostbusters	TM 102%	Istjerivači duhova
2	ghostbusters	TM 85%	istjerivači duhova

CAT Info	Concordance Search	Dictionaries
Match difference No differences found		
Translation Memory <b>ghostbusters</b>		
Unit created 08/29/2018 10:40 AM		

**Slika 15. CAT kartica**

Osim klasičnog prenošenja dokumenata sa računala, postoje i dodatne opcije. Korisnik može prenijeti tablice i odabrati koji dio će se prevoditi, podesiti način kreiranja segmenata i slično. Nadalje, mogu se prenijeti višejezični dokumenti, PDF dokumenti, ali i slike. Također se mogu podesiti lokalizacija i izgled prijevoda i koristiti oznake oblikovanja pri prenošenju različitih dokumenata za lokalizaciju ili višejezičnih dokumenata. Konačno, treba napomenuti da Smartcat podržava Trados pakete – prevoditeljski softver za prevoditelje i davatelje jezičnih usluga. Preko Trados paketa korisnici mogu jedni drugima prosljeđivati sve materijale i informacije potrebne za rad, kao što su prijevodne memorije, broj riječi, višejezične datoteke projekta, rokovi, komentari i slično. Još jedna zanimljiva značajka ovog alata je i mogućnost rada bez internetske veze. Sve aktivnosti koje se u tom periodu odrade spremaju se na korisnički račun kada korisnik bude imao pristup internetskoj vezi.

Uvid u status projekata olakšava i svojstvo prikaza statistike. Prikaz je zapravo ukupan broj simbola ili broj riječi na temelju kojeg se prikazuje opseg i količina posla, kao i podudarnosti iz prijevodne memorije, rječnika i ponavljanja u tekstu. Poslodavci mogu izračunati statistiku projekta na temelju originalnih dokumenata i podudarnosti s prijevodnom memorijom. Statistiku je moguće izračunati za svakog člana koji je radio na projektu i ona sadrži informacije o količini posla koja mu je bila dodijeljena i količini posla koju je odradio. Na taj način vidljivi su podaci o trošku projekta, ali i količini odrađenog posla. Statistika prikazuje ukupan broj stranica, segmenata, simbola, broj riječi, broj novih segmenata i

podudarnosti prijevodne memorije, broj ponavljanja u svakom dokumentu i ponavljanja unutar različitih dokumenata istog projekta. Mjerne jedinice na kojima se izračun statistike temelji su riječi, stranice (gdje je jedna stranica jednaka 250 riječi), znakovi s prazninama i znakovi bez praznina. Statistiku je moguće preuzeti u Trados XML formatu (za korištenje u SDL Trados programu) ili Smartcat XLSX formatu (za korištenje u MS Excelu ili bilo kojim drugim tablicama) te se generira poseban dokument za svaku osobu koja je radila na projektu i za jedan jezični par.

Pri izračunavanju potrebne isplate za rad na projektu koristi se automatski izračun u Smartcatu. Poslodavac može prilagoditi neto stopu projekta i pridružiti ju određenom korisniku, a ukoliko to ne napravi, onda se odabirom korisnika za rad na projektu automatski odabire i njegova neto stopa. Na temelju odabrane neto stope za određeni projekt određuju se i riječi koje se naplaćuju. Taj se trošak izračunava automatski na temelju umnoška broja riječi koje se naplaćuju u originalnom dokumentu, a sukladno prevoditeljevoj tarifi, imajući na umu da se podudarnosti sa prijevodnom memorijom i ponavljanja manje naplaćuju. Samo se potpuno dovršeni projekt u konačnici naplaćuje, a punu uplatu dobit će samo onaj prevoditelj koji prvi potvrdi segment. Sustav izračunava broj prevedenih segmenata za svakog prevoditelja i određuje je li segment koji je potvrđen preuzet iz prijevodne memorije što je od velike važnosti kada na projektu radi više prevoditelja. Primjerice, jedan prevoditelj preveo je segment i sada je on pohranjen u prijevodnoj memoriji. Zatim je drugi prevoditelj umetnuo taj prijevod u svoj segment jer je postotak podudarnosti bio velik. U tom slučaju će drugi prevoditelj biti plaćen manje od prvog. Kada dva segmenta ne bi bila jednaka i nikakva prijevodna memorija se ne bi koristila, onda bi oba prevoditelja bila plaćena po punoj cijeni.

Na početku projekta sustav izračunava očekivani trošak koji se kasnije, pri završetku projekta, može razlikovati od konačne cijene. Razlog tome mogu biti smanjene cijene, odnosno popusti, zbog ponavljanja i korištenja sugestija prijevodne memorije. Također, prevoditelji imaju različite tarife zbog čega trošak projekta varira od korisnika do korisnika. Nadalje, na promjene u samom izračunu utječu i pravilo da prvi prevoditelj koji potvrdi segment dobiva punu uplatu, dok su ostali prevoditelji plaćeni manje te promjene u zadacima projekta, kao primjerice dodavanje novih dokumenata ili prijevodnih memorija. Nakon što je projekt dovršen, upravitelj projekta ga prihvaća ili traži reviziju. Ukoliko je projekt prihvaćen, onda se prevoditelju vrši uplata sredstava direktnim prijenosom s računa na račun, preko Pay Pal-a ili bankovne kartice, ovisno o odabranoj metodi plaćanja, a valuta se automatski prilagođava korisniku. Obavijesti o uplatama šalju se i na *e-mail* adresu primatelja. Smartcat



omogućava poslodavcima i dodjeljivanje penala prevoditeljima. To znači da poslodavac može prevoditelju platiti manje od predviđenog ako se određena pravila nisu poštovala, primjerice ako je odrađeno manje od 30% posla i prevoditelj nije u mogućnosti nastaviti rad na projektu ili ako je prevoditelj koristio strojne prijevode, ali ih nije korigirao dovoljno. Prevoditelj uplatu dobiva unutar 5 dana od dovršenja projekta, a moguće je izvršiti uplatu i za projekte koji nisu odrađeni u Smartcatu. Međutim, tada izrada računa i ispunjavanje papirologije nisu automatski. Za projekte odrađene preko platforme, Smartcat uzima proviziju od 5%, dok za projekte za koje se samo vrši uplata preko Smartcata 3 – 8%, ovisno o poslu, zemlji u kojoj se nalazi poslodavac i zemlji u kojoj se nalazi prevoditelj. S druge strane, zabranjeno je plaćanje prevoditelju u gotovini, odnosno zaobilazeći Smartcat. Osim što postoji veći rizik od prevare i mrežne krađe identiteta, korisnicima može biti i uskraćen pristup nekim svojstvima platforme, primjerice mogu se prestati pojavljivati u pretragama poslodavaca ili mogu izgubiti pristup pretraživanju i zapošljavanju.

Treba spomenuti još jednu značajku ovog sustava, a to je Smartcat Mastercard kreditna kartica koju korisnici mogu zatražiti i putem koje kupovati, plaćati račune *online* i *offline*, podizati gotovinu s bankomata te, najvažnije, primati uplate preko Smartcata bez provizije. Kartica se izdaje preko *ePayments* kompanije koja je internacionalni pružatelj usluga plaćanja. Korisnik kartice ne plaća proviziju ako prebacuje novac sa Smartcat računa na *ePayments* račun, prebacuje novac s *ePayments* računa na Smartcat Mastercard karticu ili koristi karticu za plaćanje, bilo *online* ili *offline*. Jednokratna dostava kartice naplaćuje se 5.95 dolara te ju je moguće platiti u dolarima ili eurima bez provizije. Ukoliko se korisnik odluči na plaćanje u nekoj drugoj valuti, onda plaća dodatnih 2.6% od početnog iznosa. Isto tako, za podizanje gotovine s bankomata u dolarima ili eurima, naplaćuje se provizija od 2.6 dolara, a moguća je i dodatna naplata od strane banke preko čijeg se bankomata podignuo novac. Nadalje, ako korisnik podigne gotovinu s bankomata u nekoj drugoj valuti, naplaćuje se provizija od 2.6%.

Uzevši u obzir velik obujam podataka kojima raspolaže, nije iznenađujuće što Smartcat svojim korisnicima garantira sigurno i bezbrižno korištenje bez straha od zlouporabe informacija. Napredne tehnologije koriste se za zaštitu podataka i prevenciju neautoriziranog pristupa korisničkim računima. Informacije se sigurno pohranjuju na oblaku, a podaci o korisnicima i korisnički računi su izolirani, točnije limitirani su samo na korisnika računa i vidljivi samo njemu osim ako se korisnik ne odluči za javan profil. Autorizacijom se upravlja na različitim razinama te se na taj način ograničavaju prava pristupa korisnicima. Podaci su

pohranjeni u visoko osiguranim centrima koji se nalaze na različitim lokacijama, a njihov prijenos između korisnikovog računala i Smartcat servera zaštićen je HTTPS/TLS Protokolom. Osim toga, podaci su šifrirani 256-bit Advanced Encryption Standardom (AES), dok je prijenos podataka o uplatama zaštićen TLS 1.2 protokolom i RSA algoritmom. Uz sve to, Smartcat se brine i o konstantnoj pohrani informacija u svojim bazama podataka (Smartcat Platform Inc., 2018.).

## 5. PREVOĐENJE POMOĆU GOOGLE TRANSLATOR TOOLKITA

Korpus korišten u analizi rada sa Google Translator Toolkitom sastoji se od četiri dokumenta koji su preuzeti sa službene stranice Code Cluba, a pokrivaju cjelinu o HTML i CSS jezicima za izradu i uređivanje web stranica. Korpus se sastoji od dva projekta, „*Sretan rođendan*“ („*Happy birthday*“) i „*Ispričaj priču*“ („*Tell a story*“) te od bilješki spomenutih projekata za voditelje kluba. U projektima se djeca uče služiti HTML i CSS programskim jezicima, dok se u bilješkama detaljnije navode ciljevi projekta, ishodi učenja i korišteni resursi.

Projekti su učitani u Toolkit preko računala, no prije toga su pretvoreni iz .pdf formata u .doc<sup>9</sup> format jer Toolkit ne podržava .pdf formate dokumenata, a projekti nisu dostupni u drugim formatima. Bilješke za voditelje kluba obrađene su u .html formatu. Jedne su bilješke prenesene u sustav preko računala, dok su druge učitane direktno u sustav kopiranjem poveznice. Prvotna ideja bila je prevesti još jedan projekt i njegove bilješke koristeći prijevodne memorije i rječnike prva dva projekta. Međutim, jedini način da se koristi više prijevodnih memorija istovremeno je da se kreira nova memorija, a zatim se sa računala u nju dodaju već postojeće prijevodne memorije. Iz nekog razloga, Toolkit nije dopuštao preuzimanje prijevodnih memorija iz sustava te se od prevođenja ovog projekta odustalo zbog nemogućnosti isprobavanja novih funkcionalnosti i mogućnosti.

S obzirom da su materijali predviđeni za učenje programiranja, u njima se nalaze i dijelovi kôda. Kôd nije pisan direktno u dokumentu, već se nalazi na slikama. Točnije, većina slika koja se nalazi u dokumentima zapravo su snimke zaslona kôda pisanog na *trinket* platformi<sup>10</sup>. Toolkit ne prepoznaje tekst na slikama pa se zato spomenuti dijelovi kôda nisu preveli. Toolkit ih nije procesirao u sustavu te nisu bili ponuđeni u obliku segmenata za prevođenje. Treba napomenuti i kako se pri preuzimanju prevedenog dokumenta slike nisu učitale i nisu bile vidljive iako bi u pravilu trebale biti prikazane.

U tablici 2 je navedena statistika korpusa te su dokumenti označeni rednim brojevima: „*Sretan rođendan*“ (1.), bilješke „*Sretan rođendan*“ (2.), „*Ispričaj priču*“ (3.), bilješke „*Ispričaj priču*“ (4).

---

<sup>9</sup> Stranica za pretvorbu dokumenata: <https://pdf2doc.com/> (dohvaćeno: srpanj 2018.)

<sup>10</sup> Stranica za pisanje kôda u trinketu: <https://trinket.io/> (dohvaćeno: kolovoz 2018.)

**Tablica 2. Statistika korpusa**

TEKST					PROSJEK
	1.	2.	3.	4.	
<b>BROJ RIJEČI</b>	1193	462	960	413	757
<b>BROJ SEGMENTATA</b>	137	74	112	71	98,5
<b>BROJ SLIKA</b>	20	0	16	0	9
<b>VRIJEME PREVOĐENJA</b>	45 min	24 min	27 min	19 min	28 min

Treba naglasiti da se proces prevođenja i analiza dokumenata prilično razlikuje od analize u Smartcat alatu. Razlog tome je prvenstveno činjenica da se i dva alata sami po sebi veoma razlikuju. Odnosno, Toolkit je pojednostavljen i nema ni približno onoliko svojstava, funkcionalnosti i mogućnosti koje pruža Smartcat. Zato nije bilo moguće sagledati alate iz iste perspektive. Ipak, u nekim su aspektima jednaki no detaljnije će o tome biti govora u sljedećem poglavlju.

### **5.1. PROJEKTI „SRETAN ROĐENDAN“ I „ISPRIČAJ PRIČU“**

Projekti „*Sretan rođendan*“ i „*Ispričaj priču*“ analizirani su simultano jer se proces i rezultati njihovog prevođenja minimalno razlikuju. Oba su dokumenta korištena u .doc formatu te se za svaki napravila zasebna prijevodna memorija i rječnik.

Prvi korak ka prevođenju dokumenta jest njegovo učitavanje u sustav. Toolkit omogućava korisniku prijenos dokumenta sa računala, direktno unošenje teksta u polje za unos, upisivanje URL adrese web stranice, prijenos članka s Wikipedije ili prijenos Youtube videa kako bi se preveli natpisi u videu. U ovom su se slučaju dokumenti učitali sa računala te su im dodijeljeni nazivi i odabrani su izvorni i ciljni jezik. Projektu se može pridružiti prijevodna memorija i rječnik, no potrebno ih je izraditi prije učitavanja samog dokumenta. Slika 16 prikazuje postupak kreiranja rječnika, a slika 17 postupak kreiranja prijevodne memorije. U oba projekta korištena je interna, tj. osobna, prijevodna memorija.

## Translator Toolkit

---

### Upload a glossary

You can create and upload glossaries for use with your translations

[« Back to Translator Toolkit](#)

### Select a file:

ha...sv

### What do you want to call it ?

happy\_birthday\_TB

Slika 16. Dodavanje rječnika

## Translator Toolkit

---

### Add a translation memory

You can create and upload translation memories for use with your translations

[« Back to Translator Toolkit](#)

### Select a file: (optional)

Nije oda...atoteka.

### Translation memory name

happy\_birthday\_TM

### Sharing

If a [translation memory](#) is shared with everyone, its translations are searchable by all users. Otherwise, its translations are searchable only by you and users with whom you explicitly share access. Once you create a translation memory, you cannot change this setting. [Learn more about how Google uses your translation data.](#)

Shared with everyone  Not shared with everyone

Slika 17. Kreiranje nove prijevodne memorije

Prijevodnu memoriju moguće je kreirati u nekoliko jednostavnih klikova u sustavu dok je za rječnik potrebno prvo napraviti nekakvu vrstu tablice u .csv formatu. Rječnici oba dokumenta kreirani su koristeći *Google Sheets*, a zatim preuzeti na računalo u odgovarajućem

formatu. Rječnik obavezno mora sadržavati zaglavlje u kojem se nalaze kôdovi za izvorni i ciljni jezik te se zatim termini i njihovi prijevodi upisuju u odgovarajućim stupcima. Stupci za unošenje opisa i vrste riječi opcionalni su. Nakon tih postavki, dokumenti su bili spremni za prevođenje. Isječak iz rječnika projekta „*Sretan rođendan*“ prikazan je na slici 18.

1	en,hr		
2	webpage,web stranica		
3	tag,oznaka		
4	bold,podebljano		
5	paragraph,odlomak		
6	property,svojstvo		

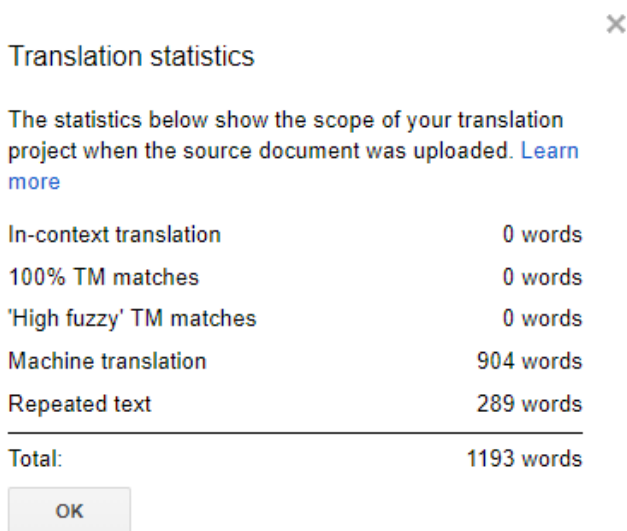
**Slika 18. Isječak iz rječnika „*Sretan rođendan*“**

Prevođenje projekta „*Sretan rođendan*“ trajalo je 45 minuta, što je relativno brzo i čak iznenađujuće jer je pretpostavka bila da će biti potrebno mnogo više korekcija kao rezultat Toolkitove jednostavnosti. Drugi je projekt, „*Ispričaj priču*“ još brže preveden – u samo 27 minuta. Tome je vjerojatno pomogla i činjenica da se sa radom GT alatom upoznalo pri prevođenju prvog dokumenta te je kasnije bilo lakše i brže navigirati alatom i koristiti se njegovim značajkama. Nadalje, GT alat se pokazao iznimno koristan u kontekstu strojnih prijevoda. U svaki segment ciljnog jezika umeće se strojni prijevod generiran Google Translate-om osim kada u prijevodnoj memoriji postoji prijevod za taj segment. Prijevodi su bili više nego korisni te su olakšali sam proces prevođenja. Velik broj segmenata nije trebalo izmjenjivati, već su odmah potvrđeni. Dio segmenata zahtijevao je sitne preinake, no unatoč tome prevođenje se odvijalo brzo. Nekim segmentima dodao se komentar u svrhu isprobavanja te funkcionalnosti koja je naročito korisna kada na projektu radi više osoba. U gornjem desnom kutu alata prikazan je postotak dovršenosti prijevoda, a za svaki segment je prikazano koliko znakova sadrži. Sustav je prikazivao termine iz rječnika, no potrebno ih je ručno upisati.

Pri završetku rada na projektima, dokumenti su preuzeti na računalo u .doc formatu. S obzirom da su dokumenti konvertirani iz .pdf formata u .doc format, javile su se greške u izgledima projekata. Originalni dokument „*Sretan rođendan*“ je, nakon pretvorbe, imao 10 stranica, dok se nakon prevođenja i preuzimanja na računalo taj broj povećao na 46. Isto tako, projekt „*Ispričaj priču*“ je sa 10 stranica skočio na 43 stranice. Uz to, dokumenti na ciljnom

jeziku nisu nalikovali dokumentima na izvornom jeziku. Osim različitog broja stranica, došlo je i do promjena u formatiranju i izgledu. Većina slika uopće se nije prikazala, a one koje su bile vidljive jedva su se mogle razaznati. Nakon što je prvi projekt cijeli preveden, Toolkit je javljao 76% dovršenosti projekta. Isto se dogodilo i sa drugim projektom te je postotak njegove dovršenosti iznosio 74%. Pretpostavka je da se to dogodilo jer se već spomenute slike, odnosno tekst na slikama, nisu procesirale u Toolkitu kao segmenti te nisu ponuđeni za prevođenje. Osim prevedenog dokumenta, moguće je preuzeti i prijevodnu memoriju projekta te ju i podijeliti sa drugim korisnicima. Ipak, pri pokušaju preuzimanja prijevodne memorije projekata pojavila se greška te ih nije bilo moguće preuzeti na računalo.

Statistika projekta „*Sretan rođendan*“, odnosno procjena prijevoda prije prevođenja, navodi kako su u dokumentu strojno prevedene 903 riječi, dok se 289 riječi ponavljalo. Ukupan broj riječi dokumenta iznosi 1193. Projekt „*Ispričaj priču*“ imao je nešto manje riječi, točnije 960. 673 riječi bile su strojno prevedene, dok se 287 riječi ponavljalo. Slike 19 i 20 prikazuju procjenu statistike svakog projekta prije početka prevođenja.



Translation statistics	
The statistics below show the scope of your translation project when the source document was uploaded. <a href="#">Learn more</a>	
In-context translation	0 words
100% TM matches	0 words
'High fuzzy' TM matches	0 words
Machine translation	904 words
Repeated text	289 words
<b>Total:</b>	<b>1193 words</b>

OK

**Slika 19. Statistika projekta „*Sretan rođendan*“**

### Translation statistics ×

The statistics below show the scope of your translation project when the source document was uploaded. [Learn more](#)

In-context translation	0 words
100% TM matches	0 words
'High fuzzy' TM matches	0 words
Machine translation	673 words
Repeated text	287 words
<hr/>	
Total:	960 words

OK

**Slika 20. Statistika projekta „Ispričaj priču“**

Toolkit ne prikazuje statistiku pogrešaka niti pogreške kao takve, točnije sustav ne detektira potencijalne greške poput Smartcata, već korisnik samostalno odlučuje je li segment ispravan ili ne. Iz tog razloga nije se provela nikakva analiza pogrešaka. Ipak, u tablici 3 su prikazane podudarnosti oba projekta sa prijevodnom memorijom, strojnim prijevodom i rječnicima, nakon procesa prevođenja. Treba uočiti kako je strojnih prijevoda koji su bili upotrebljivi i smisleni bilo jako mnogo za oba dokumenta. U statistiku su ubrojani i strojni prijevodi poput „*Pronađite ovaj kôd*“ kao prijedlog za „*Find this code*“ dok sam se ja, zbog konzistentnosti, odlučila prevesti „*Pronađi ovaj kôd*“ i zadržala neformalni stil. Isto tako, Toolkit je iznenadio svojim prijevodima duljih konstrukcija rečenica poput „*Možete li uočiti druge oznake?*“ („*Can you spot any other tags?*“). Očekivano je bilo da će poneke riječi savršeno prevesti, no alat se pokazao izvrstan i u baratanju sa cijelim rečenicama. Primjerice, „*Let's start by editing the HTML content and css style of the story webpage*“ savršeno je preveo kao „*Počnimo s uređivanjem HTML sadržaja i CSS stila web stranice priče*“. Što se tiče količine podudarnosti sa prijevodnim memorijama, rječnicima i prijedlozima strojnog prevoditelja, više ih se javilo pri prevođenju prvog projekta. No, razlike nisu velike. Drugi projekt imao je i nešto manje riječi od prvoga što je također moglo utjecati na statistiku.



**Tablica 3. Prikaz podudarnosti u projektima „Sretan rođendan“ i „Ispričaj priču“**

	<b>STROJNI PRIJEVOD</b>	<b>POTPUNO PODUDARANJE</b>	<b>SLIČNO PODUDARANJE</b>	<b>RJEČNIK</b>
<b>PROJEKT „SRETAN ROĐENDAN“</b>	44	25	13	30
<b>PROJEKT „ISPRIČAJ PRIČU“</b>	35	25	8	28

## **5.2. BILJEŠKE PROJEKATA „SRETAN ROĐENDAN“ I „ISPRIČAJ PRIČU“**

Bilješke projekata „*Sretan rođendan*“ i „*Ispričaj priču*“ korištene su u sustavu obliku .html dokumenata. URL bilješki „*Sretan rođendan*“ upisan je direktno u Toolkit, dok su se bilješke „*Ispričaj priču*“ preuzele na računalo i učitale u sustav. Nakon toga, postupak postavljanja okruženja za rad bio je isti kao i u prethodna dva projekta. Dokumenti su imenovani i odabrani su izvorni i ciljni jezici, a zatim su im se pridružile interne prijevodne memorije i rječnici projekata. Treba primijetiti kako je .html format najbolje koristiti ako je cilj prevoditi web stranicu kao takvu. U ovom slučaju nije bilo u interesu prevoditi nazive izbornika, dostupne jezike i slično, no zbog formata u kojima su se bilješke nalazile bilo je potrebno to učiniti jer je Toolkit i te dijelove web stranica učitao kao segmente.

Bilješke projekta „*Sretan rođendan*“ prevodile su se 24 minute, a bilješke projekta „*Ispričaj priču*“ 19 minuta. Prema procjeni prije prevođenja, od 462 riječi bilježaka „*Sretan rođendan*“ 25 riječi se 100% podudaralo sa prijevodnom memorijom, 2 riječi bile su *high fuzzy* podudarnosti, 424 riječi strojno su prevedene, a 11 riječi se ponovilo. U bilješkama „*Ispričaj priču*“ nalazilo se 413 riječi od kojih se također 25 podudaralo 100% sa prijevodnom memorijom, 6 riječi bile su *high fuzzy* podudarnosti, strojno je prevedena 371 riječ, a ponovilo ih se i u ovom slučaju 11. Pri završetku prevođenja, prve su bilješke pokazivale dovršenost od 68%, a druge su, sa 70%, bile nešto bolje. Slike 21 i 22 prikazuju procjenu prijevoda za oba projekta prije prevođenja.

### Translation statistics

The statistics below show the scope of your translation project when the source document was uploaded. [Learn more](#)

In-context translation	0 words
100% TM matches	25 words
'High fuzzy' TM matches	2 words
Machine translation	424 words
Repeated text	11 words
<hr/>	
Total:	462 words

OK

**Slika 21. Statistika bilješki projekta „Sretan rođendan“**

### Translation statistics

The statistics below show the scope of your translation project when the source document was uploaded. [Learn more](#)

In-context translation	0 words
100% TM matches	25 words
'High fuzzy' TM matches	6 words
Machine translation	371 words
Repeated text	11 words
<hr/>	
Total:	413 words

OK

**Slika 22. Statistika bilješki projekta „Ispričaj priču“**

U tablici 4 se nalazi broj podudarnosti oba projekta sa strojnim prijevodom, prijevodnom memorijom i rječnicima. Tablica prikazuje statistiku projekata nakon prevođenja. U bilješkama je primijećeno više djelomičnih podudarnosti nego li stopostotnih podudarnosti sa prijevodnom memorijom. Također, važno je reći da bi brzina i kvaliteta prijevoda bili bolji, a broj podudarnosti veći da je pri prevođenju bilješki „Ispričaj priču“ bilo moguće koristiti i prijevodnu memoriju bilješki „Sretan rođendan“. S obzirom da su dva

dokumenta vrlo slična, proces bi nedvojbeno bio brži jer bi se većina teksta podudarala sa već prevedenim tekstom prijašnjeg projekta.

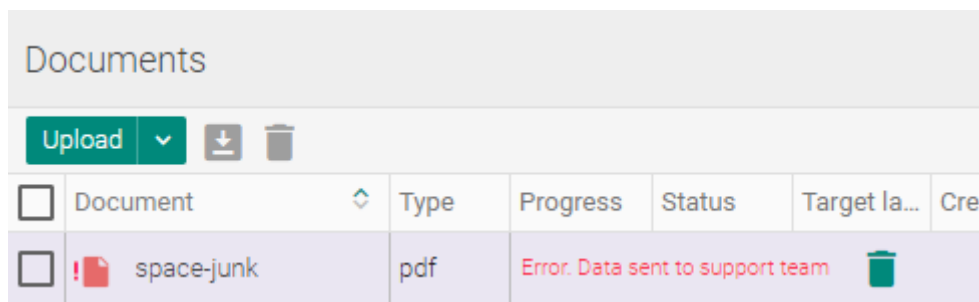
**Tablica 4. Prikaz podudarnosti u bilješkama projekata „Sretan rođendan“ i „Ispričaj priču“**

	<b>STROJNI PRIJEVOD</b>	<b>POTPUNO PODUDARANJE</b>	<b>SLIČNO PODUDARANJE</b>	<b>RJEČNIK</b>
<b>BILJEŠKE „SRETAN ROĐENDAN“</b>	26	8	8	6
<b>BILJEŠKE „ISPRIČAJ PRIČU“</b>	24	9	14	4

## 6. PREVOĐENJE POMOĆU SMARTCATA

Korpus korišten u analizi rada sa Smartcatom sastoji se od pet dokumenata koji su preuzeti sa službene stranice Code Cluba, a pokrivaju cjelinu o učenju programiranja u Scratchu. Tri su dokumenta projekti naziva „*Istjerivači duhova*“ („*Ghostbusters*“), „*Mačke!*“ („*Cats!*“) i „*Smeće iz svemira*“ („*Space Junk*“) osmišljeni kao kreativne radionice tijekom kojih djeca uče kako kreirati varijable za pohranu podataka te ih, uz petlje, koristiti za izradu različitih igara. Dva dokumenta su bilješke projekata za voditelje kluba, „*Istjerivači duhova*“ i „*Smeće iz svemira*“, u kojima su navedeni ishodi učenja, kratak opis projekta te svi resursi, zadaci i dodatne upute. Projekt „*Mačke!*“ nema dostupnih bilješki za voditelje kluba.

Dokumenti su preneseni na Smartcat platformu preko računala. S obzirom da se na stranicama Code Cluba projekti mogu preuzeti samo u .pdf formatu, prvotna zamisao bila je taj format koristiti pri prevođenju u Smartcatu. Međutim, iako Smartcat podržava .pdf format dokumenata, treba naglasiti da dokumenti ne smiju biti dulji od 10 stranica. Ukoliko je .pdf dokument dulji od 10 stranica, onda ga korisnik može prenijeti na platformu samo uz nadoplatu. Moguće je zatražiti dodatnih 20, 50, 100, 200 ili 300 stranica, a cijene se kreću od 1 dolar (za 20 stranica) do 18 dolara (za 500 stranica). Kupljene stranice vrijede godinu dana od dana kupnje, no bitno je napomenuti da su potrošne. Primjerice, korisnik je odlučio kupiti 20 stranica koje mu sada stoje na raspolaganju za rad na .pdf dokumentima duljima od 10 stranica. Zatim je započeo rad na projektu u .pdf formatu koji ima, primjerice, 15 stranica. Tada je korisniku ostalo još samo 5 stranica od kupljenih 20. Iz tog razloga odabir projekata za analizu bio je donekle ograničen njihovim brojem stranica. Naravno, moguće rješenje je i podijeliti dokument na više dokumenata, no s obzirom da su na raspolaganju bili projekti duljine do 10 stranica, oni su i korišteni u sustavu. Dakle, svi korišteni projekti su odgovarajuće duljine, odnosno nemaju više od 10 stranica. Međutim, samo je projekt „*Mačke!*“ u sustavu korišten u .pdf formatu. Projekti „*Istjerivači duhova*“ i „*Smeće iz svemira*“ u sustavu su korišteni u .doc formatu. Razlog tome je što je sustav iz nekog razloga imao problema sa prenošenjem .pdf dokumenata. Projekt „*Mačke!*“ prenesen je s lakoćom no nakon toga Smartcat više nije dopuštao prijenos .pdf dokumenata te je konstantno javljao grešku (slika 23).



**Slika 23. Greška pri prijenosu .pdf dokumenta u Smartcatu**

Alternativno rješenje bilo je pretvoriti .pdf dokumente u .doc format, što je i učinjeno preko *online* stranice za pretvorbu dokumenata. Dokumenti sa bilješkama za voditelje kluba dostupni su samo na web stranici Code Cluba te ih je moguće preuzeti samo u .html formatu. Nasreću, Smartcat podržava rad i s tim tipom dokumenata. Konačno, za prevođenje su bila spremna dva dokumenta u .doc formatu („*Istjerivači duhova*“ i „*Smeće iz svemira*“), jedan dokument u .pdf formatu („*Mačke!*“) i dva dokumenta u .html formatu (bilješke projekata „*Istjerivači duhova*“ i „*Smeće iz svemira*“). Slike projekata su, kao i kod projekata prevedenih pomoću Toolkita, snimke zaslona kôda pisanog u Scratchu. Za razliku od GTT-a, Smartcat ima mogućnost prepoznavanja teksta na slikama te ga prikazuje u sustavu u obliku klasičnih segmenata. To je veoma korisno u ovakvim situacijama, no treba naglasiti kako spomenute slike moraju biti dobre kvalitete. U protivnom će Smartcat loše prepoznati tekst i prenijeti ga krivo u sustav. Iz tog razloga, pri prijenosu u sustav neki dijelovi teksta na slikama nisu nalikovali tekstu koji se nalazio u originalnom dokumentu te nisu prevedeni jer nije bilo moguće utvrditi o kojem dijelu kôda se radi. Oni dijelovi teksta koji su bili prepoznatljivi preveli su se na isti način kao i ostali segmenti.

U tablici 5 je navedena statistika korpusa te su dokumenti označeni rednim brojevima: „*Istjerivači duhova*“ (1.), bilješke „*Istjerivači duhova*“ (2.), „*Mačke!*“ (3.), „*Smeće iz svemira*“ (4.), bilješke „*Smeće iz svemira*“ (5.).

**Tablica 5. Statistika korpusa**

TEKST						PROSJEK
	1.	2.	3.	4.	5.	
<b>BROJ ZNAKOVA</b>	4564	1819	8209	4778	1220	4118
<b>BROJ RIJEČI</b>	994	222	1671	1071	208	833,20
<b>BROJ SEGMENTATA</b>	149	41	296	153	42	136,20
<b>BROJ SLIKA</b>	16	1	23	23	0	12,6
<b>VRIJEME PREVOĐENJA</b>	56 min	25 min	1 h 45 min	54 min	11 min	50 min

\*Broj znakova bez praznina

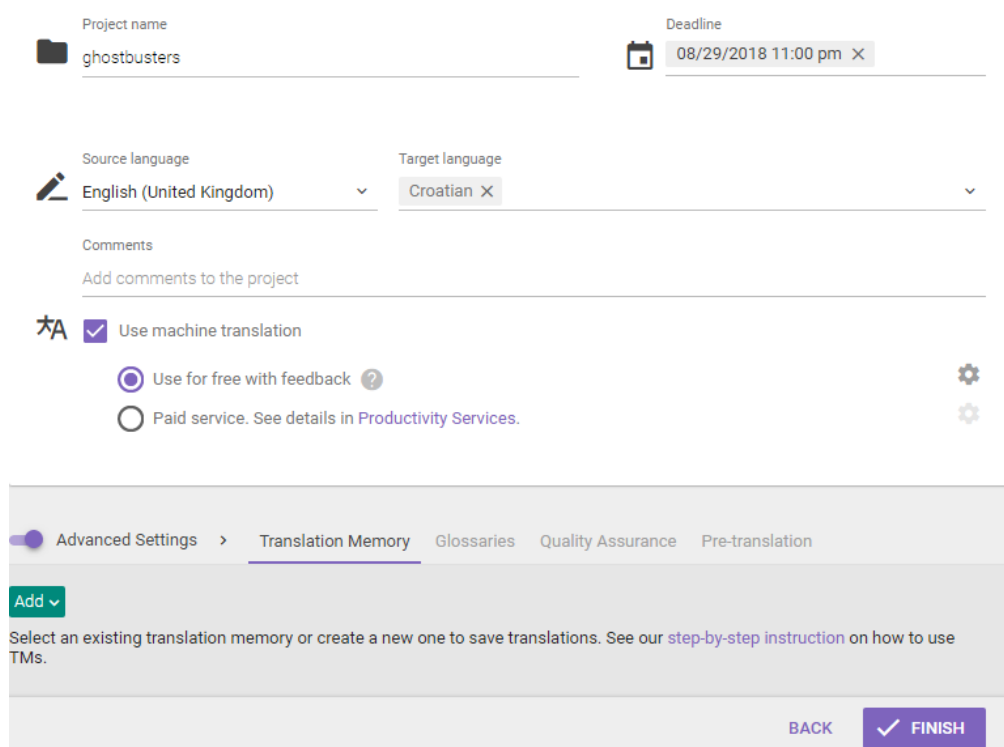
## **6.1. PROJEKT „ISTJERIVAČI DUHOVA“**

Prvi korak u pripremanju okruženja za rad bio je prijenos projekta u .doc formatu na Smartcat platformu. S obzirom da je ovo prvi projekt na kojem se radilo, nikakva prijevodna memorija nije se dodala s računala jer kao takva nije niti postojala. U drugom koraku Smartcat je automatski imenovao projekt na temelju naziva dokumenta. Uz to, samostalno je određen rok za dovršavanje projekta, odabrani su izvorni i ciljni jezik (engleski i hrvatski) te prikazivanje prijedloga strojnog prijevoda. U naprednim postavkama kreirala se nova prijevodna memorija, kao i rječnik. Naziv i jezici automatski se pridodaju prijevodnim memorijama i rječnicima, no korisnik je u mogućnosti izmijeniti njihova svojstva.

Nadalje, podešavaju se svojstva kontrole kvalitete prijevoda<sup>11</sup>. QA navodi različite kategorije i potkategorije mogućih grešaka. Osnovne kategorije vezane su uz terminologiju, prijevodne memorije, datume i brojeve, interpunkciju, pravopis, oznake oblikovanja i ostalo. Unutar svake kategorije postoje potkategorije u kojima su detaljnije navedene greške. Za svaku je grešku moguće odabrati želimo li da nas sustav o njoj obavijesti ili da ih ignorira, a neke je greške moguće označiti kao krucijalne što znači da sustav neće dopustiti potvrđivanje segmenta dok se greška ne ispravi. U ovome projektu QA postavke nisu se mijenjale.

<sup>11</sup> Engl. *QA = Quality Assurance*

Moguće je postaviti i automatski prijevod, odnosno pravila po kojima će se u segmente ciljnog jezika automatski umetnuti prijedlozi iz prijevodne memorije, prijedlozi strojnog prijevoda i datumi i brojevi. Također se može postaviti da se ti segmenti automatski potvrđuju. U projektu „*Istjerivači duhova*“ nije korišteno automatsko prevođenje. Na slici 24 prikazano je pripremanje okruženja za rad na projektu, odnosno dodjeljivanje naziva, roka, jezika i podešavanje naprednih postavki.



**Slika 24. Pripremanje okruženja za rad u Smartcatu**

Prevođenje projekta „*Istjerivači duhova*“ trajalo je 56 minuta. Sustav je automatski umetao brojeve i ponavljajuće segmente, odnosno segmente koji su se već pojavili u tekstu te su potvrđeni. Također, crvenom linijom podvučene su riječi koje imaju pravopisne pogreške iako treba naglasiti da često uopće nisu krivo napisane. S obzirom da je .pdf dokument pretvoren u .doc dokument, razmještaj nekih segmenata je pomalo neobičan i rascjepkan. Primjerice, u jednom segmentu nalazi se „*when this sprite clicke*“, a u sljedećem segmentu samo slovo „*d*“. Problem se javlja jer segmente nije moguće spojiti osim ako se nalaze u istom odlomku, što ovdje nije bio slučaj. Iz tog razloga segmenti su se morali prevesti odvojeno pa je prijevod prvog segmenta glasio „*kada je lik kliknu*“, a drugog samo „*t*“.

Klikom na „Done“ svi se segmenti spremaju i projekt se označava kao dovršen. Tada je moguće preuzeti dokument na izvornom jeziku i u originalnom formatu, dvojezični dokument u .doc formatu te dokument na ciljnom jeziku u .doc formatu. Dvojezični dokument formatiran je u obliku tablice sa četiri stupca u kojem se nalazi popis svih segmenata dokumenta. U prvom stupcu naveden je redni broj segmenta, zatim segment na izvornom jeziku, segment na ciljnom jeziku i zadatak (slika 25). Zadatak pojašnjava je li segment na ciljnom jeziku ručno preveden ili je odabran neki od prijedloga strojnog prevoditelja ili prijevodne memorije.

## Smartcat

№	Source (EN-GB)	Target (HR)	Task
1	Ghostbusters	Istjerivači duhova	Translation
2	Introduction	Uvod	Translation (MT)
3	You are going to make a ghost-catching game!	Napravit češ igru hvatanja duhova!	Translation
4	Step 1:	Korak 1:	Translation
5	Animating a ghost	Animacija duha	Translation
6	Activity Checklist	Zadatci	Translation
7	Start a new Scratch project, and delete the cat sprite so that your project is empty.	Započni novi Scratch projekt i obriši lik mačke tako da projekt ostane prazan.	Translation
8	You can find the online Scratch editor at <a href="http://jump.to/cc/scratch-">jump.to/cc/scratch-</a>	Online Scratch nalazi se na linku <a href="http://jump.to/cc/scratch-">jump.to/cc/scratch-</a>	Translation (MT)

Slika 25. Izgled dvojezičnog dokumenta

U projektu „*Istjerivači duhova*“ upozorenje o greškama javljalo se u skoro svakom segmentu. U dokumentu sa 149 segmenata javilo se 85 grešaka. Popis grešaka nalazi se u tablici 6.

Tablica 6. Popis grešaka u projektu „*Istjerivači duhova*“

TIP GREŠKE	BROJ GREŠAKA
OZNAKE	0
TERMINOLOGIJA	1



PRIJEVODNE MEMORIJE	0
DATUMI I BROJEVI	3
INTERPUNKCIJA	1
PRAVOPIS	70
OZNAKE OBLIKOVANJA	10
OSTALO	0

Najviše grešaka pojavilo se u kategoriji pravopisa pri prevođenju glagola iako riječi najčešće nisu bile krivo napisane. Također se pravopisne greške javljaju kod dijelova segmenata koji su identični kao izvorni jezik, a nisu se preveli. Neke od grešaka bilo je moguće ispraviti, a to su najčešće višak ili manjak praznina između riječi ili različit interpunkcijski znak u izvornom i ciljnom jeziku. Neka upozorenja mogu se ignorirati. Primjerice, sustav javlja grešku kada je broj „0“ preveden riječima „nula“, no takav prijevod stvar je stilskog izražavanja, subjektivnosti i konvencionalnosti jezika.

U zbrajanju grešaka ubrojane su i greške koje se ponavljaju. Isto tako vodilo se računa o podudarnostima između ručnog prijevoda i prijedloga strojnog prevoditelja te je zabilježeno 18 slučajeva kada se strojni prijevod slagao s mojim prijevodom. Treba napomenuti da su se tu našle i dvije iznimke: u jednom segmentu strojni prevoditelj je za „code“ ponudio „kod“ dok sam se ja odlučila za „kôd“. Također, za izraz „how big is it?“ ponudio je prijevod „koliko je velik?“ dok sam ja odlučila umetnuti i subjekt tog pitanja pa je moj prijevod glasilo „koliko je objekt velik?“. Unatoč tim malim razlikama, strojni prijevodi tretirali su se kao 100% točni. Uočeno je i da strojni prevoditelj sve brojeve prevodi riječima, osim kada se radi o negativnim brojevima ili brojevima u kombinaciji sa nekim drugim znakom, primjerice „= 0“. Također, prevoditelj je u nekim segmentima nudio prijevod polovično na hrvatskom, a polovično na ruskom jeziku pa je tako za izraz „more randomness“ ponudio „više рандома“.

Popis pogrešaka može se preuzeti u obliku Excel tablice u kojoj su navedeni tip pogreške, podtip, segment na izvornom jeziku, segment na ciljnom jeziku, naziv dokumenta, kôd za jezik na koji se prevodi te poveznica na svaki segment (slika 26).

1	Error type	Error subtype	Source	Target	Document	Language	Segment	Link
2	Dates and Numbers	numbers in source and	game is started (by	potrebno postaviti	ghostbusters	hr	75	<a href="#">go to segment</a>
3	Dates and Numbers	Dates and numbers in	Whenever a ghost is	Svaki put kad je duh	ghostbusters	hr	82	<a href="#">go to segment</a>
4	Dates and Numbers	Dates and numbers in	The game should stop	Igra prestaje kada vrijeme	ghostbusters	hr	98	<a href="#">go to segment</a>
5	Other	Source and	new.	new.	ghostbusters	hr	9	<a href="#">go to segment</a>

Slika 26. Izgled dokumenta sa popisom grešaka

Nadalje, prikaz statistike projekta vidljiv je u sustavu, ali se isto tako može preuzeti na računalo u .xlsx formatu ili .xml formatu za Trados pakete. Osim već spomenutog originalnog dokumenta, dvojezičnog dokumenta i prijevoda, mogu se preuzeti i prijevodne memorije projekta u .tmx formatu, ali i .xliff datoteka koja omogućava korištenje projekta i u drugim CAT alatima.

## 6.2. PROJEKT „MAČKE!“

Projekt „Mačke!“ za rad je pripremljen na isti način kao i projekt „Istjerivači duhova“. Nakon prijenosa dokumenta u sustav, napravljena je nova prijevodna memorija i novi rječnik, a QA postavke ostavljene su na automatskim podešenjima. Specifičnost ovog dokumenta je njegov .pdf format. Vrlo je važno napomenuti da kvaliteta .pdf dokumenta iznimno utječe na njegov prikaz u Smartcat sustavu. Projekt „Mačke!“ bio je lošije kvalitete, točnije slike su bile mutne, pa je loše i učitano u sustav. Sukladno tome, velik broj segmenata u ovom projektu uopće nije imao smisla ili je bio veoma izmijenjen. Iz tog razloga mnogo se vremena izgubilo na prevođenje jer je za prijevod bilo potrebno stalno provjeravati originalni dokument kako bi se otkrilo što je točno trebalo pisati u segmentu. Tako je, primjerice, segment „delete this clone“ postao „delete this done“, „change y by -2“ je postao „V by C“, a „move 10 steps“ je dobio besmislen ekvivalent „CD steps“.

Prevođenje projekta trajalo je 1 sat i 54 minute. S obzirom da su postavke bile iste kao i u prethodnom projektu, sustav je automatski umetao brojeve i ponavljajuće segmente te označavao nepravilno napisane riječi. I u ovom je projektu pronađen velik broj grešaka – čak 268 (tipovi grešaka zabilježeni su u tablici 7). Time je projekt „Mačke!“ postao dokument sa najviše zabilježenih grešaka. Ipak, to nije začuđujuće uzme li se u obzir već spomenuta loša kvaliteta dokumenta i činjenica da velik broj segmenata u dokumentu nije imao nikakvoga smisla. Također, ovaj dokument imao je i najviše segmenata – ukupno 296.

**Tablica 7. Popis grešaka u projektu „Mačke!“**

<b>TIP GREŠKE</b>	<b>BROJ GREŠAKA</b>
OZNAKE	0
TERMINOLOGIJA	4
PRIJEVODNE MEMORIJE	0
DATUMI I BROJEVI	0
INTERPUNKCIJA	19
PRAVOPIS	216
OZNAKE OBLIKOVANJA	0
OSTALO	29

Najviše upozorenja i ovoga je puta bilo u kategoriji pravopisa i također je velika većina njih bila lažna. Također se mnogo puta javljalo upozorenje kako između riječi ili slova nedostaje razmak, što je bilo besmisleno jer u izvornom segmentu također nije postojala praznina. Neke su se greške ispravile, poput djelomičnog podudaranja segmenata u izvornom i ciljnom jeziku koji se uglavnom javljao u segmentima gdje su se nalazile poveznice na web stranice. Iz tog razloga dio segmenta bio je preveden, a dio (poveznica) ne. Podudarnosti mojih prijevoda sa strojnim prijevodima bilo je 56, a prevoditelj je i ovoga puta nudio prijedloge na ruskom jeziku. Izraz „*by User Nan*“ preveo je kao „*Korisnik НЭН*“, a za slovo „y“ nudio je „*узрек*“. I ovdje treba napomenuti da se u tih 56 podudaranja nalazi i prijevod izraza „*open the starter project*“ kojeg je prevoditelj preveo kao „*otvorite početni projekt*“, dok sam se ja odlučila na neformalni stil i prevela ga kao „*otvori početni projekt*“.

### **6.3. PROJEKT „SMEĆE IZ SVEMIRA“**

Projekt „*Smeće iz svemira*“ također je trebao biti procesiran u .pdf formatu, no iako ima 10 stranica, pokušaj prijenosa dokumenta bio je neuspješan te je također pretvoren u .doc format i kao takav korišten u Smartcat sustavu. Postavke za ovaj projekt bile su nešto drugačije od postavki prethodna dva projekta. Cilj je bio ispitati na koji način će se odvijati proces prevođenja ako se iskoriste mogućnosti poput dodavanje postojeće prijevodne

memorije i rječnika, podešavanje QA opcije i slično. Pretpostavka je bila da će prevođenje biti brže, sam proces olakšan, a prijevod konzistentan sa ostalim materijalima.

U prvom koraku se, nakon prijenosa dokumenta, odabrala postojeća prijevodna memorija projekta „*Istjerivači duhova*“ sa računala. Zatim su se, u sljedećem koraku, postavili izvorni i ciljni jezik, rok završetka projekta i naziv, a dopustilo se i prikazivanje prijedloga strojnog prevoditelja. Pretpostavka je bila da će projekt postati nedostupan za rad ako ga se ne dovrši u zadanom roku ili da će sustav slati obavijesti o približavanju krajnjeg roka za završetak projekta, no to nije bio slučaj. Sa projektom se po tom pitanju ništa nije dogodilo te postavljanje roka nije utjecalo na rad. Postavljanje roka je jako korisna značajka kada se radi o stvarnom poslovnom odnosu gdje poslodavac očekuje da se posao dovrši u određenom vremenskom roku. U ovom slučaju, gdje se prevodilo prvenstveno u svrhu analize sustava, postavljanje roka za završetak projekta nije igralo nikakvu ulogu.

Iako je jedna prijevodna memorija pridružena projektu, u svrhu poboljšanja rezultata pridodana je i prijevodna memorija projekta „*Mačke!*“. Zatim su odabrani svi rječnici dostupni u sustavu, dakle rječnici oba prethodna projekta, a napravljen je i novi isključivo za projekt „*Smeće iz svemira*“. U QA postavkama nije se puno toga promijenilo – sve je ostalo na postavkama sustava osim upozorenja o identičnom izgledu segmenata u izvornom i ciljnom jeziku<sup>12</sup> i upozorenja o djelomično jednakom izgledu segmenata u izvornom i ciljnom jeziku<sup>13</sup>. Pretpostavka da će prijevod ovog projekta biti brži i jednostavniji djelomično se pokazala točnom. Proces je doista bio lakši jer je sustav nudio prijevode segmenata iz prijevodnih memorija pridruženih projektu. Osim stopostotnih podudaranja, koja su se jednostavno umetala u tekst, prikazivala su se i djelomična podudaranja sa prijevodnim memorijama. Ta je značajka također bila korisna jer je olakšavala prijevod sličnih segmenata jednostavnim izmjenjivanjem dijelova koji se razlikuju u projektu i prijevodnoj memoriji. Primjerice, u nekoliko se navrata pojavio segment koji objašnjava na kojoj web stranici se može pristupiti Scratch projektu. U tom slučaju prijevod segmenata je uvijek bio isti, samo se mijenjala navedena poveznica. Također, termini koji se nalaze u nekom od rječnika označeni su u segmentima, a u CAT kartici navedeni su njihovi prijevodi. Ukoliko korisnik odluči na drugačiji način prevesti termin koji se nalazi u rječniku, onda sustav javlja upozorenje da se prijevod razlikuje od predloženog<sup>14</sup>. Unatoč svemu tome, ukupno vrijeme prevođenja iznosilo je 54 minute, što je samo dvije minute kraće od procesa prevođenja projekta „*Istjerivači*

---

<sup>12</sup> Engl. *source and target are identical*

<sup>13</sup> Engl. *source and target partially coincide*

<sup>14</sup> Engl. *translation of a term does not match the glossary*

*duhova*“. U usporedbi sa projektom „*Mačke!*“, koji se prevodio čak sat vremena duže, ovaj je projekt ipak polučio puno bolje rezultate u kontekstu vremena. Ipak, treba imati na umu da je projekt „*Mačke!*“ imao i više segmenata.

Projekt sadrži 153 segmenta među kojima se javilo ukupno 100 grešaka. Od 100 grešaka skoro sve su bile u kategoriji pravopisa – čak njih 93. Ostalih 7 grešaka raspodijelilo se unutar kategorija terminologije, datuma i brojeva i interpunkcije. U tablici 8 nalazi se detaljniji prikaz grešaka projekta.

**Tablica 8. Popis grešaka u projektu „*Smeće iz svemira*“**

TIP GREŠKE	BROJ GREŠAKA
OZNAKE	0
TERMINOLOGIJA	2
PRIJEVODNE MEMORIJE	0
DATUMI I BROJEVI	2
INTERPUNKCIJA	3
PRAVOPIS	93
OZNAKE OBLIKOVANJA	0
OSTALO	0

Podudarnosti mojih prijevoda sa strojnim prijevodima je u ovom projektu bilo 17, a unutar njih zabilježene su dvije iznimke. Prijedlog za prijevod segmenta „*2 players*“ glasio je „*2 igrača*“, dok sam ja odlučila napisati broj 2 riječima te je moj prijevod glasio „*Dva igrača*“. Također, umjesto da segment „*More space junk*“ prevedem „*Više svemirskog smeća*“, kao što je to bilo predloženo, ja sam izostavila riječ „*svemirskog*“ i prevela ga samo „*Više smeća*.“ S obzirom da su se u ovom projektu koristili unaprijed pripremljene prijevodne memorije i rječnici, analizirala su se i ta dva aspekta prevodenja. U kontekstu prijevodnih memorija, sustav je pronašao 52 podudarnosti sa prijevodnim memorijama projekata „*Istjerivači duhova*“ i „*Mačke!*“. Uz to, javilo se 12 djelomičnih podudarnosti, odnosno podudarnosti manjih od 100%. Također, sustav je 25 puta predložio prijevode termina koji se

nalaze u rječnicima projekata „*Istjerivači duhova*“ i „*Mačke!*“. U tablici 9 se nalazi prikaz podudarnosti sa prijevodnim memorijama i rječnicima.

**Tablica 9. Prikaz podudarnosti u projektu „*Smeće iz svemira*“**

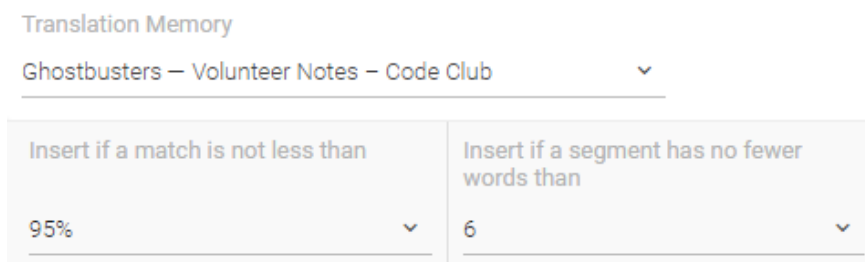
	<b>STROJNI PRIJEVOD</b>	<b>POTPUNO PODUDARANJE</b>	<b>SLIČNO PODUDARANJE</b>	<b>RJEČNIK</b>
<b>BROJ PODUDARNOSTI</b>	17	52	12	25

#### **6.4. BILJEŠKE PROJEKTA „*ISTJERIVAČI DUHOVA*“**

Za prevođenje bilješki projekta za voditelje kluba „*Istjerivači duhova*“ potrebno je bilo preuzeti web stranicu u .html formatu te ju na isti način kao i prethodne projekte prenijeti u Smartcat sustav. Dokumentu je pridružen rječnik i prijevodna memorija projekta „*Istjerivači duhova*“ te je kreirana nova prijevodna memorija specifično za ovaj projekt. Razlog stvaranja nove prijevodne memorije je zato što su bilješke svih projekata relativno slične te je pretpostavka da će ova prijevodna memorija biti korisna pri prevođenju posljednjeg dokumenta (bilješki projekta za voditelje kluba „*Smeće iz svemira*“).

QA postavkama pristupilo se kao i u prethodnim projektima – ostavljene su početne postavke osim upozorenja o identičnom izgledu segmenata izvornog i ciljnog jezika te upozorenja o djelomičnom podudaranju segmenata u ciljnom i izvornom jeziku. Ipak, iskušano je još jedno svojstvo Smartcata – automatsko pretprevođenje<sup>15</sup>. U postavkama automatskog pretprevođenja moguće je odabrati takozvana pravila po kojima će se u segmente na ciljnom jeziku automatski umetati prijevodi. Korisnik može birati želi li da se automatski umetnu prijevodi koji se podudaraju sa prijevodnom memorijom i uz to odrediti koji minimalni postotak podudaranja mora biti te, kao dodatna opcija, koja je minimalna duljina segmenta u koji se umeće prijevod. Primjerice, korisnik može odlučiti da se u tekst automatski umetnu oni segmenti koji se s prijevodnom memorijom podudaraju najmanje 95% i koji sadrže najmanje 6 riječi (slika 27).

<sup>15</sup> Engl. *pre-translation*



**Slika 27. Podešavanje automatskog prevođenja**

Također, moguće je odabrati automatsko umetanje datuma i brojeva, kao i prijedloga strojnog prevoditelja. U ovom su projektu odabrana sva tri pravila, a minimalni postotak podudaranja prijevoda iz prijevodne memorije sa segmentima u ciljnom jeziku postavljen je na 100%.

S obzirom da je ovaj projekt puno kraći od prethodnih te sadrži samo 41 segment, prevođenje je trajalo 25 minuta. Ukupan zbroj grešaka je 89, s naglaskom na greškama koje se tiču prijevodnih memorija. Točnije, sustav je javljao upozorenje da je nekoliko identičnih podudarnosti pronađeno u prijevodnim memorijama za isti segment<sup>16</sup>. To nije začuđujuće s obzirom da je projektu pridružena prijevodna memorija projekta „*Istjerivači duhova*“. Sustav automatski pohranjuje prijevode segmenata u obje prijevodne memorije te zbog toga pronalazi identične prijevode u različitim prijevodnim memorijama. Primjerice, segment „*Use basic programming constructs to create simple programs*“ nalazi se samo u bilješkama „*Istjerivači duhova*“, no Smartcat je njegov prijevod pohranio u prijevodnoj memoriji projekta i bilježaka te pri završetku prevođenja navodi da se pojavila „greška“, odnosno da je pronađena identična podudarnost sa segmentom u više od jedne prijevodne memorije. Uz to, javile su se još i greške u pravopisu i interpunkciji. U tablici 10 prikazan je popis upozorenja.

**Tablica 10. Popis grešaka u bilješkama projekta „*Istjerivači duhova*“**

TIP GREŠKE	BROJ GREŠAKA
OZNAKE	0
TERMINOLOGIJA	0

<sup>16</sup> Engl. *several exact/context matches found for the segment*

PRIJEVODNE MEMORIJE	41
DATUMI I BROJEVI	0
INTERPUNKCIJA	15
PRAVOPIS	33
OZNAKE OBLIKOVANJA	0
OSTALO	0

Strojni prijevodi 16 puta su se podudarali sa potvrđenim prijevodima, a u to su ubrojene i iznimke kada su segmenti „*Frequently Asked Questions*“ i „*Privacy & Terms*“ prevedeni kao „*Često Postavljena Pitanja*“ i „*Privatnost I Uvjeti*“ dok je u potvrđenim segmentima samo prvo slovo bilo tiskano. Također, pronađeno je 9 podudarnosti sa prijevodnom memorijom i 3 djelomične podudarnosti. Uz to, sustav je 3 puta upozorio na podudarnost termina sa terminom u rječniku. Treba napomenuti da su se u nekim slučajevima javljala preklapanja između prijedloga prijevodne memorije, termina u rječniku i strojnog prijevoda. Primjerice, za riječ „*Introduction*“ pronađena je stopostotna podudarnost u prijevodnoj memoriji („*Uvod*“), slično podudaranje u prijevodnoj memoriji („*Uvod:*“) i podudarnost sa terminom u rječniku („*uvod*“). Tablica 11 prikazuje podudarnosti sa prijevodnom memorijom i rječnicima.

**Tablica 11. Prikaz podudarnosti u bilješkama projekta „*Istjerivači duhova*“**

	STROJNI PRIJEVOD	POTPUNO PODUDARANJE	SLIČNO PODUDARANJE	RJEČNIK
<b>BROJ PODUDARNOSTI</b>	16	9	3	3



## 6.5. BILJEŠKE PROJEKTA „SMEĆE IZ SVEMIRA“

Proces prevođenja bilješki projekta za voditelje kluba „*Smeće iz svemira*“ odvijao se na isti način kao i proces prevođenja samog projekta. Materijalima su pridružene četiri prijevodne memorije: projekta „*Istjerivači duhova*“ (prenesena sa računala), projekta „*Mačke!*“, bilješki projekta „*Smeće iz svemira*“ te je kreirana nova prijevodna memorija. Nadalje, korišteni su rječnici sva tri projekta. U QA postavkama je, osim ignoriranja identičnih segmenata u izvornom i ciljnom jeziku i djelomično istih segmenata u izvornom i ciljnom jeziku, odabrano zanemarivanje grešaka vezanih uz prijevodne memorije. Točnije, postavljeno je ignoriranje segmenata u kojima je pronađeno nekoliko identičnih prijedloga u prijevodnim memorijama i segmenata čiji se prijevodi razlikuju od identičnih prijedloga u prijevodnim memorijama<sup>17</sup>. Također su se koristila sva tri pravila automatskog prevođenja, točnije dopušteno je automatsko umetanje brojeva i datuma, podudarnosti iz prijevodne memorije koje iznose minimalno 100% te strojnih prijevoda.

Prevođenje ovog projekta nedvojbeno je trajalo najkraće – samo 11 minuta. Razlog tome je vjerojatno činjenica da je ovaj projekt imao najviše dodatnih resursa kao pomoć. Pridružene su mu tri prijevodne memorije od kojih je jedna preuzeta iz projekta veoma sličnom ovome te su se javile mnoge podudarnosti, kako stopostotne tako i djelomične. Isto tako, tri su rječnika pomogla konzistentnosti i bržem odabiru pravilnog prijevoda termina. U kontekstu vremena, postavke automatskog prevođenja igrale su veliku ulogu. Sustav je u sve segmente umetnuo stopostotne podudarnosti sa prijevodnim memorijama, a ako njih nije bilo, tada se u segmente umetao strojni prijevod. Iako se strojni prijevod nije uvijek koristio, bilo je mnogo brže i jednostavnije samo izmijeniti dijelove prijevoda i potvrditi segment, umjesto da se cijeli prijevod segmenta ručno piše. Uzevši sve to u obzir, treba napomenuti i da je u ovom dokumentu, sa ukupno 42 segmenta, bilo najmanje grešaka, točnije 33. Greške su se i ovoga puta pretežno javile u kategoriji pravopisa i interpunkcije. Popis svih grešaka nalazi se u tablici 12.

**Tablica 12. Popis grešaka u bilješkama projekta „*Smeće iz svemira*“**

TIP GREŠKE	BROJ GREŠAKA
OZNAKE	0
TERMINOLOGIJA	1

<sup>17</sup> Engl. *translation is different from exact/context matches*

PRIJEVODNE MEMORIJE	0
DATUMI I BROJEVI	1
INTERPUNKCIJA	13
PRAVOPIS	18
OZNAKE OBLIKOVANJA	0
OSTALO	0

Potpunih podudarnosti sa prijevodnom memorijom bilo je 24, dok je djelomičnih bilo 11. Prijedlozi strojnog prevoditelja podudarali su se 12 puta sa potvrđenim prijedlozima, dok je sustav 5 puta upozorio na prijevod termina u rječniku. Tablica 13 prikazuje popis podudaranja.

**Tablica 13. Prikaz podudarnosti u bilješkama projekta „Smeće iz svemira“**

	STROJNI PRIJEVOD	POTPUNO PODUDARANJE	SLIČNO PODUDARANJE	RJEČNIK
<b>BROJ PODUDARNOSTI</b>	12	24	5	11

## 6.6. USPOREDBA PRIJEVODA

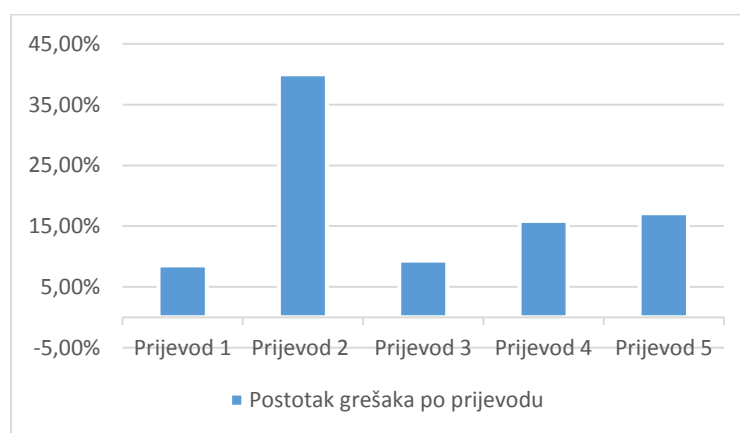
Iako je Smartcat nedvojbeno odradio vrlo dobar posao u procesu prevođenja te su se njegove značajke pokazale iznimno korisnima, neki od prijevoda polučili su veće uspjehe od drugih. U slučaju ovog korpusa, možemo zaključiti da je Smartcat bolje upravljao .doc formatima dokumenata nego .pdf formatom. Na to je utjecala i lošija kvaliteta samog dokumenta pa stoga treba pripaziti kakav se .pdf dokument planira koristiti za rad. Iz tog razloga neki od segmenata u projektu „Mačke!“ nisu imali nikakvog smisla. Primjerice, u izvornom dokumentu pojavili su se segmenti poput „sacua“ i „v bsh\_si“ za koje nije bilo

moгуće ni definirati gdje u originalnom dokumentu pripadaju ni što bi trebali značiti. Ipak, treba napomenuti da su se ove situacije javljale najčešće kod tekstova koji se nalaze na slikama u dokumentu. Kao što je već spomenuto, slike su zapravo snimke zaslona Scratch programa i sadrže uglavnom blokove naredbi ili prikaze korisničkog sučelja pa su u .pdf dokumentu neki dijelovi teksta mutni, a slova preblizu te ih Smartcat zato teško prepoznaje.

U svrhu bolje i kvalitetnije analize i usporedbe načina rada Smartcat alata, u postavkama okruženja za rad pojedinog dokumenta koristile su se različite kombinacije Smartcatovih značajki. Tako su u projektima „*Istjerivači duhova*“ i „*Mačke!*“ kreirane nove prijevodne memorije i novi rječnici. QA postavke ostavljene su na automatskim podešenjima. Pri radu sa projektom „*Smeće iz svemira*“ korištene su prijevodne memorije i rječnici prethodna dva projekta, a kreirana je i još jedna prijevodna memorija specifično za ovaj projekt. Na taj način isprobana je funkcionalnost korištenja prijevodne memorije sa računala, kao i onih prijevodnih memorija koje se nalaze u sustavu. Iskušan je princip kreiranja novog rječnika, ali i dodavanja postojećeg. Podešene su i QA postavke pa su se tako zanemarila upozorenja o identičnim segmentima u izvornom i ciljnom jeziku, kao i upozorenja o djelomično istim segmentima u izvornom i ciljnom jeziku.

Bilješkama za voditelje kluba projekta „*Istjerivači duhova*“ pristupljeno je kao i samom projektu, uz nekoliko iznimki. Bilješkama je pridružen rječnik projekta i prijevodna memorija projekta, prenesena sa računala. Također je napravljena i nova prijevodna memorija koja će se koristiti u bilješkama projekta „*Smeće iz svemira*“. Testirana je i još jedna mogućnost alata, a to je automatsko preprevođenje, pa je tako Smartcat u segmente na ciljnom jeziku automatski umetao datume i brojeve, stopostotne podudarnosti iz prijevodnih memorija ili strojne prijevode. Postavke QA izgledale su kao i u projektu „*Smeće iz svemira*“. Bilješke projekta „*Smeće iz svemira*“ su, uz prijevodnu memoriju projekta i bilješki „*Istjerivači duhova*“ i prijevodnu memoriju projekta „*Mačke!*“, koristile i vlastitu prijevodnu memoriju. Također, na raspolaganju su bili svi do sada kreirani rječnici. Osim zanemarivanja upozorenja o identičnim segmentima u izvornom i ciljnom jeziku, kao i upozorenja o djelomično istim segmentima u izvornom i ciljnom jeziku, ignorirale su se i obavijesti o više identičnih podudarnosti u prijevodnoj memoriji i različitim prijevodima segmenta i prijevodne memorije. U automatskom prevođenju implementirana su sva tri dostupna pravila, odnosno automatsko umetanje datuma i brojeva, podudarnosti iz prijevodne memorije koje iznose najmanje 100% i strojnih prijevoda u slučaju kada ne postoji podudaranje sa prijevodnom memorijom.

Uzevši u obzir različite postavke dokumenata, logično je zaključiti kako su bilješke projekta „*Smeće iz svemira*“ bile najuspješnije. Njihovo prevođenje oduzelo je najmanje vremena, a rezultiralo je kvalitetnim i konzistentnim prijevodom te je uz to projekt imao najmanje upozorenja o greškama (33). Daleko najrazočaravajuće rezultate polučio je projekt „*Mačke!*“ koji se najduže prevodio i uz to je neke segmente bilo nemoguće prevesti. Ovaj projekt imao je i najviše upozorenja o greškama (268). Ipak, to se pripisuje već spomenutoj činjenici da je dokument bio loše kvalitete te ga Smartcat nije niti mogao kvalitetno obraditi. Imajući na umu omjer broja riječi i grešaka koje su se javile u prijevodima, bilješke projekta „*Istjerivači duhova*“ rezultirale su najvećim postotkom grešaka (40,01%), dok je, s druge strane, sam projekt sadržavao najmanji postotak grešaka (8,55%). Postotak grešaka iznosio je 9,34% za projekt „*Smeće iz svemira*“, 15,86% za bilješke projekta „*Smeće iz svemira*“ i 17,11% za projekt „*Mačke!*“. U grafikonu 1 prikazan je postotak grešaka svih projekata u odnosu na njihov broj riječi. Prijevod 1 odnosi se na projekt „*Istjerivači duhova*“, prijevod 2 na bilješke projekta „*Istjerivači duhova*“, prijevod 3 na projekt „*Smeće iz svemira*“, prijevod 4 na bilješke projekta „*Smeće iz svemira*“, a prijevod 5 na projekt „*Mačke!*“.



**Figure 1. Postotak grešaka po prijevodu**

Treba napomenuti da su se, uzevši u obzir sve dokumente, najučestalije javljale pravopisne greške. Ipak, većina tih grešaka bila je neosnovana jer su segmenti pravilno prevedeni. Također, Smartcat je često javljao greške u interpunkciji kada ih nije bilo. Najčešće se to događalo kada je segment sadržavao poveznicu na web adresu. Tada bi Smartcat upozoravao da, primjerice, prije ili nakon točke nedostaje razmak, što nije imalo smisla jer ni u izvornom segmentu razmak ne postoji. Neka od upozorenja bila su opravdana, no nije bilo potrebe za njihovim ispravljanjem. Primjerice, Smartcat je javljao grešku ako je u

ciljnom jeziku broj napisan riječima. Podaci o greškama preuzeti su preko Smartcat platforme dok su podaci o podudaranjima sa prijevodnim memorijama, rječnicima i strojnim prijevodima ručno obrađeni. Također, svim se projektima pristupalo preko *freelancer* korisničkog računa te je moguće da su neke od opcija drugačije ako se koriste drugačiji tipovi korisničkih računa.

U nastavku je tablično prikazana detaljnija statistika projekata. U tablici 14 navedene su korištene funkcionalnosti za pojedini projekt. Tablica 15 prikazuje broj detektiranih grešaka po kategorijama za svaki projekt te njihov ukupan zbroj, a tablica 16 navodi broj podudaranja sa prijevodnim memorijama, rječnicima i strojnim prijevodima za svaki dokument. U tablicama su dokumenti označeni rednim brojevima: „*Istjerivači duhova*“ (1.), bilješke „*Istjerivači duhova*“ (2.), „*Mačke!*“ (3.), „*Smeće iz svemira*“ (4.), bilješke „*Smeće iz svemira*“ (5.).

**Tablica 14. Korištene funkcionalnosti po projektu**

KORIŠTENA FUNKCIONALNOST	PRIJEVOD				
	1.	2.	3.	4.	5.
QA	Da	Da	Da	Da	Da
PRIJEVODNA MEMORIJA*	Ne	Da	Ne	Da	Da
RJEČNIK*	Ne	Da	Ne	Da	Da
ROK ZA PROJEKT	Da	Da	Da	Da	Da
AUTOMATSKO PREVOĐENJE	Ne	Da	Ne	Ne	Da

\* Odnosi se na već postojeće prijevodne memorije i rječnike

**Tablica 15. Broj grešaka po kategorijama za pojedini dokument**

PRIJEVOD	BROJ SEGMENTA	BROJ GREŠAKA PO KATEGORIJAMA								UKUPAN BROJ GREŠAKA
		OZNAKE	TERMINOLOGIJA	PRIJEVODNA MEMORIJA	DATUMI I BROJEVI	INTERPUNKCIJA	PRAVOPIS	OZNAKE OBLIKOVANJA	OSTALO	
1.	149	0	1	0	3	1	70	0	10	85
2.	41	0	0	41	0	15	33	0	0	89
3.	296	0	4	0	0	19	216	0	29	268
4.	153	0	2	0	2	3	93	0	0	100
5.	42	0	1	0	1	13	18	0	0	33
<b>UKUPNO</b>	681	0	8	41	6	51	430	0	39	

**Tablica 16. Broj podudaranja sa prijevodnim memorijama, rječnicima i strojnim prijevodima za pojedini dokument**

PODUDARANJA	PRIJEVOD					UKUPNO
	1.	2.	3.	4.	5.	
STROJNI PRIJEVOD	18	16	56	17	12	119
POTPUNO PODUDARANJE	0	9	0	52	24	85
SLIČNO PODUDARANJE	0	3	0	12	5	20
RJEČNIK	0	3	0	25	11	39

## 7. USPOREDBA ODABRANIH ALATA

Google Translator Toolkit i Smartcat nedvojbeno su korisni alati za prevođenje. U nekim aspektima su veoma slični, dok se u drugima prilično razlikuju. U samom je startu moguće reći da je Smartcat složeniji i kompleksniji te nudi više mogućnosti. Korisnik može jednostavno koristiti više prijevodnih memorija i rječnika na jednom tekstu, sustav tijekom prevođenja upozorava na razne greške, a moguće je preuzeti prikaz statistike projekata, detaljan popis grešaka te dvojezični dokument sa tabličnim prikazom segmenata na izvornom i ciljnom jeziku. Sve to Toolkit manjka. Ipak, jednostavnost Toolkita nije uvijek nedostatak, već se ponekad može pokazati i korisnijim ako je u pitanju jednostavniji proces prevođenja kojem je jedini cilj na brži način prevesti tekst.

Također, Toolkit se pokazao superiornim u kontekstu strojnih prijevoda. Većina ponuđenih strojnih prijevoda u potpunosti se slagala sa izvornim tekstom te ih nije bilo potrebe korigirati. Izvršni prijevodi nisu bili ograničeni na riječi ili jednostavne izraze, već je Toolkit iznenađujuće kvalitetno prevodio i dulje, složenije rečenice. S druge strane, kvalitetni Smartcatovi strojni prijevodi uglavnom su se bazirali na jednostavnim riječima i izrazima dok se za niti jednu rečenicu ne može reći da je savršeno prevedena. Razlog puno boljih strojnih prijevoda Google Translator Toolkita može biti i činjenica da se prijevodne memorije u Toolkitu mogu dijeliti i globalno. Točnije, korisnici iz cijelog svijeta dijele svoje prijevodne memorije te je mnogo više konstrukcija rečenica na raznim jezicima dostupno. Sukladno tome, podaci koji se unose u Toolkit kvalitetnije su obrađeni. Nadalje, Toolkit generira strojne prijevode pomoću Google Translate-a kojeg koriste milijuni ljudi svakodnevno pa se i njihovi unosi procesiraju i pamte.

Oba alata nisu se pokazala praktičnima pri korištenju konvertiranih .pdf formata dokumenata. Smartcat je u nekim situacijama u sustav prenio krivo organizirane segmente, a slike koje su sadržavale tekst loše je učitao, no treba napomenuti da je originalni konvertirani dokument bio lošije kvalitete. Unatoč tome, pokazao se boljim nego Toolkit koji je u potpunosti poremetio prevedeni dokument. Dokumenti u .doc formatu imali su 10 stranica, no nakon prevođenja i preuzimanja prijevoda, taj broj se popeo na 40-ak stranica. Segmenti su bili razdvojeni, sa velikim prazninama između njih, neke su stranice sadržavale samo jedan ili dva segmenta, a većina slika uopće nije bila učitana. S druge strane, bilješke korištene u .html formatu izgledale su identično u dokumentima na izvornom i ciljnom jeziku.

Iako već spomenuta Toolkitova jednostavnost ne mora biti loša stvar, ponekad je zbog nje cijeli proces prevođenja donekle ograničen. Primjerice, više prijevodnih memorija moguće je koristiti jedino ako se preuzmu na računalo, a zatim prenesu u novo kreiranu prijevodnu memoriju. S obzirom da je sustav javljao grešku prilikom pokušaja preuzimanja prijevodnih memorija, nije bilo moguće isprobati kako se prijevodi ponašaju kada im je pridruženo nekoliko različitih prijevodnih memorija. U kontekstu vremena koje je bilo potrebno za prevesti dokumente, Toolkit se pokazao korisnijim upravo zbog kvalitetnih strojnih prijevoda zbog kojih nije bilo potrebno previše korigirati tekst, već se potvrđivanje segmenata odvijalo prilično brzo.

Iako su neke mogućnosti i nemogućnost jasne, treba napomenuti da su neke preference ipak subjektivne. Osobno, bolje sam se snašla sa radom u Smartcatu prvenstveno jer mi je korisničko sučelje bilo vizualno privlačnije i organiziranije. Iako sučelje oba alata relativno slično izgleda, dobila sam dojam da je Toolkit na neki način nepregledniji. Također, Toolkit je u svakom segmentu umetao jako mnogo oznaka oblikovanja što je bilo zbunjujuće i nezgodno za prevoditi. Uzevši u obzir sve do sada navedeno, presuda ide u korist Smartcata zbog njegove organiziranosti, vizualne privlačnosti te kvalitetnog načina obrade teksta. Ipak, treba pohvaliti Toolkit zbog iznenađujuće kvalitetnih strojnih prijevoda i uštede vremena.



## 8. ZAKLJUČAK

U doba kada se profesija prevoditelja sve više razvija i potražuje, utješno je znati da tehnologija ide u korak s lingvistikom. Brojni su alati dostupni prevoditeljima u njihovom poslu, a mnoštvo ih se još uvijek razvija i adaptira. Zato je nužno da prevoditelji uoče korist tehnologije te ju nauče iskoristiti za vlastito poboljšanje i napredak. Kako bi se pobliže ispitala korist *CAT* alata u procesu prevođenja te analizirala kvaliteta prijevoda, u ovom diplomskom radu prevodili su se dokumenti uz pomoć Google Translator Toolkita i Smartcata.

Svrha rada bila je sagledati na koji način ova dva *CAT* alata pospješuju, odnosno koče, proces prevođenja. Za prevođenje su se koristili materijali za učenje programiranja. U Smartcatu se prevelo pet dokumenata, a u GTT-u četiri. Pretpostavka da će bolje rezultate polučiti Smartcat temeljila se na njegovim značajkama i mogućnostima koje su u nekim aspektima kvalitetnije nego li značajke Toolkita. To se odnosi na mogućnost provjere kvalitete prijevoda, detaljnog postavljanja automatskog prijevoda, jednostavnog korištenja više prijevodnih memorija i rječnika te mogućnosti preuzimanja dvojezičnih dokumenata, kao i statistike projekta. Ipak, treba uzeti u obzir da Toolkitova jednostavnost nije uvijek nedostatak.

Iako se Smartcat pokazao boljim u manipuliranju tekstem te je tijekom procesa prevođenja pokazao koliko su korisne značajke poput kontrole kvalitete prijevoda, Toolkit je generirao iznimno kvalitetne rezultate u kontekstu strojnih prijevoda. Dok su se precizni strojni prijevodi u Smartcatu javljali najčešće kod jednostavnih riječi i konstrukcija, GTT je pokazao nadmoć na tom području te je čak i složenije rečenice savršeno prevodio. To je omogućilo iznimno brzo prevođenje i uštedilo mnogo vremena. Unatoč tome, Smartcatovi nedostaci bili su nezamjetni u odnosu na Toolkitove. Glavna zamjerka GTT-u je njegovo korisničko sučelje koje je, iako jednostavno i minimalističko, vizualno neprivlačnije od Smartcatovog. Također, Toolkit u svakom segmentu postavlja jako mnogo oznaka oblikovanja što prevođenje teksta čini nezgodnim i zbunjujućim te se teško snalaziti pri radu na segmentima. Uzevši sve u obzir, zaključak je da su Smartcatove funkcionalnosti i privlačnije korisničko sučelje u konačnici presudili u njegovu korist.

## LITERATURA

- A Step-by-Step Guide on How to Use the Google Translator Toolkit.* (21. svibanj 2016.).  
Dohvaćeno: 14. rujan 2018. iz Translation Rules: <http://www.translationrules.com/a-step-by-step-guide-on-how-to-use-the-google-translator-toolkit/>
- Craciunescu, O., & Gerding-Salas, C. (Srpanj 2004.). *Machine Translation and Computer-Assisted Translation: a New Way of Translating?* Dohvaćeno: 15. kolovoz 2018. iz Translation Journal: <http://translationjournal.net/journal/29computers.htm>
- Delpech, E. M. (2014.). *Comparable Corpora and Computer-assisted Translation.* London: ISTE Ltd.
- Frederking, R. E., & Taylor, K. B. (2005.). *Machine Translation: From Real Users to Research.* Washington DC: Springer Science + Business Media, Inc.
- Freelanceri? Tko su freelanceri?* (13. lipanj 2012.). Dohvaćeno: 23. kolovoz 2018. iz franjinoblog: <https://franjinoblog.wordpress.com/2012/06/13/tko-su-freelanceri/>
- Garcia, I., & Stevenson, V. (Rujan 2009.). Google Translator Toolkit: Free web-based translation for the masses. *MultiLingual*, str. 16-18.
- Google Inc. (2018.). *Translator Toolkit Help.* Dohvaćeno: 28. srpanj 2018. iz Google Translator Toolkit: <https://support.google.com/translatortoolkit/?en=#topic=6306364>
- Google Translator Toolkit.* (13. kolovoz 2018.). Dohvaćeno: 23. kolovoz 2018. iz Wikipedija: [https://en.wikipedia.org/wiki/Google\\_Translator\\_Toolkit](https://en.wikipedia.org/wiki/Google_Translator_Toolkit)
- Kučiš, V. (2010.). Prevodilački alati u funkciji kvalitete prijevoda. *Informatologia*, 43 (1), 19-33.
- Raspberry Pi Foundation. (2018.). *Code Club Projects.* Dohvaćeno: srpanj 2018. iz Code Club: <https://www.codeclub.org.uk/>
- Smartcat.* (9. srpanj 2018.). Dohvaćeno: 29. srpanj 2018. iz Wikipedija: [https://en.wikipedia.org/wiki/Smartcat#Document\\_formats](https://en.wikipedia.org/wiki/Smartcat#Document_formats)
- Smartcat Platform Inc. (2018.). *Smartcat Help Center.* Dohvaćeno: 20. srpanj 2018. iz Smartcat: <https://help.smartcat.ai/hc/en-us>

## POPIS SLIKA

Slika 1. Početna stranica Toolkita .....	14
Slika 2. Ekran alata.....	15
Slika 3. Prikaz prijevodne memorije u prilagođenoj pretrazi prijevoda.....	17
Slika 4. Prikaz rječnika u prilagođenoj pretrazi prijevoda .....	18
Slika 5. Strojno preveden segment .....	19
Slika 6. Izmijenjen segment .....	19
Slika 7. Rezultati pretrage prijevodne memorije.....	19
Slika 8. Izgled Toolkita .....	20
Slika 9. Prilagođena pretraga prijevoda .....	20
Slika 10. Kreiranje novog projekta .....	27
Slika 11. Pripremanje projekta za rad .....	28
Slika 12. Napredne postavke .....	28
Slika 13. Podešavanje automatskog prevođenja .....	29
Slika 14. Ekran alata.....	29
Slika 15. CAT kartica .....	30
Slika 16. Dodavanje rječnika .....	36
Slika 17. Kreiranje nove prijevodne memorije .....	36
Slika 18. Isječak iz rječnika „ <i>Sretan rođendan</i> “ .....	37
Slika 19. Statistika projekta „ <i>Sretan rođendan</i> “ .....	38
Slika 20. Statistika projekta „ <i>Ispričaj priču</i> “ .....	39
Slika 21. Statistika bilješki projekta „ <i>Sretan rođendan</i> “ .....	41
Slika 22. Statistika bilješki projekta „ <i>Ispričaj priču</i> “ .....	41
Slika 23. Greška pri prijenosu .pdf dokumenta u Smartcatu.....	44
Slika 24. Pripremanje okruženja za rad u Smartcatu.....	46
Slika 25. Izgled dvojezičnog dokumenta .....	47

Slika 26. Izgled dokumenta sa popisom grešaka.....	49
Slika 27. Podešavanje automatskog prevođenja .....	54

## POPIS TABLICA

Tablica 1. Obilježja korisničkih računa u Smartcatu .....	22
Tablica 2. Statistika korpusa .....	35
Tablica 3. Prikaz podudarnosti u projektima „Sretan rođendan“ i „Ispričaj priču“ .....	40
Tablica 4. Prikaz podudarnosti u bilješkama projekata „Sretan rođendan“ i „Ispričaj priču“ .....	42
Tablica 5. Statistika korpusa .....	45
Tablica 6. Popis grešaka u projektu „Istjerivači duhova“ .....	47
Tablica 7. Popis grešaka u projektu „Mačke!“ .....	50
Tablica 8. Popis grešaka u projektu „Smeće iz svemira“ .....	52
Tablica 9. Prikaz podudarnosti u projektu „Smeće iz svemira“ .....	53
Tablica 10. Popis grešaka u bilješkama projekta „Istjerivači duhova“ .....	54
Tablica 11. Prikaz podudarnosti u bilješkama projekta „Istjerivači duhova“ .....	55
Tablica 12. Popis grešaka u bilješkama projekta „Smeće iz svemira“ .....	56
Tablica 13. Prikaz podudarnosti u bilješkama projekta „Smeće iz svemira“ .....	57
Tablica 14. Korištene funkcionalnosti po projektu .....	60
Tablica 15. Broj grešaka po kategorijama za pojedini dokument .....	61
Tablica 16. Broj podudaranja sa prijevodnim memorijama, rječnicima i strojnim prijevodima za pojedini dokument .....	61

## **POPIS GRAFIKONA**

Figure 1. Postotak grešaka po prijevodu .....	59
---	----